CONSISTENCY AND VARIATION IN CLASSROOM PRACTICE:

A MIXED-METHOD INVESTIGATION BASED ON

CASE STUDIES OF FOUR EFL TEACHERS OF

A DISADVANTAGED SECONDARY SCHOOL IN HONG KONG

James Yue-on KO, MSc.

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

JAN 2010

For my teachers, students and friends

ABSTRACT

This mixed methods study was based on teacher case studies examining classroom practices of four EFL teachers of the same department of an underperforming secondary school in a socially-disadvantaged area in Hong Kong. Beside two international classroom observation instruments used for the quantitative classroom observations, extensive qualitative field notes were collected concurrently. Confirmatory factor analyses using the lesson as the unit of analysis generated a six-factor and a three-factor of model teaching behaviours respectively. For both instruments, results showed strong validity and reliability for strongly correlated underlying dimensions of teaching practices Considerable differential teaching effectiveness in terms of inconsistency in observed teaching behaviours of the four teachers was noted across the various dimensions and across contexts. The qualitative field notes provided evidence that increased understanding of the variation in observed practice. Two teachers showed teaching behaviours more inconsistent across dimensions and lessons, though their effectiveness in certain dimensions in some lessons was found. Their fluctuating teaching effectiveness seemed to be under the influences of student year groups, class composition, subject content, school policy on learning, rather than class size. Themes emerged from the interviews with these teachers, the department head and the school principal suggested that cultural and school contexts might result in inconsistent teaching behaviours and revealed challenges and contradictions at individual, department, school, and system levels. This study was significant in demonstrating that both the generic and differentiated theories of teacher effectiveness may be required to account for the full spectrum of observed teaching behaviours. It also contributed to testing validity and reliability of two classroom observation instruments as it indicated that the high-inference instrument used by the inspectors might be slightly better in predicting overall judgment of lesson quality, while the lower inference instrument developed by the academics tended to generate underlying dimensions that were more distinguishable.

ACKNOWLEDGEMENT

I have dedicated this thesis to all my teachers, students, and friends as I feel I am indebted to them the knowledge and life experience that have not only changed the course of my life, but made it fuller and more enjoyable.

I missed a chance to complete my MSc in linguistics and pursue a PhD in 1987, when I was perplexed whether I would be happy with an academic career. Then, after another four years in business, I started teaching in the school where I did this PhD research. Although my initial teaching experience was not easy, my early students were supportive and wished me to be a different breed.

My early devotion to research can be dated back to my sixth form years at St Paul’s College, where I learned from Mr. K. W. Lo and Ms Josephine Ho and developed a learning attitude that goes beyond the practical aim of passing the high stake exams.

Various professors at the University of New Brunswick stimulated my mind in 1983. The professors at three departments between 1984 and 1987 at the University of Alberta were great teachers with scholarship and unique personalities: Dr Brue Derwing, Dr Matthew Dryer, Dr John Hogan, Dr Gary Prideaux, and Dr. Lois Stanford of Linguistics, Dr Leo Mos of Psychology and Dr Raymond Morrow of Sociology. They helped maintain my life-long interests in these subjects and I can still hear their voices when I read books on these subjects.

My interest in Education studies owe much to the teaching of Prof David Coniam at the Chinese University of Hong Kong and Prof Alan Walker at the Hong Kong Institute of Education when I studied my DipEd at CUHK in 1995. However, the MEd program of the University of Melbourne ran with HKIEd in 2005 provided me a solid foundation for the PhD. There were scholars with diverse specialties, Dr Lawrie Drysdale and Dr David Gurr on educational management, Prof Richard Teese on educational policy, Prof Patrick Griffin on educational assessment, and Prof Brian Caldwell on school leadership and strategic management.

I decided to do the MSc in Psychological Studies at the University of Glasgow before starting my PhD at Nottingham. I believed that this allowed me to develop a stronger base on quantitative research. I received much encouragement from Dr. F. C. Ho and Dr Kenneth Sin at the HKIEd in extending my project supposed to be done in Hong Kong.

Dr Peter Gates delivered a productive research training program at the School of Education, but I received more thorough training on quantitative research through the various clinics run as the Advanced training in Quantitative Methods for the Social Sciences at the Methods and Data Institute. Prof. Cees van der Eijk has inspired me much and solved many puzzles for me in these clinics.

I am grateful to the comments by Prof Daniel Muijs and Dr Qing Gu on the thesis. However, my supervisors, Prof Chris Day and Prof Pam Sammons, have inspired me much with their Effective Classroom Practice Project, which set all the groundwork for this thesis. My deepest gratitude is for Pam as she has been particularly kind and supportive, making my PhD journey more tolerable. I could not survive without her detailed comments in the several revisions of the thesis and in the preparation for the viva. Her valuable advice in the countless number of emails was my beacon when I was in Hong Kong or when she left for her new post at Oxford.

My fellow students at Nottingham, Mr Musa Khaldi and Mr Nadeem Khan, have maintained a constructive dialogue with me. My ex-colleagues at Ming Tak, Mr Yuen, Bear Chung, Fiona, Angel, Ms Hui, Kitty, Deby, Michael, and Ms Chan gave me the rapport I longed for. Special thanks are given to the four case study teachers too. My high school buddies, Auster, Richard, Lam, and Chun, have been supporting me like bridges over troubled water for thirty years. Our brotherhood is vintage wine.

Finally, I have been blessed with a lovely and sympathetic wife, Teresa, who has been taking care of the family to free me from worries and distractions. My son, Eldon, has also taken up lots of chores to keep my work going. As a widow, my mother should also be proud of her raising me as an educated man single-handedly. All of them are my treasures in this life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CFA	*confirmatory factor analysis*
CMI	*Chinese as medium of instruction*
CVCP	***c***onsistency and ***v***ariation in ***c***lassroom ***p***ractice; the present study
DEE	*Dynamic model of educational effectiveness*
DTE	*Differentiated model of teacher effectiveness*
ECP	*Effective Classroom Practice*
EFA	*exploratory factor analysis*
EMB	*Education and Manpower Bureau of Hong Kong*
EDB	*Education Bureau of Hong Kong*
EER	*Educational effectiveness research*
EFL	*English as a foreign language*
EMI	*English as medium of instruction*
ESRC	*Economic and Social Research Council*
GTE	*generic theory of teacher effectiveness*
HKCE	*Hong Kong Certificate Examination*
ISTOF	*International Schedule for Teacher Observation and Feedback*
ICSEI	*International Congress for School Effectiveness and Improvement*
L1	*first language, mother tongue; that is, Cantonese in the CVCP study*
L2	*second language; that is, English in the CVCP study*
MM	*mixed methods*
MOI	*medium of instruction*
MORE	*Methodology of Research in Effectiveness (MORE) group*
ORACLE	*Observational Research and Classroom Learning Evaluation*
PEE	*Probability model of educational effectiveness*
PISA	*The OECD Programme for International Student Assessment*
QoT	*Quality of Teaching Schedule*
SEN	*special education needs*
SER	*school effectiveness research*
SSPA	*Secondary School Places Allocation*
TER	*teacher effectiveness research*
TIMSS	*Trends in International Mathematics and Science Study*
VIF	*variance inflation factor*

CHAPTER 1 : INTRODUCTION

1.1 Introduction

In Hong Kong, the need to understand, assess, and monitor teacher effectiveness has been increasing in its current context of education reforms that emphasise accountability and quality assurance (Cheng & Tsui, 1996, 1999; see Fok, 2004; Education Bureau (hereafter EDB, 2010a) for details of the education reforms). Yet, there is a gap between this growing need to address teaching quality in teacher evaluation and development and the lack of knowledge bases on teacher and school effectiveness in Hong Kong (Lee, Lam, & Li, 2003). This research is intended to enhance those knowledge bases by examining the consistency and variation in classroom practice of four teachers of the same department in a school and linking their interplay with the internal factors of that school and the wider-contextual factors in Hong Kong. This research responds to the debate on effective teacher and effective teaching between the differentiated model of teacher effectiveness (e.g., Campbell, Kyriakides, Muijs, & Robinson, 2004) and the dynamic model of educational effectiveness (e.g., Creemers & Kyriakides, 2008). To achieve this, for the purpose of this study, effectiveness is defined as perceived effectiveness based on mainly the results of the use of two rating scales, supplemented with teachers' self-reported practices collected through a survey and an interview (see Section 3.5.3 for details) and thus differs from a value-added definition[1] (see next section for details). The mixed-method (hereafter MM) approach of this research represents an attempt to study teaching as a complex process that shows multidimensionality and offer the prospect of providing a richer evidence base to promote new understandings and contributions to knowledge that would inform policy and practice than studies adopting either quantitative or qualitative approach.

In teacher evaluation and development, consistency is understood in terms of a requirement on behavioural expectations and responses for effective teachers, especially regarding their classroom management and

1 That is based on value-added measures as "a collection of complex statistical techniques that use multiple years of students' test score data to estimate the effects of individual schools or teachers" (McCaffrey, Lockwood, Koretz, & Hamilton, 2003, p., xi).

organisation (e.g., Stronge, 2007). Nevertheless, effective teachers may be expected to excel in different dimensions of teaching, though excellence across dimensions has not been addressed as consistency in performance related to effectiveness. Given that teaching is seen as a complex, multidimensional process (Kyriacou, 2007; Muijs & Reynolds, 2000), high quality teaching is expected to be consistent across dimensions (Marzano, 2003) and across contexts (e.g., different student compositions, classes and year levels). Muijs and Reynolds (2000) found that primary teachers who scored highly in various teaching dimensions tended to have stronger and stable impacts on student progress in numeracy, but they used a composite score of effective behaviours in seven dimensions rather than looking at variation across dimensions[2].

Recently, in their ESRC-funded Effective Classroom Practice (hereafter ECP) Project, Day, Sammons, Kington and their colleagues (2008) examined variation across teaching dimensions in effective teachers and found a positive correlation between teaching quality and teacher quality, because effective teachers in their sample tended to be effective in all ‘core’ aspects of teaching practices. However, there are doubts about whether effective teachers can be effective across all contexts and at all times (e.g., Campbell, et al., 2004). Day et al. (2008) also showed that variation across some teaching dimensions (e.g., catering for individual differences) was high among effective teachers and seemed to be subtly related to subjects (e.g., English vs Mathematics) and school levels (e.g., primary vs secondary schools) in quantitative comparisons. Despite these results, evidence on multidimensionality of effective teaching behaviours is limited and rarely explored in terms of consistency, whether across contexts or across dimensions. Against this background, this research has examined teacher behaviours in extent occasions and variation across different dimensions of effective teaching behaviours.

2 There were nine dimensions originally measured in nine subscales: classroom management, behaviour management, direct teaching, individual practice, interactive teaching, varied teaching, mathematical language, classroom climate and constructivist methods. Muijs and Reynolds (2000) found strong correlations between factors except that the correlations between the constructivist methods and mathematical language scales were only weak to moderate. They argued that using a composite effective teaching score in the multilevel analyses (excluding constructivist methods and mathematical language) was “to help avoid multicollinearity which could otherwise result from using the highly intercorrelated teaching scales as predictors in the analyses” (Muijs & Reynolds, 2000, p.286).

Campbell et al. (2004) regard the fundamental challenges of teacher effectiveness research as follows: first, to identify *consistency and stability* of teacher effects; and second, to attribute *variations* in teacher effects to differentiations in pupils' background characteristics, personal characteristics, professional histories, and working environments. In teacher effectiveness research (hereafter TER), consistency in teaching quality and variation in teaching strategies are not contradictory, but often seen as indicators of effective teaching because the latter may only reflect the teacher's instructional skills and abilities to cater for individual differences in students. For example, (Darling-Hammond, 2007; 2008) argues that high-quality instruction exists when the teacher's knowledge and skills match with the demands of situation, and when the conditions for instruction are appropriate and desirable. The former case places an emphasis on what the teacher can command, while the latter case stresses how the school or subject-departmental policies and the organisational contexts of the school may constrain or facilitate classroom practices (e.g., Coleman, 1987; Purkey & Smith, 1983; Sammons, Thomas, & Mortimore, 1997; Scheerens, 1990, 1992; Scheerens & Creemers, 1989).

In school effectiveness research (hereafter SER), embedded in the major theme of between-school variation was the theme of within-school variation in teaching performance. Both Scheerens and Bosker (1997) and Sammons (1999) summarise numerous research that showed positive correlation, but considerable variation in school performance across subject and years (e.g., Luyten, 1994; Ma, 2001; Mortimore, Sammons, Stoll, Lewis, & Ecob, 1988; Sammons et al., 1997; Thomas, Sammons, Mortimore, & Smees, 1997a, 1997b; Willms & Raudenbush, 1989). Stronger subject consistency was generally found between English and mathematics in primary schools (e.g., Sammons, Nuttall, & Cuttance, 1993; Sammons, West, & Hind, 1997) than their secondary counterparts (e.g., Cuttance, 1987; Luyten, 1994; Sammons, Mortimore, & Thomas, 1996; Thomas, Pan, & Goldstein, 1994; Thomas et al., 1997a, 1997b).

In relating consistency to effectiveness in classroom and in school, (Creemers, 1994, p.95; see also Section 2.4.4) elaborates the concept of consistency as an integrative principle of instruction: “the same characteristics of effective teaching should be apparent in the different components [of instruction, i.e., the curriculum, group procedures and teacher behaviour]”. Thus, at the classroom level, effective teachers are expected to be consistent in the quality of their teaching while varying their teaching strategies with respect to the situational demands, such as the ability and level of the students or the subject and the lesson topic. To achieve departmental effectiveness, teachers of the same department need to maximise their own effectiveness with consistent and stable effective teaching behaviours as well as minimise the variation in performance among themselves. Thus, there is a need to examine whether the same principle applies to the school level, whether a school would implement the consistency principle as reflected in its leadership and teaching and learning policies may affect the effectiveness of a department and the effectiveness of individual teachers of the department.

The following section begins with defining some of the key terms used in this research, as these definitions would help explain the present focus on teachers’ classroom practices. Then, in Section 1.3, teacher effects are shown to be the focus in various domains of educational effectiveness research. This serves to explain why it is crucial to address educational effectiveness at departmental and school levels with reference to the teacher level. In particular, the rationale behind the distinction made between teacher effectiveness and teaching effectiveness is clarified. It is then argued that a research strategy to study consistency and variation of the teaching effectiveness of individual teachers of a single subject department is justified as it would inform the school’s provision of learning on the subject. Finally, the rationale of studying teaching effectiveness through classroom observation is discussed with a conclusion that there is a need to use different classroom observation schedules. Before the concluding summary outlining the thesis structure, Section 1.4 clarifies and examines the aims and context of inquiry of this research

1.2 Defining and measuring teaching effectiveness, consistency and variation in classroom practice

1.2.1 Teaching effectiveness versus teacher effectiveness

While there are researchers like Scheerens (2004, 2008), who would regard terms such as instructional effectiveness, teacher effectiveness and teaching effectiveness interchangeable, the present thesis calls for a distinction between teaching (or instructional) effectiveness and teacher effectiveness for the sake of conceptual clarity. In its most restrictive sense, *teaching effectiveness* refers only to *the effectiveness of observable behaviours occurred during a classroom observation*. This definition *excludes a range of instruction-related* activities that teachers normally do *inside* the classroom (e.g., set up the facilities) and *outside* the classroom (e.g., lesson planning and marking). In contrast, teacher effectiveness in its the broadest sense includes behaviours, activities and classroom practices of teachers that contribute to better student learning and outcomes such as attainment, motivation and engagement behaviour. This definition does not exclude a range of *non-instructional* activities *that teachers normally do outside classroom* to facilitate student learning like student learning enhancement programs, extra-curricular activities, consultations with parents on students' work, and even clerical work and paper work that in principle could be taken up by other non-teaching staff. Campbell et al. (2004, p.15) argues that a model of teacher effectiveness "would be anachronistic" if it ignores the increase in working time of English teachers and focuses just on their relatively constant 18 hours a week of instruction time in the classroom.

In this research, the term teaching effectiveness is used restrictively to refer to *the effectiveness attributable to observable teacher behaviours and practices in the classroom*. It has been suggested that "the main criterion of an effective teacher is the extent to which his/her students achieve *specific educational goals*" (Campbell et al., 2004, p.61). It follows that teaching effectiveness can be further defined as *certain generic characteristics of classroom practices observed during classroom observation which are believed to have enhanced students' learning the designated educational* *goals of the lesson observed*[3]. By no means, this definition of teaching effectiveness excludes or contradicts the value-added approach based on relative student progress widely used in TER and SER.

In the present context, the distinction between teacher effectiveness and teaching effectiveness is also based on a practical concern to limit the scope of the study. Since Chapters 4, 5 and 6 will discuss those results primarily using the lesson as the unit of analysis, only teaching effectiveness is addressed. When the case study results are presented from Chapter 7 onwards, they will indicate how the *consistency and variation in the teaching* effectiveness of individual teachers may eventually affect their *teacher* effectiveness. As indicated above, teacher effectiveness is a much broader term that should include many other roles of the teachers nowadays. Investigations on these different roles are important but beyond the limits of present doctorial research. Since different studies may have different units of analysis, either the teacher or the lesson, it is advantageous to limit the scope to a term that would focus on teacher behaviours in the classroom when resources are limited, as in the case of the current thesis research. Teacher effectiveness is thus hereby restricted to a limited scope of comparing variation in teaching behaviours or classroom practices among teachers. Teacher behaviours in the classroom are under-researched in Hong Kong, comparing to the amount of research in other countries like the U.S.A., the Netherlands, and the U.K. Nevertheless, restricting the research focus on teacher behaviours in the classroom would highlight the current interest on their consistency and variation across dimensions and across lessons, no matter whether the teacher or the teaching is examined.

1.2.2 Measuring effectiveness by classroom observation instruments and by global indicators

It is imperative to clarify that the term "*effectiveness"* used in this research was restricted to the *observed measures of teaching effectiveness*. It was measured in terms of the ratings in two classroom observation instruments and the ratings of two global indicators which quantify the

3 Observations of learning were recorded mainly in the field notes or as the overall judgement of student participation.

impressions concerning the overall teaching quality of the lesson and the level of individual involvement by the pupils. Details of these measurement tools are discussed in Chapter 3 on the methodology. *Effectiveness* is thus predefined in the instruments and in the global indicators. For example, "*starting the lesson on time*" is predefined as an effective classroom practice in a particular instrument and it is easy to observe. However, in practice, during classroom observation, many teaching behaviours may not be as easily observable as "*starting the lesson on time.*" For example, "*the appropriateness of a teacher's correction of a students' misbehavior*" can be evaluated only when it actually occurs. Its appropriateness may vary context by context, dependent on the frequency and impact of the misbehavior. This has highlighted a general problem of using a classroom observation instrument that contains many items that require the inference of the observer/rater to reach a judgment. When more inferences are required in the judgment process, observed teaching effectiveness is more likely to be subjective (see also Section 9.6.3 for the limitations).

Each classroom observation instrument consists of several groups of descriptive statements about teaching behaviours hypothesised to represent different aspects of teaching, such as classroom management, clarity of presentation, and others. Scoring highly in a particular group would thus indicate the teacher's strength in a particular hypothesised aspect of teaching. Certainly, it is an empirical question whether the hypothesised aspects of teaching of an instrument would resemble the underlying dimensions of the observed teaching behaviours. Accordingly, employing reliable and valid instruments that have been used and tested in different contexts to capture multidimensionality of teaching behaviours is one of the main characteristics of this research study.

1.2.3 Focusing on the lesson versus the teacher

While teaching practices in *individual lessons during the classroom observation period* are the focuses in the first part (Chapters 4 to 6) of this research, the unit of analysis is *the individual lesson*. However, consistency and variation is explored in the individual case study teacher's practices across multiple lessons of the observation period in the second part

(Chapters 7 and 8) of the research. It is important to note that *the individual teacher* is the unit of analysis in both the case and the cross-case analyses. This contrasts with the common practice of using *the teacher* as the focus of analysis in most TER and teacher evaluation (or appraisal), where the teacher is usually the primary research focus. Since both TER and teacher evaluation involve using lesson observation for appraisal, it is essential to highlight the subtle difference between the current approach and that adopted in teacher evaluation. Thus, in this research the term 'teaching effectiveness' applies to lesson comparisons (as discussed earlier on p.6) and the term 'teacher effectiveness' is based on the overall evidence of teaching effectiveness derived from the lesson observations and then used to categorise the four teachers' practices.

1.2.4 Defining and measuring consistency and variation

Consistency is not a new concept in SER. According to Mortimore et al. (1988), consistency among teachers was among the key factors contributing to school effectiveness. Creemers (1994) regarded consistency as the integration of effective characteristics of instructional components in the classroom and as a key formal principle in his theoretical model of educational effectiveness. According to Creemers (1994), the school has to combine and coordinate different effectiveness variables like teaching behaviour, teaching materials, and group composition to produce lasting effect on student achievement. However, research evidence has not showed a strong support for consistency as a predictor of student achievement (Driessen & Sleegers, 2000; Kyriakides, 2008). This may be because Creemers (1994) and these researchers have not adopted an adequate operational definition for the concept that can be linked with the measurement of effectiveness more directly.

Contrary to the definitions in previous research, *consistency* is operationally defined in this research as *little variability found in the patterns of underlying dimensions of observed teaching behaviours*, while *variation* as *apparently large variability found in the patterns of underlying dimensions of observed teaching behaviours.* Although consistency cannot be equated with effectiveness defined in value-added terms in this research, highly effective teachers in Day et al. (2008) tended to be rated highly across different underlying dimensions of observed teaching behaviours. Consistency and variation exist basically across different underlying dimensions of observed teaching behaviours in five circumstances: 1) across lessons of all teachers observed; 2) across all lessons of each individual teacher observed; 3) between teachers; 4) between instruments in their association to teaching effectiveness; and 5) across samples of different countries/cultural contexts. The first circumstance is crucial for establishing the multidimensionality of teaching behaviours. The second and the third circumstances are important for understanding each teacher observed as well as their similarities and differences. The fourth circumstance sheds light on the similarities and differences between instruments and their reliability and validity in measuring the multidimensionality of teaching behaviours. The fifth circumstance is relevant when the generalisability of the multidimensionality of teaching behaviours is of interest. Further discussion on consistency and its relation with the theoretical frameworks can be found in Section 2.5.4. As the current research focused on ***c***onsistency and ***v***ariation in ***c***lassroom ***p***ractice, an acronym called CVCP is used hereafter to refer to the main study.

1.3 Localising teacher effects in the web of educational effectiveness

1.3.1 The roles of the teacher and the conceptualisation of teacher effectiveness

As modern schools are part of a country's hierarchical education system, both Creemers and Scheerens (1989) and Bosker and Scheerens (1989, 1994) define the terms 'instructional effectiveness', 'school effectiveness', and 'educational effectiveness' in respect to the relative impacts on student outcomes by factors that operate at the classroom, the school and the education system levels. Educational effectiveness can be used in two other senses, either broadly referring to effectiveness at different levels or, as in the present research, restrictively referring to only "the interactions between the school, classroom, and individual student levels and their contributions to students' performance" (Campbell et al., 2004, p.3).

The term 'instructional effectiveness' is not exactly equivalent to 'teacher effectiveness', which originated in educational psychology, has a longer history in the traditional TER than SER (Teddlie, 1991). In a rather restrictive sense, Campbell et al. (2004) refer it to the impact on student's performance by classroom factors like teaching methods, teacher expectations, classroom organisation, and use of classroom resources. In a recent research synthesis of approaches to evaluating teacher effectiveness, Goe, Bell, and Little (2008, p.8) have proposed the following five-point definition of effective teachers:

- Effective teachers have high expectations for all students and help students learn, as measured by value-added or test-based growth measures, or by alternative measures.
- Effective teachers contribute to positive academic, attitudinal and social outcomes for students such as regular attendance, on-time promotion to the next grade, on-time graduation, self-efficacy, and cooperative behaviour.
- Effective teachers use diverse resources to plan and structure engaging learning opportunities; monitor student progress formatively, adapting instruction as needed; and evaluate learning using multiple sources of evidence.
- Effective teachers contribute to the development of classrooms and schools that value diversity and civic-mindedness.
- Effective teachers collaborate with other teachers, administrators, parents, and education professionals to ensure student success, particularly the success of students with special needs and those at high risk of failure.

This definition contrasts Medley's (1982, pp.1894-1895) definition, which was based on the literature of TER prior to the value-added approach in SER:

> ..... the results a teacher gets to the amount of progress the pupils make toward some specified *goals* of education. One implication of this definition is that teacher effectiveness *must be defined, and can only be assessed, in terms of behaviours of pupils, not behaviours of teachers*. For this reason, and because the amount that pupils learn is strongly affected by factors not under the teacher's control, teacher effectiveness will be regarded *not as a stable characteristic* of the teacher as an individual but as *a product of the interaction between certain teacher characteristics and other factors that vary according to the situation in which the teacher works*… [emphasis added]

Medley's emphasis of assessing pupil behaviours or learning outcomes and interaction of teacher characteristics and situation factors is based on a process-product model of teacher effectiveness (see Section 2.4). However, his proposal for the structure of teacher effectiveness is more complicated than his emphasis as it includes nine interrelated components: pre-existing teacher characteristics, teacher competence, teacher performance/ behaviour, students' learning experience, student behaviour or learning outcomes, teacher training, external teaching context, internal teaching context and individual student characteristics.

Later, Cheng (1995, 1996) proposes to add two more components, namely teacher evaluation and professional development, to conceptualise what he called 'total teacher effectiveness' (Cheng & Tsui, 1996). Similarly, Campbell et al. (2004, p. 81) argue that TER should attempt to measure teacher effectiveness with respect to "the different criteria for measuring effectiveness in the various duties of teachers in the 21st century". The duties of the teacher, and the school, have been overstretched probably because family and possibly the church cannot fulfil their traditional functions. Regarding potential cultural impacts on the conceptualisations of teaching effectiveness, Pratt, Kelly and Wong (1999) identify contrasts the views of effective teaching between western and Chinese educators. Putting forward a model of teacher effectiveness based on outstanding teachers in Hong Kong, Cheung, Cheng and Pang (2008) argue that the coexistence of personal attributes and professional qualities as well as contextual factors contribute much to the success or effectiveness of a teacher.

Muijs argues (2006, p.54) that the limitation of the simple process-product model and the typical academic outcome measure is that they cannot "take into account sufficiently the fact that teachers' roles are broader than their classroom practice and includes management roles, pastoral roles, and relationships with parents and community as well as classroom practice." Although these definitions incorporate broader aspects of the teacher's various roles, the value-added approach still prevails in the recent TER literature and usually uses academic achievement as the typical outcome measure and corrects it by "using earlier scores on the same measure as a predictor" (Muijs, 2006, p.57).

Recently, Day, Sammons, Stobart, and Kington (2006, 2007) propose a definition of teacher' effectiveness that is based on two related measures: teachers' perceived effectiveness (i.e., self-perception of teachers of their own practice) and relative effectiveness determined in terms of value-added measures of pupil attainment. Based on this definition, they find that teachers' effectiveness is not simply a consequence of age or experience, but influenced by variations in their work, lives and identities that directly affect teachers' senses of professional identity in their various professional life phases. In turn, teachers' senses of professional identity influence their relative commitment and resilience as well as their capacities to manage these variations to sustain effectiveness.

These findings by Day and his colleagues are important in two ways. First, they suggest that studies that simply control for age and teaching experience would miss important roles of personal, situated and contextual factors that shape professional identity of teachers and their capacities to manage variations and sustain their effectiveness. Second, teacher effectiveness is not an isolated characteristic in the teacher but a consequence of many interacting factors. This suggests that a teacher may be effective in different circumstances and at different times and thus, there is a need to examine the factors that affect teachers' observed teaching behaviours, their overall teaching effectiveness, and their variation and stability over time.

On the one hand, these various definitions by different researchers suggest that 'teacher effectiveness' is a malleable concept that its complexity changes with the expectations of what a teacher can do and should do as determined by personal factors, situated factors and professional factors. On the other hand, these definitions apparently extend a traditional TER construct on teacher behaviour and learning outcomes to an extent that probably no single researcher can accomplish to study. Accordingly, in the present context, both teacher effectiveness and teaching effectiveness are assumed to be based on teacher behaviours or classroom practices, and while a distinction between the two is made (as proposed in Section 1.3.1), the latter is assumed to constituent only part of the former. This restricted conceptualisation is justified on the fact that the major role of a teacher remains to be instructional.

1.3.2 Teacher effectiveness in TER and SER

Although teacher effectiveness is the shared focus of both TER and SER traditions, their approaches to study the impact of different classroom practices and processes on student outcomes are quite different. Goe and her colleagues (Goe et al., 2008; Goe & Croft, 2009) regarded classroom observation and value-added approach as the two most widely used teacher evaluation methods. Their comparison of the advantages and limitations of these two approaches are summarised in Table 1.1.

Table 1.1: **A comparison between two widely used teacher evaluation methods**

	Classroom Observation	**Value-added Models**
Descriptions	• Used to measure classroom processes, including specific teacher practices, holistic aspects of instruction, and interactions between teachers and students. • Can measure board, overarching aspects of teaching or subject-specific aspects of practice.	• Used to determine teachers' contributions to students' test score gains. • May also be used as a research tool (.e.g., determining the distribution of "effective" teachers by students or school characteristics).
Research	• Some highly researched protocols have been found, though and to link to student achievements are sometimes modest. • Research and validity findings are highly dependent on the instrument used, sampling procedures, and training of raters. • There is a lack of research on observation protocols as used in context for teacher evaluation.	• Little is known about the validity of value-added scores for identifying effective teaching, though research using value-added models does suggest that teachers differ markedly in their contributions to students' test score gains. • However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded mixed results and few significant findings. • Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.
Strengths	• Provides rich information about classroom behaviours and activities. • Is generally considered a fair and direct measure by stakeholders. • Depending on the protocol, can be used in various subjects, grades, and contexts. • Can provide information useful for both formative and summative purposes.	• Provides a way to evaluate teachers' contribution to student learning, which most measures do not. • Requires no classroom visits because linked student/teachers data can be analysed at a distance. • Entails little burden at the classroom or school level because most data is already collected for NCLB purposes. • May be useful for identifying outstanding teachers whose classrooms can serve as "learning labs" as well as struggling teachers in need of support.

	Classroom Observation	Value-added Models
Cautions	• Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters. • Classroom observation is expensive due to cost of observers' time; intensive training and calibrating of observers adds to expense but is necessary for validity. • This method assesses observable classroom behaviours but is not as useful for assessing beliefs, feelings, intentions, or out-of-classroom activities.	• Models are not able to sort out teacher effects from classroom effects. • Vertical test alignment is assumed (i.e., tests essentially measure the same thing from grade to grade). • Value-added scores are not useful for formative proposes because teachers learn nothing about how their practices contributed to (or impeded) student learning. • Value-added measures are controversial because they measure only teachers' contributions to student achievement gains on standardised tests.

Note: Adapted from Goe et al. (2008, pp. 18-19).

Given that most researchers in the traditional TER field, especially in the US, have educational psychology backgrounds, they tend to avoid school effect labels by studying teacher effectiveness in quasi-experimental studies or in naturalistic studies that involve quantifiable, low-inference data collection methods like systematic classroom observation instruments (Teddlie, 1991; Teddlie, Stringfield, & Burdett, 2003). In contrast, school effectiveness researchers look at the characteristics of effective teachers and effective schools through measuring the *differential impacts* between teachers, departments, and/or schools on their pupils' educational outcomes, while taking into account those differences in the prior attainments and other characteristics of the pupil intakes (Mortimore et al., 1988). However, in both TER and SER traditions, effective teachers are referred to those who can have a positive impact on promoting students' cognitive progress, academic as well as psychological outcomes such as self-esteem, attitudes to school, and motivation to learn. In contrast, less effective teachers are often defined as those whose students show outcomes poorer than the average after the results of the prior test and background factors are taken into account in TER or the predicted *on the basis of intake* in SER.

Despite the statistical sophistications evident in SER in the last two decades, school level value-added data cannot be used to measure the relative effectiveness of individual teachers. First, effectiveness is by definition always *relative* as the predicted outcome varies with the students or classes chosen for comparison in the sample (Scheerens & Bosker, 1997)[4]. Second, unlike the measures in experiments in traditional TER, which are usually obtained immediately after the experimental conditions, measures of effectiveness in SER are *retrospective* because they are very often taken at the end of the school term or upon students' graduation. This means that much of the information will be lost as the specific effects of certain teachers and/or the specific effectiveness of certain classroom practice is not be easily identifiable. When students are taught by different teachers, it is inappropriate to use results on school effects to infer individual teacher effects.

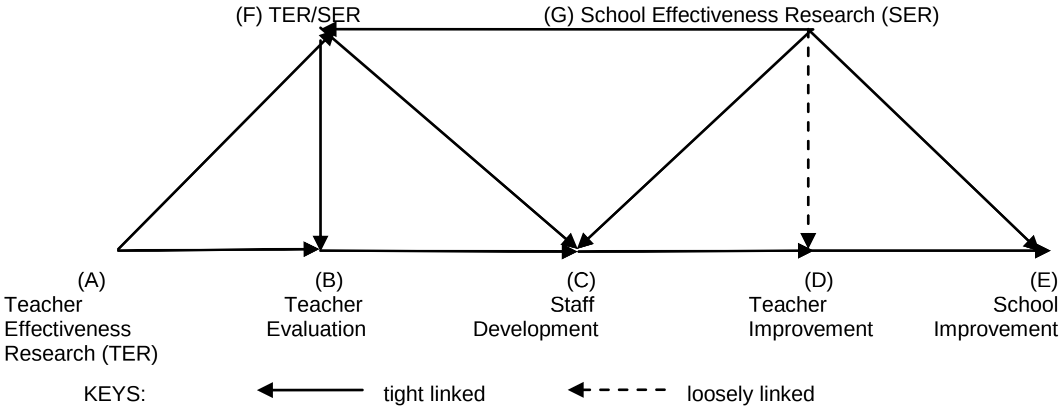
Multilevel modelling employed in SER has been used to identify the fixed effects of school, department, and teacher effects, but these results cannot inform what make the schools, departments and teachers with different results without data other than achievement data. Using both student outcomes and classroom observation data and both structural equation modelling and multilevel modelling, Muijs and Reynolds (2000) linked individual teacher effects more closely with student outcomes than the traditional SER methods that usually lack classroom observation data. This justifies the present research strategy to study individual teacher effectiveness with classroom observation data and a single subject department as a means to understand the school's provision of learning on the subject in more depth (see Section 3.3.3).

1.3.3 Classroom observation building linkages between domains

Other than experimental studies, naturalistic classroom observation is the major method of inquiry in TER and other different domains of educational research. Teddlie et al. (2003) show that there are theoretical links between SER/TER and teacher evaluation because classroom observations using variables from the TER literature may inform teacher evaluation, staff development, teacher development, and eventually teacher and school improvement. They depict the conceptual links between all these domains as in Figure 1.1 below.

4 This explains why contextualised value-added scores are often preferred as they take into account student and school background variables, rather than simply controlling student prior learning.

Figure 1.1: **Conceptual links among different domains of educational research**

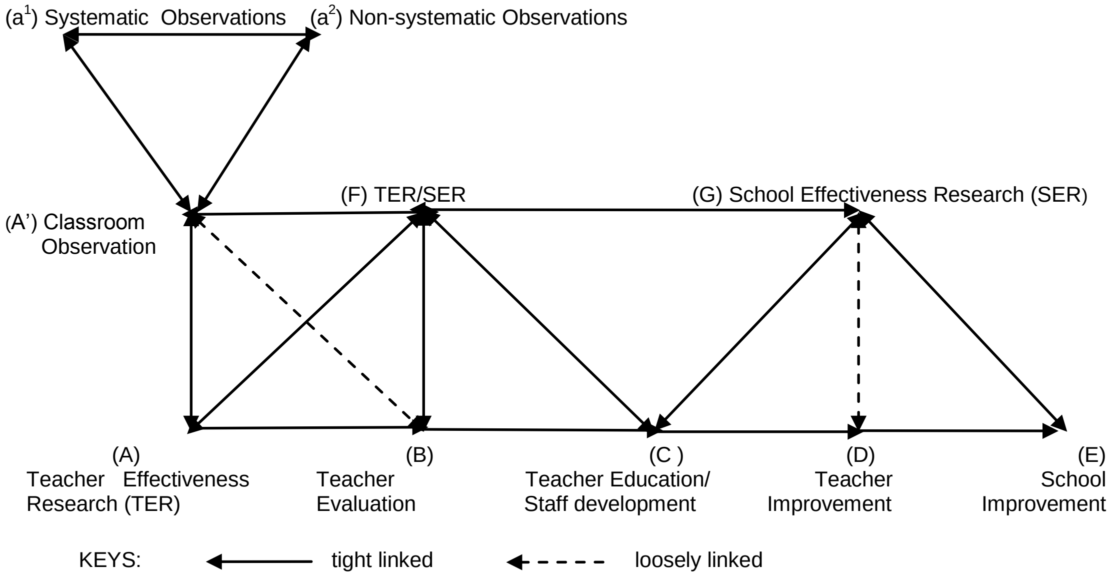


Note: Adapted from Teddlie et al. (2003, p.7).

Despite the desirability of these links, these authors recognise that there are missing links in different countries. For example, there was a lack of TER and an absence of teacher evaluation as an agenda item in educational policy and practice in the past in Hong Kong (Lee, Lam & Li, 2003). The relatively short history of TER in the UK has mitigated the subsequent linkages. A strong linkage between (A) and (B) is also noted as a consequence due to the influence of TER on the development of teacher evaluation instruments in Cyprus (Kyriakides & Campbell, 2003; Teddlie et al., 2003). Even in the US, TER and SER often represent two different paradigms in which researchers are not necessarily interested in the other's domain (Teddlie, 1991).

Accordingly, Teddlie et al. (2003) also acknowledge that establishing the missing links or strengthening the existing weak links may not be easy; for example, the general distrust to TER and severe criticisms of SER in the U.K. are considered as the major hindrances to develop successful and constructive links there (Teddlie et al., 2003). Since in both TER and teacher evaluation rely on classroom observation, the role of classroom observation and the employment of observation instruments become crucial to the development of TER, teacher evaluation and the subsequent domains. Accordingly, the conceptual graph can be revised as Figure 1.2 below to capture the role of classroom observation.

Figure 1.2: **Revised conceptual links among classroom observation and other domains of educational research**



Classroom observation can exist in two major types: systematic (classroom) observations (i.e., a^1) in which researchers may record classroom processes with some systematic instruments that are quantitative by nature and non-systematic qualitative observations (i.e., a^2) like ecological observations and ethnographic observations (Medley, 1982; see more examples in Erickson, 1985). Another major difference between the links in Figure 1.1 and Figure 1.2 is the possibility of *reciprocal* influences between some major domains. For example, it is not only the case that TER can inform teacher evaluation but also possible for teacher evaluation to inform TER. For example, in the U.K., the dissemination of school inspection reports provides the TER research a large repertoire of classroom observation evidence on teacher effectiveness obtained from an evaluative perspective (see Beard, 2000). Classroom observation brings teacher evaluation and TER together as both look for similar evidence in the classroom. Used as a lens to observe, a classroom observation instrument is selective in respect to what is selected as evidence and value-laden in respect to what evidence is regarded as important. Decisions made in the development of an instrument are much dependent on the philosophical and theoretical assumptions that are not explicit. As it is rare that classroom observation instruments are compared, the question of what is valued in a particular instrument is also often ignored. Accordingly, *employing classroom observation instruments* *developed independently in different traditions would raise interesting comparisons (see Section 3.5)*.

1.4 The aims and background of the study

There are two broad aims in the research project of this thesis:

1) to contribute to the TER knowledge base by describing, analyzing, and explaining the consistency and variation in the teaching behaviours identified in lesson observation;
2) to inform teacher evaluation and teacher development through characterising and explaining the teaching practices of four ESL teachers in an underperforming school in a socioeconomically disadvantaged area in Hong Kong.

These aims reflect an attempt to study a different cultural context with theoretical and methodological interests comparable to those found in a body of British SER and TER. Though the most immediate and comparable link of the present research can be found in Day et al. (2008), its background and context of inquiry was actually aligned with many earlier studies. The similar emphasis of differential school and teacher effectiveness was originated in Mortimore et al. (1988) and Day et al. (2006), which has examined the differential school impacts and the differential teacher effects on student learning and achievement.

Based on the findings by Muijs and Reynolds (2000) and Sammons and Ko (2008), it is assumed that teaching is a multi-dimensional process and its multidimensionality can be explored and confirmed by using observation schedules and factor analysis. In their longitudinal teacher effectiveness study, Muijs and Reynolds (2000) have provided not only justifications for employing classroom observation instruments as the major data collection method in SER/TER, but also confirmed the multidimensionality of effective teaching and the primacy effect of teacher behaviours on student achievement. Similarly, in a report of Day et al.'s (2008) study of effective teachers in England, Sammons and Ko (2008) confirmed the multidimensionality of effective teaching behaviours using two classroom observation schedules. In particular, it was shown that certain characteristics in the observed teacher behaviours of a purposive sample of teachers who were identified as typical of more effective classroom practice could be seen as generic characteristics of effective classroom practices. Based on these recent major contributions to TER, what needs to be further addressed is to explore the extent to which the level of teacher effectiveness, as defined in terms of their observed effectiveness, would match with the frequency and the strength of its positive impacts of these classroom practices on student achievement.

Although there has not been a consensus on how many dimensions of teaching behaviours may need and what they should be. Many classroom observation instruments have been developed to explore this (see Section 2.2.4 for detail), but large scale international research has showed that there were more variation in classroom or teacher level factors across different countries than the fine-grained behaviours (Creemers, Stringfield, & Guldemond, 2002). More importantly, there is a need to address the multidimensionality of teaching within a theoretical framework. In other words, findings need to be explained in terms of theoretical models which vary in their specific accounts for the nature of consistency and variation in classroom practices. In order to account for consistency and variation of observed teaching behaviours in the classroom, three theoretical frameworks are elected for comparison: the Dynamic Model of Educational Effectiveness (DEE) by Creemers and Kyriakides (2008), the Differentiated Model of Teacher Effectiveness (DTE) by Campbell et al. (2004), and the Probability Model of Educational Effectiveness by Marzano (2003). Details about these models and their specific predictions regarding consistency and variation in classroom practices are discussed in Section 2.5.

Findings on schools in challenging contexts in Hong Kong are limited. For example, only Cheng, Cheung and Tam (2002) report a comparison of an effective school and an ineffective school, both in low-socioeconomic contexts. In contrast, research on schools in challenging circumstances in the U.K. are extensive and may have implications to other cultural contexts (e.g., Harris & Chapman, 2005; Harris, James, Gunraj, Clarke, & Harris, 2006; Harris, Chapman, Muijs, Russ, & Stoll, 2006; Harris, Muijs, Chapman, Stoll, & Russ, 2003; Lupton, 2004; MacBeath, Gray, Cullen, Frost, Steward, & Swaffield, 2007). They have illuminated the depth of tensions and paradoxes in the U.K. Similarly, the ten benchmarks identified by Day (2004, 2005) to characterise the profile of headship in these schools were useful for understanding the role of leaders in their successes. By comparing findings in earlier research, the present study can then inform to what extent schools in similar socioeconomic background but in different cultural contexts. Specifically, it is imperative to address whether ineffective teaching is one of the main features of an underperforming school or whether such a school would undermine the teacher effectiveness of individual teachers equally. The distinction between the two lies in that ineffective teaching is attributed more to the teacher in the former, but more to the school contexts in the latter.

Certainly, to achieve the above aims in the Hong Kong context is a challenge, given a neglect of TER and teacher evaluation in educational policy and practice in Hong Kong in the past (Lee, Lam & Li, 2003). Although recently there have been more studies on the classroom factors, their focus and theoretical framework have been on promoting lesson study and teacher mentoring in schools (Lo, Chik, & Pang, , 2006; Lo, 2003; Lo, Pong, & Chik., 2005) or understanding teacher discourse (Marton & Tsui, 2004) in Marton's phenomenographic approach to the theory of learning (Marton, 1986a, 1986b; Marton & Booth, 1997). Thus, these studies on teacher education and teacher improvement (i.e., C and D in Figure 1.2) are not readily applicable to the present research and there is a strong need to fill the gap in the under-researched areas (i.e., A’, A, and F in Figure 1.2) in Hong Kong. Rather, the policy studies that have contributed to the understanding of education reforms in Hong Kong in the past two decades have provided an essential background for understanding features of the Hong Kong educational contexts relevant to the present research (see Section 2.2 for details).

1.5 Summary and the structure of the thesis

The above discussion presents a new perspective to view teacher effectiveness based on consistency and variation in observed teachers’ behaviours as measured with systematic classroom observation schedules. To further present the study addressing this perspective, the structure of the following thesis is summarised as follows:

Chapter 2: This chapter begins with a review of four system-wide challenges identified as challenges strongly affecting the teaching and learning in secondary schools. It proceeds with examining evidence on effective teachers, effective teaching and classroom practices, from describing and classifying fine-grained teacher behaviours, comparing with and relating to other non-behavioural and contextual factors in teaching, and then to incorporating different factors in testing theories of teaching. The focus of this discussion gradually shifts to multidimensionality of teaching and measuring consistency and variation in observed teaching behaviours.

Chapter 3: This chapter examines the epistemological and methodological issues surrounding the MM and case study approaches employed in this study. The research design and its rationale are reviewed, followed by a brief account of the implementation of the pilot study and the main study.

Chapter 4: This chapter reports the main findings using the ISTOF classroom observation instrument. Results of descriptive statistics and factor analysis are compared with those found in another English study.

Chapter 5: This chapter reports the main findings using the QoT classroom observation instrument. Results of descriptive statistics and factor analysis are compared with those found in another English study.

Chapter 6: This chapter reports the results of comparing the various factor analysis models, based on their relative predictability in relation to indicators of teaching effectiveness. Issues regarding relative strengths of instruments and cross-validation of samples are addressed.

Chapter 7: This chapter presents the mini-case studies of the four observed teachers, which are developed on the basis of their self-reports in the returned teacher survey, the post-observation interviews and the classroom processes recorded in the qualitative field notes, and the quantitative observation results. These results are triangulated.

Chapter 8: This chapter examines the challenges existed in the school and the department identified in the interviews with the school principal, the head of the department and the teachers. Issues concerning consistency and variation across teachers are reviewed in numerous quantitative cross-case comparisons.

Chapter 9: This chapter summarises the research questions, main findings and the lessons in this embedded case study. The significance, implications for future research and limitations of this study are addressed.

CHAPTER 2 : CHALLENGES IN THE EDUCATIONAL CONTEXTS OF HONG KONG AND ISSUES CONCERNING OBSERVING AND THEORISING CONSISTENCY AND VARIATION IN TEACHING PRACTICE

2.1 Introduction

The current thesis addresses issues relevant to a broad range of research related to classroom observation, teacher behaviours, effective teaching practices, teacher effectiveness, school effectiveness, school improvement and teacher evaluation. The literature presented here selected from this vast body of research has provided ample evidence on the topics of effective teachers, effective teaching, effective schools, and classroom observation for the development of different theoretical frameworks varied in terms of comprehensiveness and relative significance of variables.

Section 2.2 outlines the context of the present study. Four system-wide challenges that may affect the classroom and the school in the Hong Kong education system include school places allocation system, streaming and setting, medium of instruction policy (hereafter MOI), and examination-oriented culture. They are interrelated factors that affect the teaching and schooling in Hong Kong. They are challenges as they contribute to the negative impact of social class and inequality as determinants of students' educational and later occupational outcomes and thus the reproduction of the social order. The selection and sorting function of schools has been operating not only through, as in the UK, a 'high stake' public examination system which controls access to higher education, but also through the school places allocation system which controls the access of primary and secondary schools. While streaming is built into the school places allocation system, setting and mixed ability teaching reflect the strategies of individual schools. The MOI dilemma is a unique phenomenon due to Hong Kong's colonial history and current international outlook. Examination-oriented culture is a unique phenomenon of countries of Confucian culture. Examinations have a long history in China as the major apparatus of social selection, though which those who possess high calibre can get accesses to economic means and political power in the government. The account is not intended to be a comprehensive survey of all challenges, but to represent some distinctive features that are essential for understanding how contextual factors in Hong Kong have acted as goals, pressures and supports to teachers and schools.

Section 2.3 summarises four sets of evidence concerning effective teaching. It begins with evidence on the general characteristics of effective teachers or effective teaching identified in previous research. While the second set of evidence concerns the relative effectiveness of different teacher characteristics, the third set of research evidence compares teacher effects with school effects. Finally, after introducing what classroom observation is, evidence using classroom observation to characterise teacher effects in some major research in the U.K. is examined. Classroom observation is found particularly useful in establishing multidimensionality of teaching.

The purpose of Section 2.4 is to address the extent to which teaching practices are value-laden, reflecting the values of the agents or stakeholders, because the present research also examines the potential impacts of contextual variables on teaching practices. It is shown that contextual variables often appear to be paradoxical because they have opposite functions, acting as either constraints or facilitators of teaching practices and creating tensions for teachers. Evidence in one large-scale international cross-cultural research project using classroom observation schedules is discussed. A case study of Hong Kong in the same project showed the contrast between a more effective school and a less effective school. The portraits of these schools offer an alternative understanding on schools of low socioeconomic background in Hong Kong.

In Section 2.5, four models are selected as major attempts at theorising teacher effects in relation to other variables. These include Carroll's (1963) model of teaching and learning, Dunkin and Biddle's (1974) global model of teaching, Creemers' (1994) comprehensive model of educational effectiveness, and Cheung, Cheng and Pang's (2008) model of teacher success in the Hong Kong context. The main features and limitations of these models are examined.

Section 2.6 describes the three theoretical frameworks that would provide different predictions regarding consistency and variation of classroom practices. The dynamic model educational effectiveness by Creemers and Kyriakides (2008) provides the most up-to-date account of a generic concept of teacher effectiveness and its relation to student characteristics, classroom, school, and system level factors. In contrast, arguing that some of the differential teacher effectiveness should be understood as consequences of five different dimensions of differentiation, Campbell et al. (2004) put forward their differentiated model of teacher effectiveness with an emphasis on the differentiations found in teachers' instructional role and other non-instructional roles. Finally, Marzano's (2003) probability model of educational effectiveness is distinctive as it uses different scenarios to describe the predicted student outcomes as consequences of the teacher, the subject department and the school that may vary independently.

Implications of the literature reviewed are summarised and a set of research questions are proposed with the rationales discussed in Section 2.7.

2.2 Some system-wide challenges that may affect teaching and schooling in Hong Kong education system

2.2.1 School places allocation system in Hong Kong

At the system level, one of the main challenges to most schools lies in the impacts on school results due to the selective Secondary School Places Allocation (EDB, 2010b, hereafter SSPA) System. This is a centralised system to allocate secondary school places based on academic merits in a catchment area. That is a system in which students in the same district are streamed by abilities. This system also allows any students (including those who live in different districts) to compete for some discretionary places apart from the centrally allocated places. As in England, Hong Kong students generally go to schools in their catchment areas. However, different from their English counterparts, Hong Kong primary students are streamed into different schools ranked by their academic achievements.

Moreover, the catchment area in Hong Kong is a district that is geographically a much larger area than it normally would be in England. In any district or catchment area, there is a much larger student population and a larger number of secondary schools than is in England. According to students' order of preference and the availability of places in preferred schools, primary students of the same districts would go to one of the secondary schools they choose in their living districts. The number of secondary schools in a district varies district by district, but there are always a sufficient number of schools for streaming to make a significant difference. Those secondary schools with outstanding public examination results are generally "rewarded" by a steady intake of students who are academically more able. Similarly, certain prestigious primary schools are also "rewarded" by oversubscription if they have a higher proportion of students that can admit to secondary schools with outstanding academic results.

The meritocratic school places allocation system has huge impacts on creating differences in practices and results among schools. Compared to most developed countries, Hong Kong has an exceptional pattern of differences between schools in students' attainment levels. Lam, Wong and Ho (2002) conducted a multilevel analysis on the results of a cohort of 41,709 students from 322 schools who took the SSPA exercise (equivalent to the Key Stage 2 SAT test in England) in 1992 and the Hong Kong Certificate of Education Examination (equivalent to GCE O level in England) in 1997. They found the differences among schools in Hong Kong differed from those in the West (like those countries in Scheerens, Vermeulen, & Pelgrum, 1989) in three aspects[5].

First, there was a relatively large proportion of variance at the school level in Hong Kong secondary schools, ranging from a high value of 37% for English to a high value of 11% for mathematics. Second, there is a contextual effect because the effect of prior attainment was widened more by less effective schools in the achievement gap in the academic outcomes between the able and the less able students, suggesting that streaming may

5 It should be noted that the SSPA system which Lam et al. (2002) researched on was the old one that ranked schools in five bands. The system was revised in 2001 to make it more equitable.

results in greater differences in quality and effectiveness, making schools with more able students more effective and schools with fewer able students more ineffective. Third, there is a clear differential effectiveness for low and high ability students for most schools, more accentuated in ineffective schools than in effective schools.

SSPA not only creates ranks in schools, but may also results in unfair competitions and discriminations. Lam et al.'s (2002) study indicates that the contextual factors that influence the effectiveness of schools in a mixed system in the West are not duplicated in a streamed-school system like that in Hong Kong. Schools in Hong Kong take in students of similar rather than mixed abilities. Like many East Asian countries, Hong Kong students compete to enter the good primary schools. Primary school students compete to enter the good secondary schools, and then compete to enter universities[6].

One consequence of this highly competitive system of education is that only the best students can enter the highest attaining schools, the next best students to the next best schools, and the weaker students to the less preferred schools. Students are streamed to schools of different bands (i.e., ranks) based on their student's results in the public examinations. These schools with more able and motivated students who work harder to face the keen in-class competitions would have a better chance to sustain effectiveness, while those schools with more less able students studying in poorer learning atmospheres would struggle with a spiral downturn. This is a general picture of the traditional, examination-driven, selective Confucian education systems mixed with market-driven streaming policy.

The less able primary students who become academically weaker in lower attaining secondary schools are forced to compete with more able counterparts who perform better academically in higher attaining secondary schools when they take the public examinations years later. There seems to be evident among schools a systemic Matthew effect, under which the initial

6 The streaming system in Hong Kong is more like that in Singapore than Japan, where it was reported variation across classroom and across school was small in Japan (Husen & & Postlethwaite, 1994; Kaya & Rice, 2009), but large in Singapore (Tan, 1998; Kaya & Rice, 2009). Hong Kong has a similar highly competitive school system and merit-based streaming.

academic differences in pupils when they start school would increase with the progress of their school career (see Luyten, Cremers-van Wees, & Bosker, 2003; Stanovich, 1986; Walberg & Tsai, 1983). Such a competitive system may also have a negative effect on students' academic self-concept (Cheng, 2000; also see Marsh & Hau, 2003; Marsh, Kong, & Hau, 2000).

2.2.2 Different groupings by streaming, setting and mixed ability

The literature on mixed ability grouping in the form of streaming and setting[7] shows mixed evidence as both streaming/setting and mixed ability grouping had their own drawbacks (for a review, see Suknanadan & Lee, 1998; Gamoran, 2000; Harlen & Malcom, 1997). Criticisms against ability grouping are based on its detrimental effects on equal learning opportunities and increasing inequality in students' achievement levels over time (Gamoran, 2002). In contrast, as mixed-ability classes are generally hard to manage and teach, teachers may aim lessons at the middle of the ability range, sometimes treating mixed-ability groups as though they were low-ability streams (Gamoran, 2002). Although whole-class teaching methods are generally considered as inappropriate for mixed-ability groups, they are frequently used by teachers who are experienced in teaching with mixed-ability classes (Harlen & Malcom, 1997).

At the system level, the Hong Kong government has tried to reduce the adverse effects of streaming in the school reforms since 2001 (EDB, 2010b) by narrowing the gaps between schools through reducing the number of bands. Before 2001, schools were divided into five bands. The so-called Band One schools are invisible labels for those schools which admit a higher proportion of primary students whose Chinese and Mathematics are on the top 20% in their catchment areas[8]. To moderate the discrimination against schools of lower bands, five bands became three bands. This means that many schools are more likely to have students with mixed abilities than

---

7 Streaming refers to where students are divided by ability regardless of subjects, while setting is where students are divided on a subject-by-subject basis.

8 Within in each band, schools vary in the proportion of the top 20% of primary students in their intakes, from 50% to 100%. The actual academic standard of students in different districts also vary, so Band One schools in some districts, especially high SES areas, may have more able students than the whole territory average and some districts can have more Band One schools as these oversubscribed schools can recruit 10-20% of their intakes from other districts through direct admission. Some students may also transfer to schools of higher bands if they can demonstrate their academic results.

before because students of the same band are more likely to vary in abilities in wider bands[9].

At the school level, the best strategy for secondary schools to tackle streaming is to maximise their chances in getting better intakes, either through partnership with local outstanding primary schools or by getting better academic results in the public examinations. The former requires better marketisation of the schools and the latter requires strategies to enhance student outcomes. The two goals are not necessarily contradicting each other because schools that perform better in the public examinations generally have better publicity. However, enhancing student outcomes is particularly difficult for schools at a lower band, because they have to be highly effective in order to counter the large unfavourable school effect created by the SSPA system. Otherwise, a static or spiral downturn is more likely to be their destiny.

Most secondary schools now also have to face with the increasing number of students with mixed abilities. Depending on their districts and sponsoring bodies, schools also vary considerably for resources that they can get to deal with the increasingly diverse needs in the students. Some districts have a much higher socioeconomic background and some sponsoring bodies are financially much stronger than others. These are schools that can have more resources to deal with mixed abilities in students. While mixed abilities in students have often been regarded as a challenge to many teachers, many schools deal with this problem by setting their students in Chinese, English, and Mathematics in the schools. The best 20% of students go to Class A, the next 20% go to Class B, and so on. Most teachers, especially those who were trained to teach elite students before the school expansion since late seventies of the last century, would prefer in-school streaming to in-class differentiation. Differentiation in the classroom is seen as inefficient and incompatible with the standardised examination-oriented teaching and learning culture.

9 The policy to reduce the number of bands among schools is most unwelcome among the Band One and Band Two schools, as it is against their interests to have students of lower academic abilities. However, local Hong Kong academics like K.T. Hau is one of the main advocates for the policy as it is believed that this would minimise the negative Big-fish—small-pond effect upon students' self concept.

Although mixed research evidence suggests that teaching quality seems likely to have more impact on achievement than classroom organisation (Harlen & Malcom, 1997), some secondary schools prefer to compensate less able students with smaller classes. In these schools, students of similar but lower abilities are often grouped together as a group smaller than the normal class size. It is argued that teachers can give students more individual attention and individual involvement may be enhanced in a smaller class of students of similar abilities. Very often, the debate about class size and streaming students are associated with the enhanced opportunity to learn in class. To those who are rarely exposed to a classroom setting, the connection between the two may sound so natural that they hardly would question the effectiveness of these practices.

Regarding the generation and allocation of resources, high attaining and low attaining schools in Hong Kong may have different strategies. On the one hand, schools with better public examination results become more successful in recruiting more able primary students and, some of which would become fee-paying Direct Subsidy Scheme schools when necessary, allowing students from the new middle class to be over-represented in high attaining secondary schools. On the other hand, most low attaining secondary schools would prefer focusing their resources on the handful more able students through more in-school streaming, rather than rendering the less able students with compensating programmes. With more funding pouring into schools recently, some schools can experiment with co-teaching in some classes in which two teachers would teach a class at the same time[10]. These practices are rarely determined by individual teachers as they require allocation of staff and resources that is usually under the control of staff of senior levels like the department heads and the principal himself/herself. No matter what strategies are more preferable, resources are always limited and the amount of resources allocated in these strategies often reflects the priority of the school authority.

10 The compensatory funding system in Hong Kong is somewhat similar to that in the Netherlands (see Bosker & Guldemond, 2009 for details), but much smaller in scale. Thus, its effectiveness is unlikely to be the same.

2.2.3 Conflicting goals of Medium of Instruction Policy

Apart from reducing the number of bands, the EMB also introduced a controversial medium of instruction (MOI) policy for secondary schools to help the majority of primary students who previously learn in their first language (i.e., Cantonese, hereafter L1) in primary schools to cope with the difficulties to use English, a second or foreign language (hereafter L2). This is because previous research evidence showed deficiencies in learning content subjects using English as MOI (Johnson, Chan, Lee, & Ho, 1985; Brimer et al., 1985). Both classroom observation and research evidence showed that teachers and learners' relied on mixed-code (i.e., Cantonese admixed with English words and phrases) in classroom instruction and interaction, especially in teaching content subjects (Ip & Chan, 1985; Johnson, 1983; Johnson & Lee., 1987; Shek, Johnson, & Law, 1991). This policy was consistent with the school reforms to narrow the gap among schools, because there might be in effect only two major groups of schools that differed only in terms of their MOIs. However, according to Tsui, Shum, Wong, and Tse (1999), this policy, like other system reforms, reset the standards, endorsed new exemplars of 'good' practices and practitioners, redefined the rules of survivals, and thus changed the whole ecology of teaching and learning.

This MOI policy was announced in 1997 (Education Department, 1997a, 1997b), shortly after the change of sovereignty of Hong Kong. According to this policy, Chinese would be the default MOI for all secondary schools from September 1998 onwards. It was meant to be a mandatory policy, so those schools that wanted to use English as the MOI instead had to apply for approval and meet the criteria. In effect, out of a total of 421 government and government-subsidised secondary school schools, only 114 (27%) schools were approved to use English as the MOI (hereafter EMI schools) and the rest 307 schools had to use Chinese as the MOI (hereafter CMI schools) (Tsui et al., 1999). All EMI schools became Band 1 schools (Choi, 2003) because they can recruit more able students who can learn other subjects using English as MOI, though 95% of the primary students come from CMI primary schools. In the policy documents and elsewhere (see Education

Commission, 2005; EDB, 2010c), it was assumed that students of the CMI schools would not be discriminated in their entrance to higher education as they would show their competitiveness through learning subjects in L1.

The government declared that the new MOI policy was grounded on sound education aims as it was intended to ensure that the majority of primary students who come from CMI primary schools can continue to learn naturally and effectively in their mother tongue when they are in the secondary schools. Given the sensitive timing of the policy, many researchers argued that the policy was motivated by political reasons (e.g., to symbolise national unity and identity and to decolonise and resinicise Hong Kong) more than educational ones (Poon, 1999; Tsui, 2007; Tsui et al., 1999). This MOI policy was a symbolic act to stand out Hong Kong's "hybrid identity" in maintaining a delicate balance as a city built on Chinese culture but with an international outlook (Tsui, 2007, p.129). It was not just controversial, but also seen as unwelcome by parents and the business sector. This is because it would deter many students' chance to enhance their live chances as well as acquisition of the cultural and linguistic capital of the English Language (Evans, 2000, 2002; 2009; Ho & Ho, 2004; Morrison & Lui, 2000), but welcome by some CMI school principals for its Egalitarian ideals and pedagogical soundness (Choi, 2003).

Most parents, middle-class parents in particular, understand that the mastery of English still signifies power and status and will not change immediately after the decolonisation of Hong Kong (Pennycook, 1995). The MOI policy was strongly criticised by academics as it did not positively promote mother tongue teaching and learning (e.g., Choi, 2003; Evans, 2000, 2003; Poon, 1999; Tsui, 2007; Tsui et al., 1999). MOI was perceived as a "high-handed, inconsistent and socially divisive" policy (Evans, 2000, p.186), which actually allows an elite group of schools to stand out from the crowd (Choi, 2003), because it allowed only "one criterion for determining which people will complete different levels of education" (Tollefson, 1991, p.8). The MOI policy has exemplified Bourdieu and Passeron's (1990, p.73) view that students have to achieve a successful level of acculturation with respect to language.

After ten years of implementation, the government found that a mandatory MOI policy was undesirable in view of new evidence from research and academic results. New research evidence showed that didactic pedagogy and passive learning that characterized teaching and learning through a foreign medium might still prevail in many EMI schools even after they had the best students selected for them through the new social selection apparatus. In a 3-year longitudinal study commissioned by the Educational and Manpower Bureau (i.e., the previous educational authority between 1997 and 2007, hereafter EMB), Tsang (2004) found that the CMI secondary students enjoyed a significant competitive advantage over the EMI counterparts in sciences and social studies (see Yip, Tsang & Cheung, 2003). However, Tsang (2004) also found the EMI secondary students outcompeted their CMI counterparts in English, a result that seemed to confirm Krashen's maximum comprehensible input hypothesis (Krashen 1981, 1982; Krashen & Terrell, 1983), according to which students who receive more and diversified English inputs in the curricula in EMI schools are more likely to excel in English than their CMI counterparts. The competitive advantage of the CMI schools in content subjects soon disappeared in a few years when the secondary graduates of the CMI schools competed head on with their counterparts of the EMI schools. These results also showed that the streaming effect of the MOI policy would be detrimental, rather than beneficial, to the bilingual education, a fear that Poon (1999, p.142) expressed in her appeal for an alternative model that was based on "streaming by subject", rather than "streaming by class" or by school.

In a follow-up study, Tsang (2009) found that the EMI students not only eventually caught up with the CMI students in science and social studies but also enjoyed a significant advantage over them in English that made them dominate the entry to the tertiary education. As Choi (2003) points out, language policy in education requires scrutiny because the school system mediates social stratification based on language use, which is often used as the "one criterion for determining which people will complete different levels of education" (Tollefson, 1991, p.8). Thus, recently, the government announced the new MOI policy did not work as they wished and allowed schools to make some adjustments that might suit the individual school situation[11] (EDB, 2010b) However, it took the government to come to this action only after they collected the unsatisfactory public examination results of six cohorts of the CMI school graduates after the implementation of the new MOI policy. If this policy was meant to diminish an elitism centred on English, it *failed* because it has *reinforced* it instead[12]. If this policy was meant to proscribe mixed-code teaching, it partially failed because students and parents opted for their preference in paid private tutorial classes where students could learn any subject including English in mixed coded MOI.

2.2.4 Interdependence of examination-oriented education and private tuition

CMI schools are unlikely to disappear in the near and distant future, even though the MOI policy is not as strict as before. Many schools have found it they need more resources to make the adjustment. Instead, schools tend to do further streaming in schools. Therefore, discriminations against the CMI schools and those students who cannot learn effectively in EMI remain unchanged and competitions among them become more intense because parents and primary students would keep an eye on those schools which have EMI classes and more potential to get the EMI status. To tackle the problem that students in EMI or CMI schools may not understand their subjects, parents and students heavily relied on private tuition in one to one basis to one to hundreds basis.

While the Confucian system seems to be working well in many East Asian countries and places like China, Japan, Korea, Taiwan, and Singapore, Hong Kong students seem to suffer from a self-esteem lower than places and countries whose students perform less well in international comparative studies (Cheng, 2000). Because of the ‘high stakes’ nature of Hong Kong public examinations as determinants of young people’s future educational and employment life chances, there are strong arguments for emphasising

11 According to this fine-tuning adjustment to the MOI policy, CMI schools are allowed to set aside a quarter of their lesson for “extended activities conducted in English”. This has provoked some strong resistance from principals, teachers, and academies who are staunch supporters of mother-tongue teaching as this is believed to result in increased workloads and intense competition among schools to offered classes.

12 Certainly, I believed that this consequence was unintended because I personally knew one of the officials in the highest rank in promoting the policy. Unfortunately, he died prematurely before I could conduct an interview with him.

academic goals. This is also the same in England (Sammons, 1999). Many educators may declare that students' academic achievement, particularly results of public examinations, should not be the only important goal of education, but in reality it remains to be the major indicator that the public can understand and use it to evaluate a school. Thus, teaching practices that are considered to enhance attainment results are most preferred. Streaming, extensive and frequent assessments like quizzes, tests and in-school exams, drills of past examination papers, after-school classes in schools, and above all, the private tutorial classes are popular as they are considered effective means to boost results. Although the teaching practices and its mixed code teaching in private tutorial classes would hardly be endorsed as best practices by the educational authority, but it was so widely practised in the region that it was considered a “Shadow Education System” (Bray, 1999), because it “provides supplementary instruction to students enrolled in the public school system” (Dang &Rogers, 2008, p.161).

As one of the most overlooked areas in the TER and SER literature, this so called shadow system may reinforce the washback effect of examination on education and may distort the true effects of schools if no attempts are made to separate and take into account of their impacts. Dang and Rogers (2008) argued that tutoring can raise the effectiveness of the education system. Time spent on private tuition can be considered as a measure of the opportunity to learn factor associated with student achievement in Carroll's (1963) model or Creemers' (1994) model. The measure of this opportunity factor was found to be closely related with student achievement in studies conducted in various countries (e.g., Brookhart, 1997; Trautwein, Koller, Schmitz, & Baumert, 2002; cf. Antoniou, 2009). Private tuition is often regarded as a common phenomenon in Confucian countries with an examination-oriented culture, such as China (including Hong Kong and Taiwan), Japan, Korea and Singapore[13].

Most often private tuition is discussed from an economic production or

13 However, in fact it is a widespread practice of many other countries with diverse economic and geographical variations like Bangladesh, Canada, Cyprus, and Greece in Bray (2006), India (Agarwal, 2006), Tukey (Tansel & Bircan, 2007), Germany (Otto, 2008), Cambodia, the Arab Republic of Egypt, Kenya, Morocco, Romania, the U.S.A., and the U.K. (Dang & Rogers, 2008). According to Otto (2008), private tutoring is also common in Germany but its purpose is more remedial and less examination-driven.

equity perspective (e.g., Bray, 1999, 2003; 2006; Bray & Kwok, 2003; Dang & Rogers, 2008). However, it can also be an important confounding variable in educational effectiveness studies in these countries if it is not accounted for (e.g., Brookhart, 1997; Trautwein et al., 2002), but assumed to be just normal differences in student backgrounds or differences between students' socioeconomic groups. The equity concern here is that students whose are willing to pay more may enjoy more advantages and high SES families generally can afford to pay more for their children (Bray, 1999; Bray & Kwok, 2003). In Hong Kong, private tuition is not only for remedial purpose as in many other countries, but also for preparing students to get entry to higher education. As it is strongly believed by parents that access to higher education would improve one's life chances, students from different backgrounds and SES groups all tried to get access to private tuition and thus changed the learning time spent and the opportunity to learn through a shadow system[14]. Thus, it is ironic that the demand of this shadow system originates from examination as a powerful instrument for national governments. "This instrument will probably become even more powerful when the national goals are established and examinations are adapted in line with these goals" (Creemers, 1994, p.22).

2.3 Evidence on effective teachers, effective teaching, and classroom practices

2.3.1 General profiles of effective teachers and effective teaching

Clearly, there would be no teaching without teachers, but effective teaching or teacher effectiveness can be defined in a much broader sense than simply teacher behaviours, or what they are observed doing in the classroom. Besides pedagogical processes, teacher effectiveness may also include managerial and organisational aspects of teaching (Harris, 1998). Over the years a large number of reviews have already synthesised robust research findings on effective teacher behaviours, for example, Bloom (1976), Brophy & Good (1986), Gage (1978), Glass (1977), Good, Biddle & Brophy (1983), Light and Smith (1971), Rosenshine (1971), Walberg (1986) and

14 Tansel and Bircan (2007) also notice that the demand for private tutoring in Tukey has increased dramatically since competitive university entry examination was introduced.

Wittrock (1986). These reviews have suggested that despite the diversity of approach, there has been some consensus in TER about what an effective teacher would look alike. For example, based on their studies and others, Porter and Brophy (1988, p.74) describe effective teachers as "semi-autonomous professionals" who:

- are clear about instructional goals;
- are knowledgeable about curriculum content and the strategies for teaching it;
- communicate to their students what is expected of them – and why;
- make expert use of existing instructional materials in order to devote more time to practices that enrich and clarify the content;
- are knowledgeable about their students, adapting instruction to their needs and anticipating misconceptions in their existing knowledge;
- teach students meta-cognitive strategies and give them opportunities to master them;
- address higher-as well as lower level cognitive objectives;
- monitor students' understanding by offering regular appropriate feedback;
- integrate their instruction with that in other subjects areas
- accept responsibility for student outcomes;
- are thoughtful and reflective about their practice.

Similarly, Mortimore et al. (1988, pp.227-231) identified a set of the effective teacher characteristics in their study of effective primary schools in England:

- teacher was responsible for ordering activities during the day for pupils, i.e. structured teaching;
- spent greater amount of time communicating with pupils about the content of their work, but not routine matters
- kept a lower level of noise and movement in pupils;
- maintained high levels of interaction with the whole class;
- kept a fairly narrow focus within individual sessions;
- spent more time on asking questions, particularly high-order questions, providing ample, challenging work;
- let pupils have some responsibility for their work and independence within these sessions;
- maintained high levels of pupil involvement in tasks appropriate for their levels of ability;
- kept a positive atmosphere in the classroom;
- had high levels of praise and encouragement

Doyle (1987, p.95) also found secondary pupil achievement can be enhanced when the teacher:

- emphasises academic goals;
- makes [goals] explicit and expect pupils to be able to master the curriculum;
- carefully organises and sequences the curriculum ;
- clearly explains and illustrate what pupils are to learn;
- frequently asks direct and specific questions to monitor pupils' progress and checks their understanding;
- provides pupils with ample opportunities to practise;
- gives prompts and feedback to ensure success;
- corrects mistakes and allow pupils to use a skill until it is over-learned and automatic;
- reviews work regularly and holds pupils accountable for their work.

The fine-grained behaviours of effective teachers in most reviews of teacher profiles are likely to be universal, as they appear to be evident in different countries (see Creemers et al., 2002). Yet, since the present research was conducted in Hong Kong, a Chinese culture setting, it is reasonable to think that there may be some conceptualisations of effective teachers and teaching more compatible with the traditional Chinese views.

Findings in Pratt et al.'s (1999) study[15] on effective teaching in higher education seem to be applicable to secondary teachers. Table 2.1 below summarises three themes in their study stressed by both Hong Kong Chinese students and faculty.

Table 2.1: **Three themes in conceptualising effective teachers and effective teaching by Hong Kong Chinese students and faculty in high education**

Appropriate roles and relationships for teachers and students of effective teachers
- Have a close, protective relationship with student like a coach or even a parent
- Strict image with high expectations and care
- Understand their difficulties and to guide them in their learning and personal development

Attributions of responsibility for effective teaching
- It is more common for learners to accept the responsibility not only for their learning, but for their teacher' s effectiveness as a teacher

The process of teaching of effective teachers
- Take students systematically through a clear set of tasks, high in structure and directed toward examination
- Provide feedback that is specific and critical to point out weaknesses or errors in the students' thinking to ensure they accurately understand what they are studying.
- Adjust the pace and sequencing to the group's level of understanding
- Slow down to provide further explanation and closer guidance as to where there may be misunderstandings or gaps in their knowledge
- Be well-prepared and organized and be able to manipulate the structure of the content when students do not understand something
- Appear to be formal and distant in class, but be more informal outside classroom
- Be concerned about more than students' academic success; think of each other as members of an extended family

15 Pratt et al.'s (1999) findings seemed to be supported in the HK higher education context, as seen in (Bailey, 2005).

Comparing to the expatriate faculty, the Chinese participants' view of the roles and social relationships of effective teacher as a figure of authority, morality and benevolence, conformed to the Confucian concept of 'ren' (Jin & Cortazzi, 1998) and the social hierarchy of teachers in Chinese society (Pratt et al., 1999). Interestingly, learners tend to think that they are also responsible for their teachers' effectiveness, because teacher effectiveness is seen as a shared responsibility between teacher and learners (Pratt et al., 1999) as *all* students, not just *good* students, are supposed to obey and pay attention to what the teacher says in China (Jin & Cortazzi, 1998).

Pratt et al. (1999) also noted that the process of effective teaching in Hong Kong has to address the learning process of Chinese learners and the instrumental values of education in the Chinese culture context. For example, effective teachers are expected to respond to the learning process of Chinese learners (Marton, Watkins, & Tang, 1997; Watkins & Biggs, 1996) in their needs of lots of structural tasks, drills and memorisation of materials until they can master the basics and develop critical thinking. Though students expect teachers' critical feedbacks on their errors more often than their praises, effective teachers, in return, are expected to pay due regards to the instrumental goal of education through their preparation and organisation of the teaching content aiming at success in examination (Pratt et al., 1999). Thus, it is argued that the Egalitarian relationship between teachers and students in the western societies and the student-led constructivist approach of learning are not necessarily highly appreciated by the Chinese participants (Pratt et al., 1999).

Recently, from their interviews with four primary and eleven secondary teachers who were winners of educational awards in the years 1998-2000, Cheung et al. (2008) have identified a set of professional qualities (in Table 2.2) contributed to the success of these teachers. While most of the professional qualities listed in Table 2.2 (*shown in italics*) seem to be observable in the teacher's classroom practices and similar to those suggested by western researchers, three qualities (shown in bold) are more similar to those mentioned by Pratt el at. (1999). The emphasis on the teacher's ability to help students to prepare for examinations and obtain good results and the expectation to be students' role model are respectively reinforced by the examination-oriented education and the traditional role of teacher in the social hierarchy in the Confucian culture. It is an example to illustrate that the goals of education affect the criteria of effectiveness (Creemers, 2001). Cheung et al. (2008) also particularly stressed the successes of the teachers were results of the interaction of personal qualities (such as respectfulness, facing adversities with courage, and not giving up easily, and attaching importance to moral education/having a positive influence on students' values and attitudes) and professional qualities with the contexts (see also Section 2.5.4).

Table 2.2: **Professional qualities of award-winning teachers in Hong Kong**

	Professional qualities	Interaction with context
Skills/abilities	*Possessing generic skills (e.g. communication skills, critical thinking skills)* *Clear and in-depth delivery of lessons* *Ability to enhance students' understanding* *Ability to arouse students' learning interests* *Basing teaching on students' abilities* *Teaching students to analyse and view things objectively* *Effectively managing the classroom* *Having good relations with students* **Helping students to obtain good academic results/high passing rates** **Teaching students the skills to prepare for examinations** **Being a role model for students**	School context variables (Principal's support, colleagues' collaboration and encouragement, and student's positive feedback)
	Ability to handle duties other than teaching	
	Grasping opportunities and making good use of resources	School context --resources
	Understanding in the needs of colleagues	School context -- colleagues
	Having good communication with parents	Context beyond school -- parents
Attitudes	Never ceasing to improve ways of teaching and classroom management Lifelong learning Teaching students both subject knowledge and attitudes Willing to face new challenges (e.g., in teaching, education reform)	
Knowledge	Reaching adequate standards in the teaching subject	

Note: Adapted from Cheung et al. (2008, p.627).

In contrast to the western account of effective teacher, the Chinese model of successful teacher by Cheung et al. (2008) seems to link the teacher's classroom behaviours more closely with student outcomes. However, this link may require further quantitative evidence and this gap is best filled by meta-analysis. Synthesising over 800 meta-analyses relating to the influences on achievement in school-aged students, Hattie (2009) recently has identified over 31 teacher and teaching factors with an effect size over 0.40, indicating their moderate to strong impacts on pupil progress (see Table 2.3).

Table 2.3 **: Mean effect-sizes from over 800 meta-analyses of various influences to achievement**

Teacher/Teaching factors	Effect size	Domain
Provide formative evaluation	.90	Teaching
Micro teaching	.88	Teacher
Comprehensive interventions for learning disability students	.77	Teaching
Teacher clarity	.75	Teaching
Reciprocal teaching	.74	Teaching
Feedback	.73	Teaching
Teacher-student relationships	.72	Teacher
Spaced vs mass practice	.71	Teaching
Meta-cognitive strategies	.69	Teaching
Self-verbalisation/self-questioning	.64	Teaching
Professional development	.62	Teacher
Problem-solving teaching	.61	Teaching
Not labelling students	.61	Teaching
Teaching strategies	.60	Teaching
Cooperative vs, individualistic learning	.59	Teaching
Study skills	.59	Teaching
Direct instruction	.59	Teaching
Mastery learning	.59	Teaching
Worked examples	.57	Teaching
Concept mapping	.57	Teaching
Goals	.56	Teaching
Peer tutoring	.54	Teaching
Cooperative vs, competitive learning	.54	Teaching
Keller's PIS	.53	Teaching
Interactive video methods	.52	Teaching
Questioning	.46	Teaching
Quality of teaching	.44	Teaching
Expectations	.43	Teaching
Behavioural organisers/adjunct questions	.41	Teaching
Matching style of learning	.41	Teaching
Cooperative learning	.41	Teaching

Note: Adapted from Hattie (2009, pp.297-298).

Hattie's findings are impressive, but it is still not clear how these teacher/teaching factors are related. For example, it is uncertain whether teachers are effective because they can manage their class well or maintain a supportive classroom climate, though conventional wisdom may tell us that a good teacher has to do both. Attempts to separate effects of instruction, management, and classroom curricular design were unsuccessful (Levy, Wubbels, Brekelmans, & Morganfield, 1997). Thus, *there is a need to establish empirically what dimensions are crucial for effective teaching and to show to what extent teacher behaviours are stable or consistent in these dimensions across lessons.*

2.3.2 Characterisation and categorization of effective teaching practices

Going beyond profiling effective teachers, some researchers have attempted to categorise different teaching behaviours and characterising the links between these categories and student achievement. Therefore, in addition to the extensive research on general teaching behaviour, much has been written about specific *effective teaching skills* (e.g., Clark & Peterson, 1986; Kyriacou, 2007; Muijs & Reynolds, 2005; Philpott, 2009; and Wragg, 1984), *different teaching styles* (e.g., Bennett, 1976; Galton & Croll, 1980; Opdenakke & Van Damme, 2006), and *different models of teaching,* which specify particular types of learning environment and approaches to teaching (e.g., informational processing models, behavioural systems family models, personal family models; see Joyce, Weil, & Calhoun, 2005 and Joyce, Calhoun, & Hopkins, 2008). On the one hand, these studies have showed that variations in teaching behaviours contribute much to teachers' effectiveness in the classroom. On the other hand, numerous research studies also reflect a high degree of consensus concerning the *generic features of effective teaching* (e.g., Bennett, 1988; Bickel & Bickel, 1986; Good & Brophy, 1999; Harris, 1998; Mortimore et al., 1988; Rosenshine, 1983; Walberg, 1986, 1990; Wang & Walberg, 1991).

However, these characterisations or classifications suggested how teaching behaviours is grouped may be subject to philosophical orientations. The most notable example is the debate on the relative effectiveness of the teacher-directed (or explicit) instruction and student-centred constructivist approaches to literacy teaching. Rowe (2006) argues that as the philosophy of constructivism has been prevailing in the content of teacher education courses, school systems in many western countries have been dominated by the various constructivist approaches to teaching under the names of whole language teaching, anchored instruction, situated learning, discovery learning, task-based learning and scaffolding, problem-based learning, and issue-based learning. Supporting evidence for the stronger effects of teacher-directed approaches on student learning (i.e., direct instruction) was also found in numerous research in the U.K. by Galton et al., (1980), Mortimore et al. (1988), and Muijs and Reynolds (2000) (for details see Section 2.3.4).

Based on the interview sample of 18 teachers of physics in Guangdong, China, Gao and Watkins (2001) developed an instrument to measure teachers' conceptions of teaching. The instrument consisted of six dimensions (Learning and learner, Nature of teaching, Role of teacher, Expected outcomes, Teaching content, and Methods of teaching) for five lower order conceptions of teaching five lower order conceptions (Knowledge Delivery, Exam Preparation, Ability Development, Attitude Promotion, and Conduct Guidance) as shown in Table 2.4.

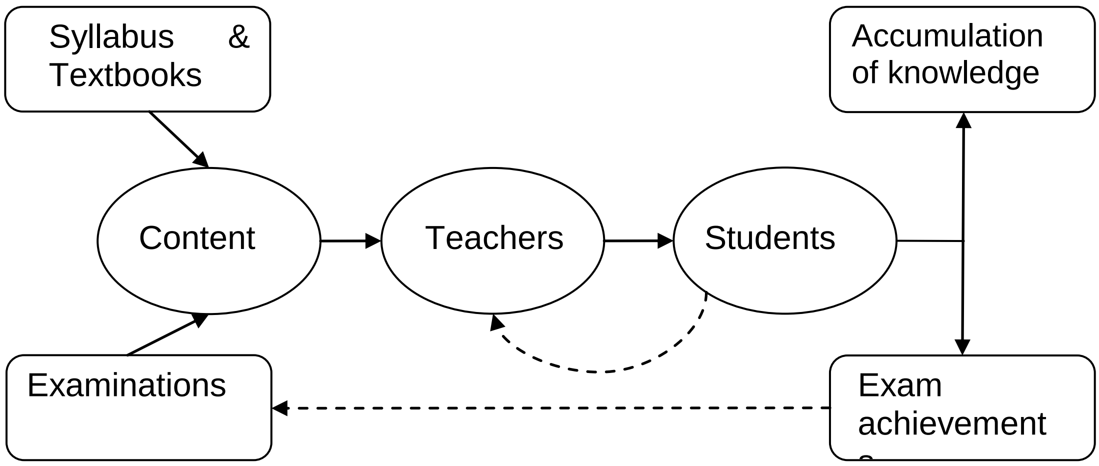
Table 2.4: **An overview of the conceptions of teaching identified in the qualitative data analysis of interviews with teachers in Gao and Watkins (2001)**

	Learning & Learner	Nature of teaching	Role of teacher	Expected outcomes	Teaching content	Methods of teaching
Knowledge Delivery	Acquiring knowledge and skills; Passive receivers	Delivering knowledge and skills	Deliverer and resource	Accumulation of knowledge and skills	Follows the textbook closely	One-way lecturing plus demonstration
Exam Preparation	Achieving exam requirements, Achievers, Competitive	Preparing for examinations; Drilling students	Trainer and director	High exam achievement	Conducted by the `baton of exams'	Classroom drilling, Effective for preparing exams
Ability Development	Internal construction; Explorers, Capable, flexible and creative	Facilitating learning	Guide, leader, and facilitator	Developing understanding and ability, knowing how to learn	Meets the needs of students and matches students' level	A variety of methods, emphasises activities & interactions
Attitude Promotion	Establishing good attitude	Promoting and fostering good attitude	Model of good learner with good attitude	Active and independent in learning	Contained implicitly in teachers' performance	Interactive and interesting; indirect manner
Conduct Guidance	Self-improvement	Facilitating and guiding good conduct	Role model of good conduct, friend of students	Qualified persons with good conduct	Related materials, contained implicitly in teachers' behaviours	Friendly and interactive, indirect manner

Note: Adapted from Gao and Watkins (2001, p.451).

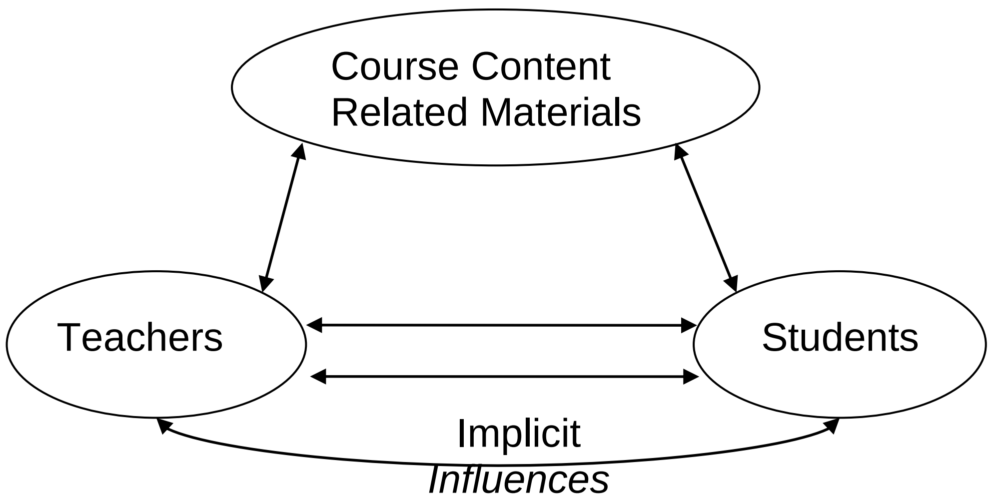
Based on this table, Gao and Watkins (2001) developed 37-item an instrument called School Physics Teachers' Conceptions of Teaching to measure teacher's conceptions of teaching and proposed two higher order dimensions concerning orientation of teaching (Moulding and Cultivating) in three theoretical models, as depicted in Figure 2.1, 2.2 and 2.3. Testing the instrument with another sample of 450 physics teachers and confirmatory factor analysis, the researchers found the instrument showed satisfactory internal consistency and fitted the expected underlying factor models.

Figure 2.1: **The proposed model of the Moulding orientation of teaching (the dotted lines indicate weak relationships)**



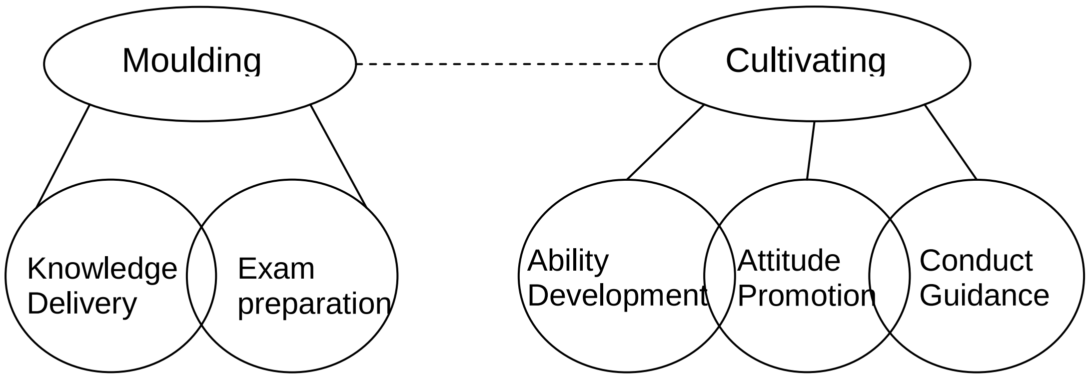
Note: Adapted from Gao & Watkins (2001, p.455).

Figure 2.2: **The proposed model of the Cultivating orientation of teaching**



Note: Adapted from Gao & Watkins (2001, p.456).

Figure 2.3: **The proposed general framework of conceptions of teaching (the dotted line indicates a weak relationship)**



Note: Adapted from Gao & Watkins (2001, p.456).

The classification and categorisation shown in Table 2.4 seem to be consistent with the results of Pratt et al. (1998) and Cheung et al. (2008) in general. The emphasis on examination in teaching content seems to suggest an unequivocal widely cultural phenomenon, spreading in all school levels and geographical contexts. The cultivation of attitude and conduct seems to subject to the teacher's personal qualities and his/her ability to be the role model, and this may a blend of personal and professional qualities going beyond behaviours that have been generally categorised as classroom management and classroom climate in the western literature.

2.3.3 Non-behavioural aspects of teacher and teaching effectiveness

There has been also a revival interest in *non-behavioural* characteristics of teachers. These would include their personality, attitude, experience, aptitude/achievement and knowledge. Earlier research on teacher effectiveness in the US in the sixties focused mainly on linking the personality characteristics in teachers and student achievement, but this type of research generally yielded insignificant and inconsistent findings (Martin, Baldwin, & Yin, 1995; Borich, 1996; cf. Costin & Grush, 1973; Levine, 1991). These research findings were often criticised because personality attributes failed to be good predictors of effective teacher behaviour for being too far remote from the actual classroom processes (Getzels & Jackson, 1963).

However, in two influential papers, Shulman (1986, 1987) criticises TER for unduly ignoring the knowledge base of teachers because it has been predominantly interested in effective teacher behaviours. Rather, he argues that teachers should possess at minimum an effective knowledge base including different types of knowledge, among which the pedagogical content knowledge or the special amalgam of content and pedagogy is hypothesised to be most important. Such an emphasis on pedagogical content knowledge has generated another body of research investigating differential teacher effectiveness beyond the classroom level (e.g., Askew, Rhodes, Brown, William, & Johnson, 1997; Thompson, 1992), on *subject knowledge* (e.g., Askew et al., 1997; Aubrey, 1997; Monk, 1994), and on *teachers' self-efficacy beliefs* (e.g., Anderson, Greene, & Loewen, 1988; Ashton, 1985; Ashton & Webb, 1986; Chan, Chan, Cheung, Ngan, & Yeung, 1992; Magno & Sembrano, 2007; Moore & Esselman, 1992; Muijs & Reynolds, 2000; Philippou & Christou, 1999; Relich, 1996; Schunk & Rice, 1993).

While findings about the actual impacts of teacher knowledge are not conclusive and some concepts are still under-conceptualised (e.g., teacher's philosophical beliefs, see Campbell et al., 2004), there has been a tendency in the government to institutionalise that knowledge base in teachers through policies to raise teacher quality by teacher education, teacher evaluation, and professional certification or qualifications (see Goldhaber, 2007; Goldhaber & Anthony, 2007; Mandeville & Liu, 1997). However, raising teacher quality through qualifications or licensing examinations may be justified on cost, but not supported by its effectiveness as it might disturb teachers' focus on teaching and it was not always the most effective teachers retained (Goldhalber, 2007; Libman, 2009).

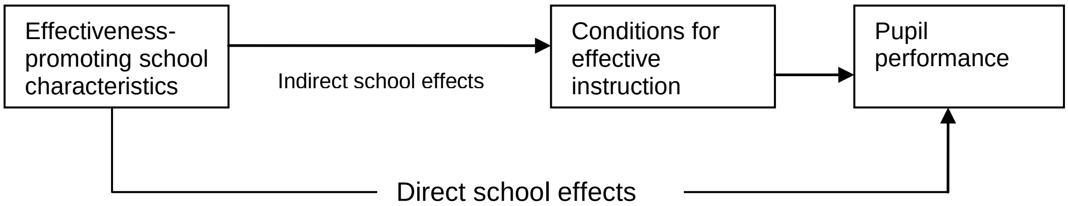
There have been few attempts to link professional knowledge and teacher behaviours in the classroom in an empirical model (e.g., Magno & Sembrano, 2007; Mcber, 2000; Muijs & Reynolds, 2002). Empirical evidence seemed to support that teacher knowledge, as a teacher characteristic, has a smaller impact on student achievements than teacher behaviours, as it is more distal and its effect is mediated through teacher behaviours (Magno & Sembrano, 2007; Muijs & Reynolds, 2002). Muijs and Reynolds (2005) warn of the dangers of overemphasising a knowledge base for teachers and prescriptive policies by the policymakers.

Regarding other aspects of teacher characteristics, in a recent mixed methods study by Halvorsen, Lee, and Andrade (2009) found that teachers whose attitudes showed higher levels of *teacher responsibility* and schools with higher collective teacher responsibility had significantly more positive impact on their students' reading ability and achievement. Individually these teachers tended to attend more conferences, spend more time of professional development and preparation. Collectively these teachers were more likely to feel that they have a high impact on policy and control over curriculum, report principal communicates vision and show supportive and encouraging behaviours. However, they were less likely to teach in schools in urban areas with higher than average of minority (76%) students, nor they have extensive *experience*. Day et al. (2006, 2007) have shown that differential teacher effectiveness is *neither static nor progressive* with teachers' professional life cycles, but a consequence of continuous dynamic interactions between teacher characteristics and contextual variables. Their research suggests that any attempt to explain differential teacher effectiveness without reference to the impacts of personal and contextual factors that affect teaching would be misleading and oversimplifying the reality.

2.3.4 Teacher and school effects and their relative significance

The contributions of SER to studying teacher effects are somewhat incidental as the focus has been on estimating the magnitude of school effects. As teachers work in schools, schools directly affect teacher effectiveness through different effectiveness-enhancing conditions, but can also have direct impacts on pupil performance, as shown in Figure 2.4.

Figure 2.4: **Step by step causal process with school and instruction conditions as malleable factors**



Note: Adapted from Scheerens and Bosker (1997, p.147).

Table 2.5 below summarises the school-level factors identified in the post-1990 reviews by Cotton (2002), Levine and Lezotte (1990), Marzano (2000), Sammons, Hillman and Mortimore(1995), and Scheerens and Bosker (1997).

Table 2.5: **Comparing school-level factors across reviews since 1990**

School-level factors	Levine & Lezotte (1990)	Cotton (2002)	Sammons, Hillman & Mortimore (1995)	Scheerens & Bosker (1997)/ Marzano (2000)	Creemers & Kyriakides (2008)
Guaranteed and viable curriculum	Focus on central learning skills	Planning and learning goals	Concentration on teaching and learning	Content coverage/ opportunity to learn	School policy on teaching and actions to improve teaching
		Time use	Maximisation of learning time	Time	
Challenging goals and effective feedback	High operationalised expectations and requirements	High expectations	High expectations	Monitoring	Evaluation of the effect of school policy on teaching and actions to improve teaching
	Appropriate monitoring	Monitoring student progress and alternative assessment	Monitoring progress	Pressure to achieve	Evaluation of learning environment
Parental and community involvement	Salient parental involvement	Parent and community involvement	Home-school partnership	Parental involvement	Partnership policy
Safe and orderly learning environment	Productive school climate and culture	Discipline and physical environment	A learning environment	School climate	Student behaviour outside the classroom
		Incentives and rewards	Positive reinforcement		Values in favour of learning
		Equity and special programs	Pupil rights and expectations		Provisions of sufficient learning resources
Collegiality and professionalism	Outstanding leadership	School-based management and instructional leadership	Professional leadership	Leadership	Leadership
		Restructuring, workplace readiness	Shared vision and goals		Provision of sufficient teaching resources
	Practice-oriented staff development	Instructional improvement, pprofessional development/ collegial learning	A learning organisation	Cooperation	Collaboration and interaction between teachers

Note: Modified after Marzano (2003, p.19) with comparison the Dynamic model of education effectiveness by Creemers and Kyriakides (2008)

Although many researchers found that the profiles of effective schools also showed characteristics of effective classroom processes contributed to fostering pupils' learning and progress, much of the research evidence to date on educational effectiveness suggests that while schools can make a difference to student achievement, a substantial portion of that difference may be attributed to teachers (e.g., Creemers, 1994; Creemers & Kyriakides, 2008; Hill & Rowe, 1996, 1998; Konstantopoulos, 2007; Kyriakides & Creemers, 2008a, 2008b; Nye, Konstantopoulos & Hedges, 2004; Muñoz & Chang, 2007; Sammons et al.,1997; Sanders, 1998, 2000; Scheerens et al., 1989; Tymms, 1993).

Using data collected in Victoria, Australia, Hill and Rowe (1996) showed that class/teacher effects ranged from 38 to 45% for literacy and 53 to 55% for numeracy, whereas the school effects shrank to 4-9%. In the U.S.A., teacher effects on adjusted student achievement were found additive, strongly cumulative, but little compensatory in the multilevel analysis performed on longitudinal data in the database of Tennessee Value-Added Assessment System (Sanders & Rivers, 1996) and in the database of the Dallas Public Schools (Jordan, Mendro & Weerasinghe, 1997). Thus, Sanders (1998, p.27) concludes that[16]:

> Of all the contextual variables that have been studied to date (indicators of school socioeconomic status, class size, student variability within classrooms, etc.), the *single largest* factor affecting academic growth of populations of students is *differences in the effectiveness of individual classroom teachers*. When considered simultaneously, the magnitude of these differences dwarf the other factors..... Also, the effects of teachers appeared to be *cumulative*. At the extreme, a high-high-high sequence [of 3-year teacher effects of 5th grade pupils] resulted in more than a 50 percentile point higher score in 5th-grade math achievement than the low-low-low sequence.
>
> In addition, students of the most effective teachers had excellent gains regardless of their prior achievement levels, while students in the least effective teachers' classrooms across the entire prior achievement spectrum did not make appropriate levels of gain. As the level of teacher effectiveness increased, students of lower achievement were the first to benefit, and only teachers of the highest effectiveness generally were effective with all students.

Accordingly, in a review of the educational effectiveness evidence in the U.S., Darling-Hammond (2000) concludes that, as the major determinant of differences in student learning, differential teacher effectiveness outweighs the effects of differences in class size and class heterogeneity.

16 It should be noted that similar cumulative school effect was also found in English primary schools (Mortimore, 1998).

The relative stronger impacts of teacher and classroom factors on student achievements are evident in meta-analyses. For example, Scheerens (1992) summarised results of some 40 school effectiveness studies conducted before 1989 and found that instructional conditions such as structure teaching and aspects of classroom management such as effective learning time have received clear empirical support, while factors predominately defined as school level factors such as recruiting staff, organisational preconditions, or school climate showed weak empirical support. Hattie (2009) shows that the top thirty most influential variables out of a rank of 138 variables affecting student achievement, nineteen variables related to teacher or teaching with an effect size above 0.5. Thus, most of the research evidence not only confirms the primary role of teachers in student progress, but also suggests the relative effectiveness of variables associated with teachers. However, given the abundance of evidence and a taxonomy of variables with differential effectiveness as Hattie's, there seems to be a lack of theories of teaching that can explain how and why different variables may differ in their effects on student progress (cf. see Section 2.5 and Section 2.6).

2.3.5 Multidimensionality of teaching practices in classroom observations

Classroom observation as a method and limitations of instruments

Classroom observation is often regarded as a naturalistic method to observe those classroom practices of teachers that are hypothesized to be effective or have positive impacts on certain student outcomes. Systematic observation has played a significant role in educational research, especially in the U.S., as it is direct, naturalistic and quantifiable. Medley (1982, p.1842) defines the term "systematic observation" as follows:

> The term "systematic observation" is used here to refer to observations of classroom behaviour made by a trained observer who records the behaviours according to an observation system. An "observation system," in turn, is a scheme that specifies both the events that the observer is to record and the procedure to be used in recording them.
>
> ….a quantitative method of measuring classroom behaviors from direct observations that specifies both the events or behaviors that are to be observed and how they are to be recorded

Sophistication of observation instruments might correspond with the increasing popularity of systematic observation in the sixties as a major research tool. The number of observation instruments developed over the years has increased enormously; for example, as cited in the historical account of classroom observation by Meehan, Cowley, Fich, Chadwick, Ermolov and Riffle (2004), the anthology of 92 classroom observation instruments with detailed accounts edited by Simon and Boyer (1967-1970) runs up to fourteen volumes. Another single volume sourcebook, Borich and Madden (1977), described only ten instruments that specifically aim at obtaining information about the teacher from an observer.

Despite recommendations in the introductory texts by Croll (1986), Wragg (1999), and Muijs and Reynolds (2005), classroom observation systems have not been popularised as they are in the American research tradition. The UK collection of classroom observation systems by dated work of Galton (1978, 1979) consisted of forty-one instruments. However, the few English studies[17] employed systematic observation systems tended to be large scale and longitudinal (e.g., the Observation Research and Classroom Learning Evaluation (ORACLE) Project by Galton et al., 1980; the Inner London Education Authority (ILEA) Project by Mortimore et al., 1988; the Primary, Assessment, Curriculum and Experience (PACE) project by (Pollard, Broadfoot, Croll, Osborn, & Abbott, 1994; the Gatsby Project by Muijs & Reynolds, 2000; the Effective Provision of Pre-School Education (EPPE) Project by Sylva, Melhuish, Sammons, Siraj-Blatchford, & Taggart, 2004; and the ECP Project by Day et al., 2008; some of them are reviewed in the next section). These studies might adopt classroom observation schedules developed by US researchers (e.g., the Flander system used in ORACLE the US). Instead, many of the British studies (e.g., those in Hargreaves & Woods, 1984) adopted a more descriptive approach in their accounts of the more flexible type of classroom and teaching methods.

In contrast to evaluative instruments adopting the inspection models developed by inspectors for teacher evaluation, the specification and

17 There are also a few studies done in Scotland; see McPake, Harlen, Powney and Davidson (1999) for a review.

categorisation of some pre-determined and agreed behaviour and practice in systematic observations originated and developed mainly by researchers are often intended to allow for comparisons, rather than for evaluations, of teachers (Rosenshine, 1970; Muijs & Reynolds 2005). Accordingly, value-oriented instruments in an inspection model generally require a value judgment from the observer to form global judgments on whether the observed behaviour was ‘excellent’ or ‘unsatisfactory’, (e.g., the performance indicators used in Hong Kong), while behavioural instruments are designed to look at the occurrence of specific behaviours without making evaluations on them (Muijs and Reynolds, 2005). This contrast is related to the distinction between high and low inference measures (Rosenshine, 1970). Rosenshine and Furst (1973) also state that a crucial distinctive feature of observation instruments is that its scale is generally for recording the frequency of the target behaviour or event.

According to Medley (1982, p.1845), as systematic observations offer process-product researchers a relatively affordable means for obtaining objective, accurate, and quantifiable records of the specific behaviours of teachers and students in classrooms, “systematic observations are still important tools for research in teacher effectiveness”. However, though systematic observation instruments are useful for exploring underlying dimensions in teacher behaviours, they are often context-specific and limited by a set of predefined categorical descriptive statements about teacher behaviours to be observed and some coding procedures that would not always inform much on teacher-student interactions, grouping procedures and lesson content and objectives. Accordingly, complementary qualitative field notes are often employed, as in the present research, to provide extra information on these areas that and to illuminate other aspects of teaching practices and classroom contexts such as their intentions and reflections on their teaching practices.

Some earlier empirical research based on classroom observation evidence in the U.K.

The significant contribution of classroom observation schedule to TER can be illustrated in the work of three groups of research in the U.K. First, there is little argument that the ORACLE studies by Galton and his colleagues (Galton, Simon, & Croll, 1980; Galton & Simon, 1980; Galton, 1995; Galton, Hargreaves, Comber, Wall & Pell, 1999) are classic classroom observation investigations on classroom practices and pupil outcomes. Their major findings on teacher styles indicated that teachers identified as 'Class Enquirers' for their excessive use of teacher-led direct instruction generated the most gains for pupils in the areas of mathematics and language, while pupils in classes of teachers identified as 'Individual Monitors' for their a highly individualised student-centred approach made least progress. While teachers of the Class Enquirers group were observed to spend significantly more time in whole class interactive teaching than those of the Individual Monitors group, a similar association was also found between progress and non-individualised interaction in further analyses (Croll, 1996; Galton et al., 1999).

While there were criticisms about the classification of teacher styles, the finding on the positive association between whole class interaction and high levels of pupil time on task was consistent with similar evidence found in on direction instruction[18] in the US (Slavin, 2006) and in Australia (Rowe, 2006, 2007). The association between high levels of whole class interaction and greater pupil task engagement was also confirmed in the later school matters research (Mortimore et al., 1988) and the later PACE study of primary schools, which investigated the impact of major educational reforms such as the introduction of curriculum and national curriculum changes (Pollard et al, 1994).

Multidimensionality of teaching behaviours and other classroom variables on pupil progress

Previous research by Levy et al. (1997) showed that different aspects of teaching tended to be highly interrelated and would not appear in isolation in the teaching of effective teachers. Similarly, using Mathematics

18 According to Hattie's (2009) synthesised result of meta-analyses, direction instruction on average has a moderately strong effect size of 0.59. However, its impact was found stronger in special education (d=0.83 in White, 1988) and reading (d=0.75 in Adams & & Engelmann, 1996), but much less weaker in comprehensive schools (d=0.21 in Borman, Hewes, Overman, & Brown, 2003).

Enhancement Classroom Observation Record (MECORS)[19] (Schaffer, Muijs, Kitson, & Reynolds, 1998 as cited in Muijs & Reynolds, 2000), Muijs and Reynolds (2000) established a seven-factor model of effective teaching behaviours. In a structural educational model, they have further shown that *whole-class teaching creates the conditions for effective teaching to occur rather than directly affecting pupil progress*. Since teaching behaviours such as asking open questions, allowing multiple answers and other explicit cognitive structuring are found among effective teaching behaviours, varied teaching strategies and constructivist teaching methods are also included as key dimensions of effective teaching. Accordingly, Muijs and Reynolds (2000, p.299) stressed that "it would be wrong to describe [the] whole-class interactive teaching style as a 'chalk and talk' drill-and-practice approach".

Using a multilevel model, Muijs and Reynolds (2000) also found that once pupil characteristics have been controlled for the pupil background effect, the aggregate effect of effective teaching behaviours explained between 61.5% and 100% of the remaining between classroom variance in test gains in written and mental math tests in their sample of Year 1 and Year 5 primary pupils. Holding all other variables constant, they estimated the difference between effective teaching behaviours by the teacher scoring highest on the effective teaching scale as opposed to the teacher scoring lowest can contribute to a difference in a pupil's scores on the test by between 10% and 25%. In contrast, other classroom or teaching variables like time on task, whole class interactive and constructivist teaching methods were found significant only in specific analyses, in either Year 1 or Year 5 level and in either written or mental test. This study by Muijs and Reynolds contributed to the understanding of the relative relationship of effective teaching behaviours and other variables like whole class teaching, constructivist teaching methods, and time on task, as well as their relative impacts on student progress.

19 This is a modified version of the classroom observation schedule *Special Strategies Observation System* (SSOS) by Schaffer, Nesselrodt, and Stringfield (1991, as cited in Meehan et al., 2004)

Multidimensionality of teaching behaviours of effective teachers as measured by different instruments

Reynolds (2006) declared one lesson learned in the International School Effectiveness Research Project (ISERP) was that there are few agreed international constructs concerning effectiveness. Different instruments may measure different constructs and instruments can vary significantly in their external validity, that is, their applicability in different contexts. Accordingly, there is a pressing need to develop a classroom observation instrument that would measure some agreed teacher effectiveness constructs. One of most neglected areas in classroom observation research is using multiple instruments to examine the multidimensionality of teaching practices. To date, only few attempts in the literature have used different instruments simultaneously and there is little recent research on this topic (e.g., Emmer & Peck, 1973; Ober, Wood, & Cunningham, 1970; Wood, Brown, Ober, & Soar, 1969).

Regarding developing an external valid classroom observation instrument, van de Grift and his colleagues (van de Grift, Matthews, Tabak, & de Rijcke, 2004; van de Grift, 2007) have attempted to establish reliability and validity a value-oriented instrument in an inspection model that requires a value judgment from the observer. They found their instrument showed strong external validity in its applications in four European countries, namely, England, Flanders (Belgium), Lower Saxony (Germany) and the Netherlands. However, their instrument differs in its approach from the behavioural instruments used by academics, as in Muijs and Reynolds (2000) for example, who measured the occurrence of specific behaviours without passing global judgments on whether the observed behaviour was 'good' or 'poor'.

Seeing an important contrast between the *evaluative* and *behavioural* instruments in the evaluative categories of practices based on the experiences of the inspectors and the pre-determined and agreed categories of teaching behaviours and practices originated in TER, the ECP research by Day et al. (2008) explored the underlying dimensions of the observed teaching behaviours of a purposive sample of typical and more effective teachers in England using one instrument of each type. In a report in that study, Sammons and Ko (2008) identified two sets of underlying factors that might define effective classroom practices, one for each of the two instruments employed. As shown in Table 2.6*,* these underlying dimensions share similar foci on climate, management, objectives/purposes, and support/feedback:

Table 2.6: **Underlying dimensions found in the ratings using the two instruments**

Evaluative Instrument	Behavioural Instrument
• *Supportive lesson **climate***	• *Clear and **coherent lesson** in a supportive learning **climate***
• *Proactive lesson **management***	• *Engaging students with assignments and activities*
• *Well **organized lesson** with **clear objectives***	• *Positive classroom **management***
• *Environmental and **teacher support***	• ***Purposive learning*** • *Quality questioning and **feedback** for students*

Sammons and Ko (2008) also found that a fifth of the sample teachers was rated relatively highly in one instrument but was also rated relatively highly in the other. High scores of their purposive sample of primary and secondary teachers were found in a number of the underlying factors and on particular items. These findings lent *support the generic concept of teacher effectiveness*, which holds that effective teachers would excel in the generic characteristics of effective classroom practices. However, the distribution patterns of the factor scores of the various confirmatory factor analysis models of the two instruments showed there was *variation across teachers in the sample for most factors*, although this was greater in some areas measured than in others. This also provided some *support for a differentiated concept of teacher effectiveness* in revealing that variation may exist in teachers' teaching behaviours when their students, working environments, subject taught are different.

Although Sammons and Ko (2008), like Muijs and Reynolds (2000), provided results that established multidimensionality of teaching behaviours, they could not estimate measurement consistency in the ECP study because the two instruments were used in two occasions. Variation in their results between the two instruments might be a result of instability of behaviours and contexts over time rather than differences between the instruments. *Further research is required to compare the instruments in more detail and in the same lessons and this is a focus for the present research.*

2.4 The significance of contexts, values, and their impacts on teaching effectiveness

2.4.1 Effectiveness promoting characteristics in department and school

As mentioned earlier in the introduction, schools seem to vary in teaching effectiveness for different subjects. The subject inconsistency in school seems to be larger in secondary schools than primary schools. Citing Luyten (1994), Scheerens and Bosker (1997) showed that departmental effects of subjects may account for 40% of the school effect, while consistency across subjects and stability across years amount only to 25%, suggesting the department effect might be stronger than the net school effect. The potential causes for this difference are attributed to curriculum and quality of teachers (Scheerens & Bosker, 1997; Sammons, 1999), because primary teachers taught both subjects, while their secondary counterparts were usually specialised in either subject.

Alternatively, Sammons et al. (1997) suggested that the roles of departments in contributing to differential effectiveness within a school, as they noted that some departments were more effective than others in some schools. They identified eight factors to explain (in)effectiveness of schools and departments in the English secondary schools :

- the importance of school and departmental histories and the impact of change;
- high expectations;
- academic emphasis – including examination entry policy and monitoring;
- shared vision/goals;
- an effective School Management Team;
- the quality of teaching (consistent for all ability groups);

Based on the multilevel results of the Scottish system, Fitz-Gibbon (1991) obtained similar findings as Sammons et al. (1997) and concluded earlier:

- it is *departments*, not schools, which vary most in any one year;

- almost all schools contain *both* effective and ineffective departments
- from year to year, the departments often *change* in effectiveness.

Accordingly, Fitz-Gibbon (1996, p.32) called for a closer examination of departmental effects:

> These findings strongly imply that quality, a good education for all students, will be best attained by close monitoring of departments. We need to learn from the most consistently effective departments and take action in the case of consistently underperforming departments.

Despite their findings, Sammons et al. (1997, p.145) argued that the school effect should not be neglected because:

> in some schools, because it was apparently 'easier' for all departments to function effectively....it was 'harder' for departments to be effective due to lack of overall leadership, shared goals and vision, poor expectations and inconsistent approaches.

Nevertheless, the above research evidence suggests a need to address teacher effectiveness within a departmental context and consistency in the department's provisions to different year levels.

2.4.2 The interplay of value and effectiveness

Earlier discussion in Section 2.3.1 suggests that the criteria of effectiveness are subject to the goals and objectives of education. Yet, arguments on goals and objectives are often based on value judgments which may vary over time. More than half a century ago, Rabinowitz and Travers (1953) rejected the view that we can statistically arrive at a set of characteristics that distinguish effective and ineffective teachers by empirically observing many teachers and by extending observations over long periods of time without explicitly or implicitly making a value judgment:

> it must be recognized that the *ultimate conception of the effective teacher is neither an empirical nor a statistical matter*. There is no way to discover the characteristics which distinguish effective and ineffective teachers *unless one has made or is prepared to make a value judgment*. The effective teacher does not exist pure and serene, available for scientific scrutiny, but is instead a fiction in the minds of men. No teacher is more effective than another except as someone so decides and designates. Teachers are real enough, and methods are available

> or can be improvised to study these real teachers. But *the effective teacher is only an abstraction*. *The process of designating any particular teaching practice as effective or ineffective inevitably stems from a reasoned judgment*. The *ultimate definition of the effective teacher does not involve discovery but decree*.

They further argued that because this value judgment concerns what are the worthwhile consequences of effective teaching, an ultimate criterion of teacher effectiveness has to be established on the basis of some goals of education. Rabinowitz and Travers' stance highlights a paradoxical relationship between value and teaching practices.

At the classroom level, the moral implications of teaching lie generally in the imbalance of institutionalised power between the teacher and the students (Sober, 1991), but particularly in teachers' competence because they possess the knowledge of pedagogy: "teachers and schools must be able to demonstrate that learning and development take place because of their expertise" (Thompson, 1995, p.32). Thus, Campbell et al. (2004a, 2004b) argue that a model of teacher effectiveness cannot be value-free as effectiveness in education carries value assumptions. Fok (2004) also argued that there were conflicting values in Hong Kong's recent school reforms that emphasized competitiveness and finance on the one hand and democratization, diversity and equity on the other hand. These conflicting values may add pressure to existing teacher-centered and exanimation-oriented teaching practices (Fok, 2004).

2.4.3 The paradoxical impacts of contextual variables on teaching practices

Teaching quality is affected by many conditions of teaching such as curriculum materials and syllabus, coherence of curriculum across year levels and subject areas, or class sizes, which are out of the control of teachers and depend on the administrative and policy systems where they work (Darling-Hammond, 2007). Findings in Pratt et al. (1999), Gao and Watkins (2001), and Cheung et al. (2008) have showed the impacts of contexts and culture on the criteria of effective teachers and teaching. Classroom is seen as an eco-system where all the components have a mutual effect on each other (Biggs, 1998). Thus, the system created between a teacher and students in one class may not be the same in another class. While a class is nested in a school which is also nested in an education system, “a very complex, multi-layered equilibrium is set up, with the culture over riding” (Gao & Watkins, 2001, p.447). This creates the characteristic `pedagogical flow’ of the classroom of a country’s schools, evident in the cross-cultural differences in approaches to teaching in the preliminary report of the IEA Third Mathematics and Science Study (TIMSS) (Schmidt, McKnight, & Raizen, 1996). Regarding this flow of teaching practices, Gao and Watkins suggested (2001, p.447):

> The nature of that `flow’ springs from socialisation practices, cultural values about education, and so on. Teaching practices thus have a contextual validity derived from the culture’s eco-system.

From a school improvement perspective, Sun, Creemers and de Jong (2006) categorise contextual variables into three categories by their three functions, as goals, pressure or support. Similarly, factors affecting teacher practices and teacher effectiveness may also exist in similar functions. For example, in the context of the public exam, an exam syllabus exists as a goal, the publicity of exam results as pressure and collaboration among teachers on teaching as supports. In their comparative analysis on the policy and practice of curriculum change in primary schools in England and in Finland, Webb and Vulliamy (1999) found that the external coercion pressurised teachers to adopt practices they did not support at a cost to their self-identity and motivation. Similarly, teachers in Hong Kong were required to teach students in English in the English-medium schools though these students were handicapped in science learning by their low levels of English proficiency, as their learning of English in the primary years was not sufficient to prepare them for a full English immersion program in secondary school (Yip, Tsang, & Cheung, 2003).

These examples showed that institutionalised practices at the system level are powerful as they may affect all teachers in the system. Most of these variables are apparatuses under the control of the government, existing in the form of a common curriculum or assessment system, educational policies, regulations and legislations, and supervisory bodies like inspection systems. These apparatuses perform their unique legislative, regulative, normative, or supportive functions. For example, while school inspection works by its regulative and normative functions, policies exercise their power through their legitimate functions. Education reforms are also multi-functional as they reset the standards, endorse new exemplars of 'good' practices and practitioners, redefine the rules of survival and may change the wider contexts of teaching and learning. They often present a paradigm shift of teaching practices (Fok, 2004). The exact role and the extent of the impacts of these apparatuses may vary in different places or countries. Since there lacks a benchmark or standard set for primary schools in Hong Kong, like National Literacy and Numeracy Strategies in the U.K., and a public channel for disseminating inspection reports, it is not clear to what extent Hong Kong teachers can draw on inspection data for improving their practices. Thus, the impact of inspection evidence in Hong Kong seems to be confined to the quality assurance framework and school accountability framework.

Best practices supported by inspection evidence are likely to be empirically-based, but it is also evaluation-oriented. Thus, one cannot take it for granted by assuming that it is unbiased or value-free. For instance, although evidence in favour of an exclusive use of English in the English as a foreign language (hereafter EFL) classroom is neither conclusive nor necessarily pedagogically sound (Auerbach, 1993; Cook, 2001, 2008; (Macaro, 1997, 2001a, 2001b), alternative EFL pedagogical methods using mixed code have been marginalised in Hong Kong in the policy discourse of medium of instruction policy. Instead, mixed-code was considered as a 'bad' practice (Education Commission, 1990). Bunton and Tsui (2002) also argued that a language benchmark test developed on a native speaker model for EFL teachers in Hong Kong is unjustified and discouraging to the teachers. It seems that the new linguistic imperialism in Hong Kong emphasised not only the necessity of an exclusive use of the second language (hereafter L2) in the EFL classroom, but also the superiority of the English-speaking teachers.

Similarly, though individual-focused instructional practices may be seen as desirable by some, they may not be sustainable if there are serious behaviour problems in class (Tam, 2009) and if the class size may not allow effective control of pupil behaviour (Galton & Hargreaves, 1996). It is not certain whether such constructivist teaching methods are compatible with the examination-oriented senior secondary curriculum and the negative washback effect of public examinations on classroom practice and the potential differences in the student outcomes (see Cheng, 1998, 2005 for the washback effect of examinations in Hong Kong).

2.4.4 Large scale research on teaching effectiveness in different cultures and in challenging contexts

To date, the most extensive results on differential teacher and school effects in different countries were from the report of the International School Effectiveness Research Project (ISERP) (Reynolds, Creemers, Stringfield, Teddlie, & Schaffer, 2002). Teacher effectiveness in that project was measured using the Virgilio Teacher Behaviour Inventory (VTBI) (Teddlie, Virgilio, & Oescher, 1990; Virgilio, Teddlie, & Oescher, 1991) and QAIT[20] (i.e., an updated version of Special strategies Observation Systems (SSOS), (Schaffer, Nesselrodt, & Stringfield, 1994; for details see Meehan et al., 2004). It was found that classroom management, classroom climate and teaching/instruction were the three factors that had statistically significant positive impacts on student academic outcomes in the U.S.A., the U.K., and Norway. However, only the climate factor was significantly associated with gain in Hong Kong schools, but it also appeared to have both positive and negative impacts in Irish schools.

Another exception was found in the negative impact of classroom management on student achievement in Dutch schools. Although some factors seemed to be less universal across countries, researchers found that “it is the fine-grained behaviours that are the same in different countries” (Creemers, Stringfield & Guldemond, 2004, p.49). These findings suggested that underlying dimensions of teacher behaviours tended to be less likely to be universal, but Teddlie, Creemers, Kyriakides, Muijs and Yu, (2006) later stressed that underlying dimensions were less universal across countries

20 The acronym stands for the different subscales, Quality of instruction, Appropriate level of interaction, Incentive and Time.

might be because the VTBI was not developed as an internationally valid instrument for assessing teacher effectiveness. Accordingly, in the new wave of ISERP, a new instrument, the International System for teacher Observation and Feedback (ISTOF) (i.e., one of the two instruments employed in the current study; see Section 3.5.3), has been developed for a better understanding of the universality of teacher effectiveness (Teddlie et al., 2006).

Another important finding in ISERP was that different teacher behaviours seemed to be associated with differential teacher effectiveness for specific student groups. For example, behavioural incentive systems, clear presentation, positive academic feedback, detailed directions and explanations and others seemed to help students in less effective schools more than students in schools of middle or high effectiveness categories. These findings are consistent with the findings in challenging contexts in the U.K. (Harris & Chapman, 2005; Harris, Gunraj, Janes, Clarke & Harris, 2005; Harris, Chapman, Muijs, Russ & Stoll, 2006; Lupton, 2004; Muijs, Harris, Chapman, Stoll & Russ, 2004). It is argued that specific teaching methods may be required to enhance teacher effectiveness at the start of a school's improvement initiative (Hopkins, 2001). For example, low-SES students are likely to benefit from more structured instruction, more positive reinforcement, and a curriculum tailored in smaller packages with subsequent rapid feedback (Brophy, 1992). However, there is also some evidence to suggest that students of low-SES backgrounds are also capable of high order thinking and need a curriculum that is as rich as that of their advantaged counterparts (Leithwood & Steinbach, 2002), but a review by Rowe (2006) indicated that direct instruction helped low SES learn better.

In the case study of Hong Kong schools in ISERP (Cheng et al., 2002), less effective schools of low-SES backgrounds generally showed more unfavourable characteristics than more effective schools of the similar SES backgrounds in many areas, such as student performance, students' participation in extra-curricular activities classroom climate, teacher satisfaction and attitudes, staff relationship, principal performance, parental influence, school organizational characteristics, and perceived school environment. Nevertheless, the curriculum content and teaching methods were found similar or the same. In particular, in the teachers' reports in more effective schools, less than 25% of class time was spent on handling student behaviour, comparing with 25 % to 50% more in less effective schools. In the less effective schools, teachers tended to resort to punishments, students perceived them as their superiors exercising the coercive power, rather than facilitators of their learning. These results indicated that challenging contexts should not be the major factor explaining the differential school effects or differential teacher effects.

2.5 Theorising models of classroom teaching in TER and SER

In Section 2.4.3, it has been argued that there seems to be a poverty of theories of teaching that can explain how and why different teacher variables may differ in their effects on student progress. Several models have been proposed to explain variation in the influences of different variables in teaching and learning. Four of these models are described below as background for introducing the three theoretical models to which the current research has specifically addressed.

2.5.1 Carroll's (1963) model of teaching and learning and its variants

Working from an educational psychology perspective, Carroll (1963) defined school learning as a function of time spent and time needed for learning in the classroom (i.e., school learning = f(time spent/time needed)). To measure time spent, Carroll proposed to look at two time related measures: first, the *opportunity to learn* in the classroom, that is, *allocated time* or the amount of time that the teacher is engaging her students on learning, and second, *perseverance* or *engagement rate*, that is, the percentage of the allocated time that students are actually on task. When defining time needed, Carroll (1963) proposed that it is a function of aptitude, ability to understand instruction, and quality of instruction.[21] Interestingly, Carroll (1963, 1989) defined aptitude as the ability that determines how fast a student to learn something in certain conditions and the ability to understand

21 The equation of school learning can be written: school learning = f ((time spent)/(time needed)) → = f((opportunity X perseverance)/(aptitude X ability to understand X quality of instruction)).

instruction as the preparedness of a student for understanding the specific material to be learned or the prerequisite knowledge. Thus, there are three time-related variables: opportunity to learn, perseverance, and aptitude and two achievement-related variables: prerequisite knowledge and quality of instruction. Of these variables, opportunity to learn and quality of instruction are expected to be under the teacher's control, while perseverance is expected to be subject to both the teacher's and the student's behaviours (e.g., the teacher's classroom managing skills). Carroll tested his model in Carroll and Spearritt (1967).

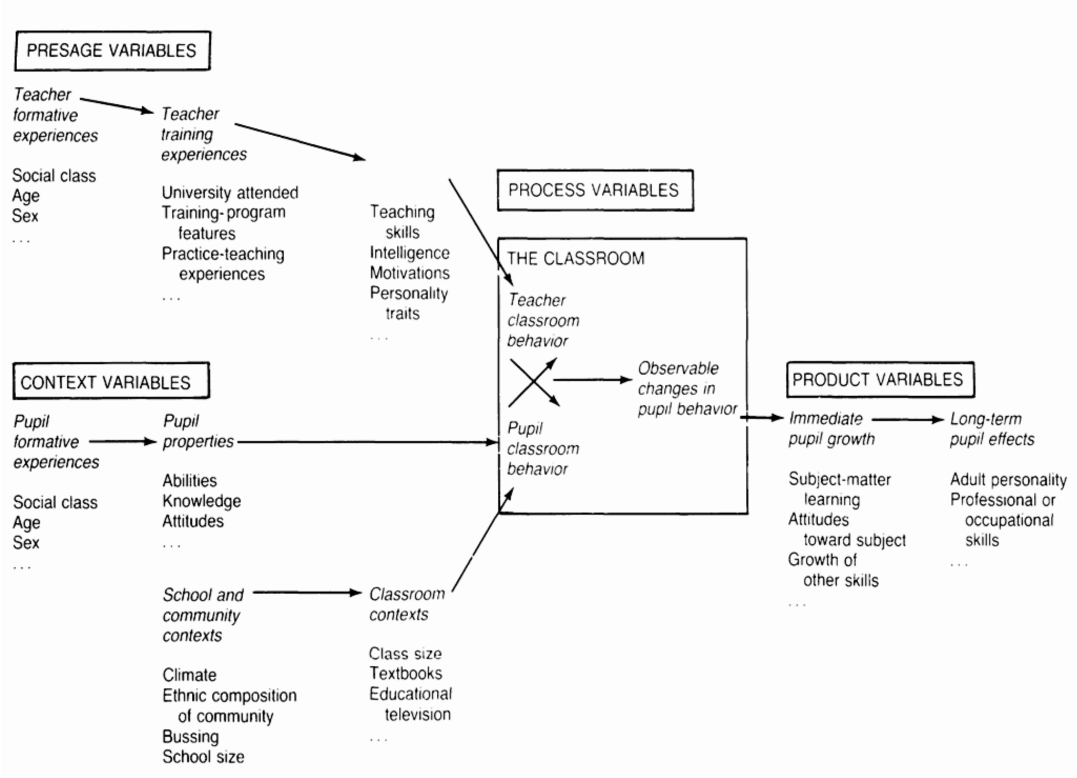
Carroll's classroom learning model has been very influential in TER, because, as a basic input-process-product model of teaching, it takes into account of the input of the students, the interactions between the teacher and the student, the relative roles of time and instruction in affecting learning.[22] Over the years, various extensions of Carroll's model have been attempted: notably, the inclusion of context variables such as home, classroom, peers and television in Walberg (1982; for others, see Carroll, 1989 and McIlrath & Huitt, 1995 for details) and the hierarchical elementary education effects model by Stringfield and Slavin (1992). In a retrospective review of his model after 25 years, Carroll (1989, p.30) emphasised that his model of school learning was intended to seek equality of opportunity, which "means not only providing appropriate opportunities to learn (*appropriate*, not necessarily *equal* for all students), but also pushing all students' potentialities as far as possible toward their upper limit". He also added that "when the variables of quality of instruction and opportunity to learn are properly managed, the variable of student perseverance—willingness to learn—will take care of itself" (1989, p.30).

2.5.2 A global model by Dunkin and Biddle (1974)

Dunkin and Biddle (1974) presents a model for classroom teaching that includes all four global groups of variables, namely, presage, context, process and product variables and numerous variables for each of these groups (Creemers, 1994), as shown in Figure 2.5 below.

22 Carroll's emphasis on quality instruction has also led to the identification of a system of instruction labelled under "Direct Instruction", which was found to be the best definition of quality instruction when the desired outcome is scored on standardized tests of basic skills (Adams & Engelmann, 1996).

Figure 2.5: **A model for the study of classroom teaching**



Note: Adapted from: Dunkin & Biddle (1974, p. 38).

However, as Creemers argues, its completeness is its drawback because it fails to specify which and to what extent these variables are important. Some context variables like pupil characteristics and pupil formative experiences can be understood as inputs variables as well. At individual level, the prior experiences and personal characteristics of each student is an input to the classroom, but students in the classroom also collectively form a context for teaching and learning. This context is distinguishable from the classroom environment. As a model within the TER tradition, its centrality lies in the classroom. Given the lack of attention paid to the different levels within education (a multilevel approach), this model ignores the role of process variables beyond the classroom and thus leave no space for school processes. This deficiency distinguishes it from the similar input-process-product models in SER.

2.5.3 Creemers' (1994) comprehensive model of educational effectiveness

In contrast to Carroll's and Dunkin and Bindle's classroom level models, various attempts in SER have utilised three-level models to separate effects at the student, the classroom/teacher, and the school levels (e.g., Creemers, 1994; Campbell et al. 2004; Creemers & Kyriakides, 2008; Kyriakides & Creemers, 2008a, b). These recognise importance of classroom/teacher effects and classroom processes (e.g., Luyten,1995; Hill & Rowe 1996, 1998). Creemers (1994) concluded that teacher behaviour is the most important factor in promoting learning in school. He elaborated Carroll's concept of opportunity to learn further by making a distinction between *available* and *actually used* time and opportunity. He also argued that 'time actually spent' on learning task should be further refined as the amount of curriculum content is covered. Classroom instructional effectiveness factors in Creemers' (1994, p.89) effective classroom model can be summarised as follows:

Classroom effectiveness factors:
Quality of instruction:
Curriculum:
- explicitness and ordering of goals and contents;
- structure and clarity of content;
- advance organisers;
- evaluation;
- feedback;
- corrective instruction.

Grouping procedures:
- mastery learning;
- ability grouping;
- cooperative learning (dependent on differentiated material, evaluation, feedback and corrective instruction);

Teacher behaviour:
- classroom management;
- homework;
- clear goal setting (restricted set of goals, emphasis on basic skills, emphasis on cognitive learning and transfer);
- structuring the content (ordering of goals and content, advance organisers, prior knowledge);
- clarity of presentation;
- questioning;
- immediate exercises;
- evaluation;
- feedback; and
- corrective instruction

Time for learning
Opportunity to learn

On the one hand, SER theorists like Creemers (1994) and Scheerens (1992), who have acknowledged the centrality of the classroom as the immediate context of learning and insisted teacher behaviours to be the most important determinant in promoting student learning, suggest that theories of learning and instruction should be placed at the core of multilevel educational effectiveness models. On the other hand, these theorists also argue that when conceptualising effective teaching, one should not limit to the teaching and learning activities within individual classrooms. School factors can facilitate classroom factors, because they provide a broader context that may affect teaching and learning. For example, while quality of teaching, the curriculum, and the grouping procedures influence time on task and opportunity to learn at the classroom level, they are also influenced by factors at the department or school level. For Creemers (1994), school effectiveness factors may include:

School effectiveness factors:
Quality:
- rules about classroom instruction;
- evaluation policy/evaluation system;
- policy on intervention, supervision, professionalisation

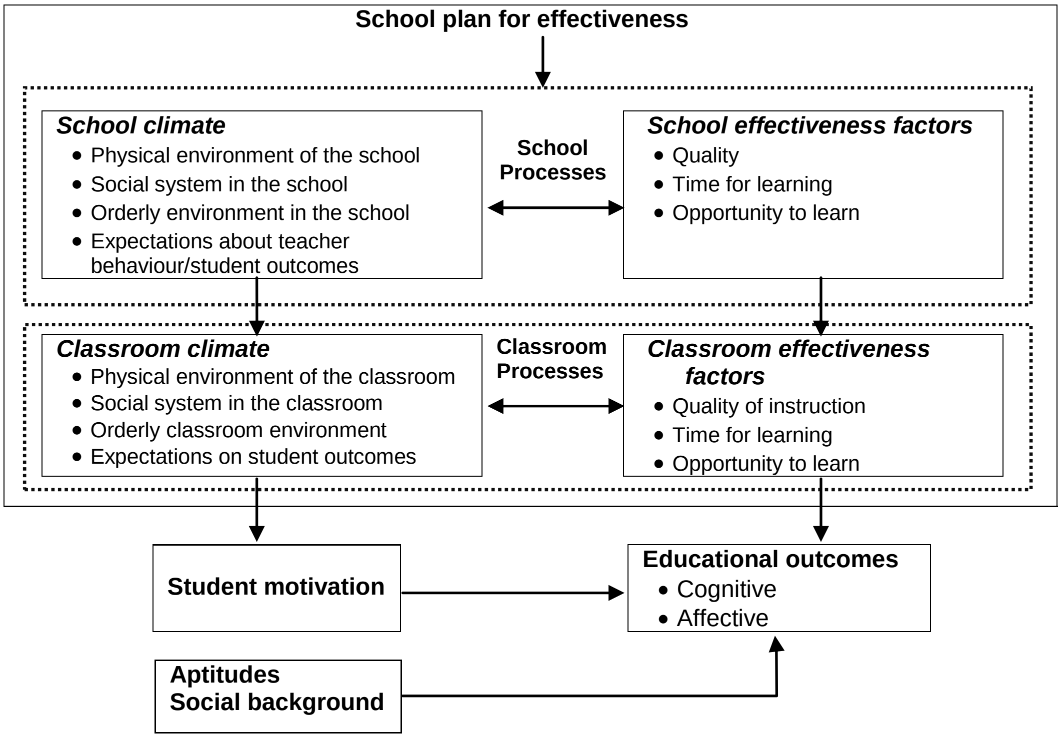
Time:
- time schedule;
- rules about time use.

Opportunity to learn:
- school curriculum;
- rules about implementation of the school curriculum

Higher-level conditions such as school leadership, policy and organisation may facilitate the lower level conditions (i.e., the quality of teaching and learning in classrooms) which, in turn, have a direct impact on pupils' academic outcomes (Bosker and Scheerens, 1994, Hill and Rowe, 1996, 1998). Here, Creemers (1994) hypothesises that student factors like aptitudes, social backgrounds and motivation affect achievements because the students can determine how much they will spend on their school tasks, how much effort they will put into work, and how much assigned work they will complete. For example, Hill and Rowe (1998) found that aptitudes, social background, and motivation also affect students' attentiveness and subsequently their performance.

Like the Hay Mcber (2000) study, climate factors also play an important role in Creemers' model of teacher and school effectiveness, as shown in Figure 2.6 below.

Figure 2.6: **Climate factors in educational effectiveness**



Note: Adapted from Creemers and Reezigt (1999, p.31).

This model represents the prevailing view in SER that the influences of schools on student outcomes are mediated through classroom factors. In particular, while classroom effectiveness factors are under the direct influences of classroom climate and school effectiveness factors, they are also indirectly affected by the school climate. The classroom climate is also postulated to moderate student motivation directly and educational outcomes indirectly. Moreover, climate factors and effectiveness factors can be ideally mutually reinforcing each other, regardless of their levels (Creemers & Reezigt, 1999). Creemers and Reezigt (1999) argue that while effectiveness factors can be superimposed onto schools or classrooms as interventions in some school improvement initiatives, climate factors could not be easily replicated and thus, sustainable change would not occur. They point out that emphasises on supportive climate and network have contributed to the successes of school improvement projects such as Barclay-Calvert project in the U.S.A. (Stringfield & Herman, 1996) and the various Success for All projects[23].

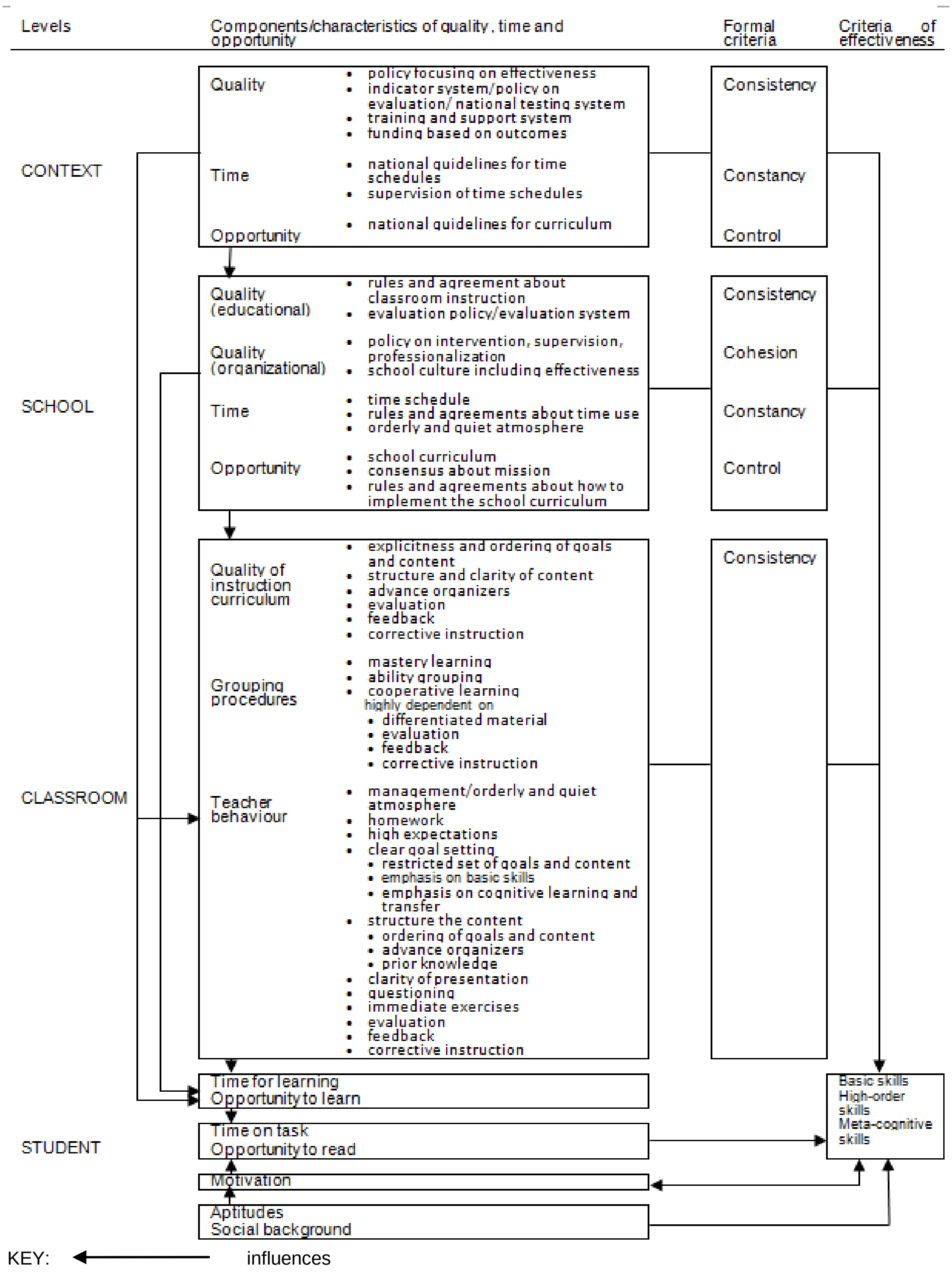
The model depicted in Figure 2.6 is still incomplete as it has left out two features of Creemers' model. First, the role of the students seems to be passive, while it is assumed to be more active. Although Creemers claims that the school and the teacher can plan for the effectiveness and climate factors to a certain degree, the student background characteristics can have strong impacts on the teacher and the school. For example, it has been depicted in Figure 2.1 Durkin and Biddle's (1974) model that classroom processes are best seen as an interaction between the teacher and the students. Therefore, while the teacher is able to influence time for learning and opportunity to learn through the quality of the instruction, the students can also determine how much time they will spend on how much attention they will pay (see attentive behaviours of students in Rowe & Rowe, 1999) and how much they will participate in the learning activities in the classroom. The parents can also affect the student learning directly through their involvement in school provision of learning or indirectly through increasing the time for learning at home (e.g., private tuition).

Creemers (1994) also introduces four formal principles in his complete model (Figure 2.7) to account for the cross-level interactions among factors. According to Creemers' *consistency* principle, factors at the different levels should support each other in order to improve students' achievement. Consistency is expected to be operated within and between levels. *Cohesion* is the second formal criterion, which implies that all teaching staff must show characteristics of effective teaching. Creemers also pays attention to the problem of schooling and argues that the school has to achieve *constancy*, meaning that effective instruction should be provided throughout the school career of the student. Finally, the model states that the school has to maintain *control*, meaning that goal attainment and the school climate should be evaluated. However, these principles have not been discussed and

23 The *Success for all* projects have been implemented successfully in different countries: in the U.S.A., (Slavin & Madden, 2001); in the U.K., (Hopkins, Youngman, Harris, & Wordsworth, 1999; Hopkins, 2001); in China (Zhou, 2008).

studied as often as other aspects of the model because Creemers has not explained how these principles can be observed, measured or evaluated.

Figure 2.7: **Creemers' comprehensive model of educational effectiveness**

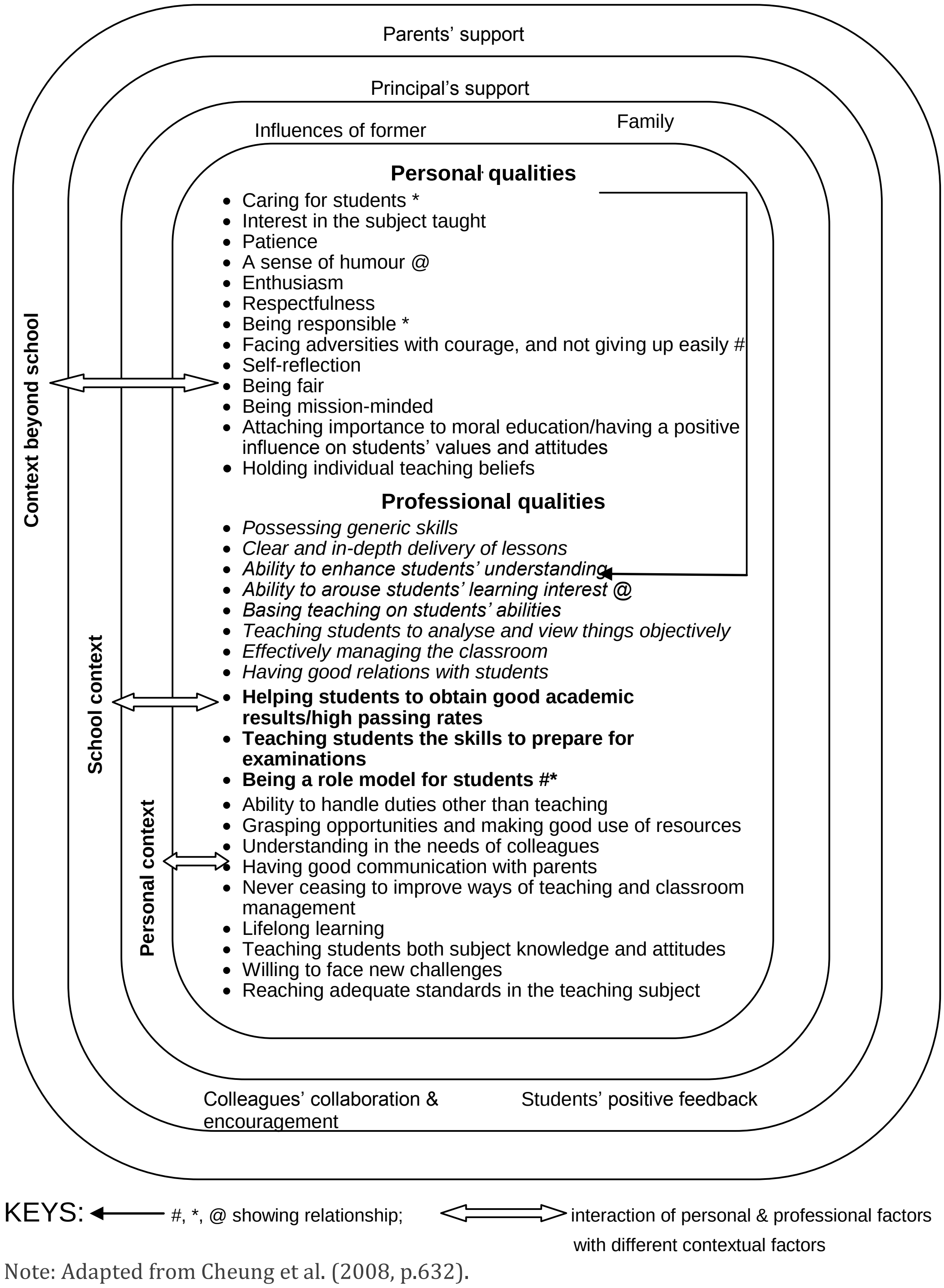


Note: Adapted from Creemers (1994, p.119).

2.5.4 A model of teacher success by Cheung et al. (2008)

Cheung et al. (2008) conceptualise teacher effectiveness in terms of teachers' success in relating to personal and professional qualities that positively interact with variables in the personal, school and beyond-school contexts, as depicted in Figure 2.8.

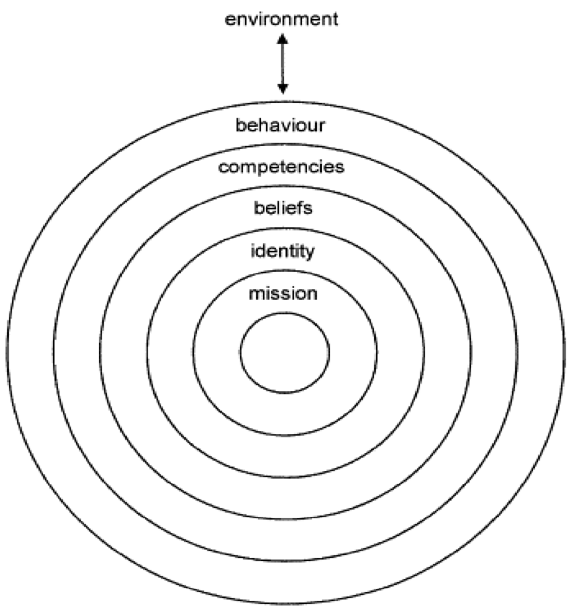
***Figure 2.8*: A model of teacher success in Hong Kong context**



Note: Adapted from Cheung et al. (2008, p.632).

As mentioned in Section 2.3.1, what distinguished professional qualities in Cheung et al.'s (2008) model from other profiles of effective teachers suggested by western researchers is the inclusion of student outcomes and role. Like Dunkin and Biddle (1974), they also consider the impacts of teacher and contextual factors. Regarding the teacher factors, the emphasis on personal qualities is reinforced by their relationships with professional qualities (e.g., 'caring for student', 'being responsible,' and 'facing adversities with courage, and not giving up easily' are associated with 'Being a role model for students'). This emphasis reflects the traditional concept of teacher in Confucian philosophy and similar to (Korthagen, 2004) descriptions of good teachers. Cheung et al. (2008) extend Korthagen's onion model of change (in Figure 2.9) to describe the interaction between teacher factors and the environment (the class, the students, and the school). The significance of the three school level factors highlighted by Cheung et al. (2008) were also similar in Day et al.'s (2006) VITAE study of the effectiveness of English teachers.

Figure 2.9: **The onion: a model of levels of change**



Note: Adapted from Cheung et al. (2008, p.625).

2.6 Current theoretical frameworks on consistency and variation of classroom practices

The four models presented in the last section lay the foundations for three more sophisticated models presented here. These three models were proposed to incorporate a range of empirical findings from both TER and SER literature. They represented the fruits of the latest theoretical development in the field. They are employed in this thesis research to account for the specific problem of characterising consistency and variation in classroom practices.

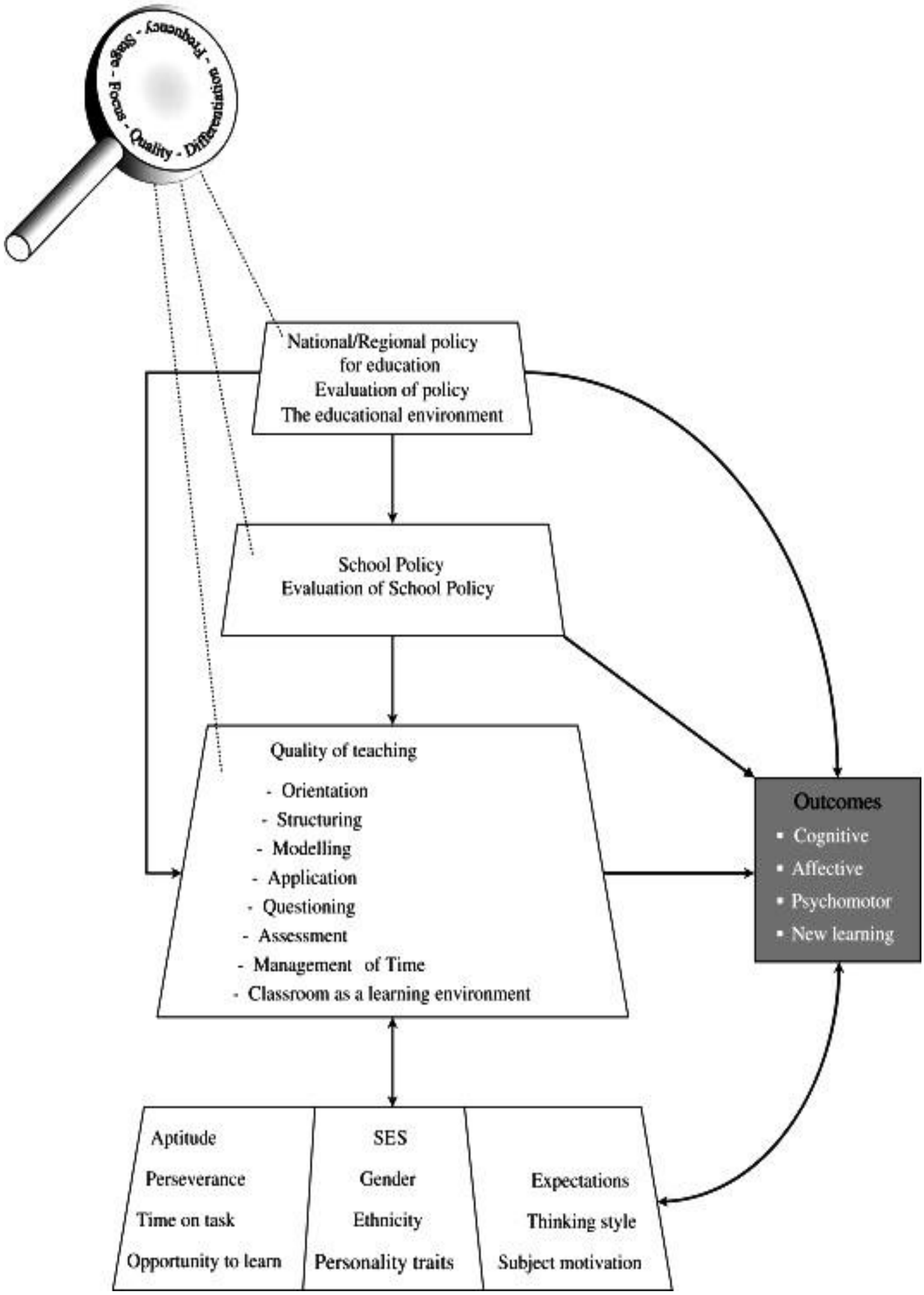
2.6.1 The dynamic model of educational effectiveness (DEE) by Creemers and Kyridakies (2008)

As a revised model of Creemers' (1994) comprehensive model of educational effectiveness discussed in Section 2.5.3, Creemers and Kyridakies (2008) put forward a comprehensive and dynamic, multilevel model that includes effectiveness factors at the student, classroom, school and context levels. As depicted in Figure 2.10, the DEE has student, classroom, school and context level effectiveness factors. These factors are dynamic as some of them are interacting with one another and cross-level interactions may occur. Creemers and Kyridakies (2008, p.149) argue that despite its multilevel structure, the DEE is parsimonious because it:

- takes into account the new goals of education and their implications for teaching;
- searches for interactions among factors operating at the same level;
- investigates the extent to which non-linear relations among some factors and student achievement may exist;
- uses different measurement dimensions to define the functioning of each effectiveness factor;
- describes the complex nature of educational effectiveness.

Creemers and Kyridakies (2008) propose to measure all effectiveness factors in the DEE in terms of five dimensions, namely, *frequency, focus, stage*, *quality* and *differentiation*. This makes the DEE different from previous process and product models, including Creemers's (1994) classroom effectiveness model. In response to the criticism that models of educational effectiveness often lack explicit operation definitions and measurement methods, Creemers and Kyridakies (2008, p.84) has described in detail the operational definitions of the five dimensions of measuring each effectiveness factor and ways of measuring each dimension. For example, differentiation is defined as "the extent to which activities associated with a factor are implemented in the same way for all the subjects involved with it" and this dimensions is measured by the extent to which "different tasks are associated with each factor provided to different groups of subjects involved with each other".

Figure 2.10: **The dynamic model of educational effectiveness (DEE)**



Note: This figure is adapted from Creemers and Kyridakies (2008, p.150).

The DEE can be seen as an ambitious research program[24] that is attractive at least in three ways. First, because of its comprehensive nature, the DEE accommodates variables and factors previously identified to be associated with school or teacher effectiveness in an umbrella fashion. It is an attempt to deal with the shortage of well-developed theoretical models for theory testing in SER and infrequent use of existing theoretical models in testing the relationships between variables (Creemers, 2002; Creemers & Kyridakies, 2008). Its usage of the term 'educational effectiveness' to emphasise the importance of conducting joint school and teacher effectiveness research and the functioning of education system as a whole reflects a pressing need in the field to conduct joint studies on both school and teacher effectiveness. This is because previous joint studies like Mortimore et al. (1988), Teddlie & Stringfield (1993), Opdenakker and Van Damme (2000), and de Jong, Westerhof and Kruiter (2004) have shown that neither level can be studied adequately without taking into account of the other.

Contrasting with the previous models (except Creemers, 1994) discussed, the dynamic model hypothesises that student outcomes are under constant influence from factors at different levels in the education system, rather than confining to factors of any particular level. Its aim of incorporating and integrating findings of research conducted in various disciplinary perspectives would ultimately enrich its comprehensiveness. *The emphasis on comprehensiveness means that it also justifies the need to understand classroom practices of Hong Kong teachers not only in the situational contexts of the classroom, but also other factors operated in their department (e.g., teacher collaboration), their school (e.g., the school's streaming and class composition policies), and the broader educational contexts in Hong Kong (e.g., medium of instruction policy, school places allocation system as mentioned earlier in Section 2.1).*

Second, it recognizes teaching and learning as "dynamic processes that are constantly adapting to changing needs and opportunities" (Creemers &

24 Lakatosian research program (Lakatos, 1970).

Kyriakides, 2008, p.9). Thus, although on the one hand the DEE is expected to be a generic model and effectiveness factors are seen as generic in nature, on the other hand it can incorporate differential educational effectiveness by maintaining that the impacts of effectiveness factors on different groups of students, teachers or schools may vary. Creemers and Kyriakides (2008, p.82) argue:

> ……*we should not overestimate the differential nature of teacher and school effectiveness….[to the extent that] the concept of differential teacher effectiveness ought not to be polarized against a generic concept. Rather, the former should be incorporated as a refinement into the latter.*

This means that the DEE not only has no conflicts with differential teacher effectiveness found in differentiation in teaching, but incorporates it. The theoretical compatibility of the DEE has strong implication in this present research because this means the research interest should not be confined to evidence and explanations for generic characteristics of teaching effectiveness, but also differential teaching effectiveness. These generic characteristics have been hypothesised to be relatively universal across contexts but coexist with variations that reflect contextual influences.

Finally, the extra attention that Creemers and Kyriakides (2008) have paid to developing instruments and testing the validity of the DEE using the Multitrait-Multimethod Matrix (hereafter MTMM) [25] offered SER/TER an attractive, feasible research program and obtained some positive empirical results to support the model (Antoniou, 2009; Kyriakides, 2005; 2008; Creemers & Kyriakides, 2008; Kyriakides & Creemers, 2008a). They have developed instruments especially for testing the five factors (i.e., frequency, focus, stage, quality, and differentiation) operating at four hierarchical levels

25 MTMM is an approach developed in 1959 by Campbell and Fiske (Campbell & Fiske, 1959) to assess the construct validity of a set of measures in a study. To my knowledge, attempts to use MTMM in TER and SER are rare because it is complicated and costly to carry out. However, according to Trochim (2006), multiple methods are not necessary to establish convergent and discriminant validation, suggesting that a modified and simpler approach can be adopted. From the results in previous research (Ko & Sammons, 2008; Muijs & Reynolds, 2000) as well as the current results in Chapter 4 and 5, it seemed that researchers might have established convergent validity (i.e., the degree to which concepts that should be related theoretically are interrelated in reality), but not discriminant validity (i.e., the degree to which concepts that should not be related theoretically are, in fact, not interrelated in reality). Although Creemers and Kyriakides (2008) claimed that they have found support for both types of validity for their constructs, they actually meant the constructs for frequency, focus, stage, quality, and differentiation, rather than constructs for the eight teacher or classroom factors.

(i.e., student, classroom, school, and context). Unlike the classroom observation instruments mentioned earlier, its items are designed to test the five factors for each of the classroom factors separately rather than together as in the instruments employed in this research (i.e., orientation, structuring, modelling, application, questioning, assessment, management of time and classroom as a learning environment). *The present research seeks to contribute to these aims of instrument testing and development and to the theoretical basis for studying features of effective classroom practice in the context of a Hong Kong secondary school.*

2.6.2 The differentiated model of teacher effectiveness (DTE) by Campbell et al. (2004)

Unlike the DEE, the main foci of the DTE are the teacher and the five dimensions of differentiation: time stability, subject consistency, differentiation by people, differentiation by working environment and teacher' expected roles. As shown in Table 2.7, the DTE is simpler than the DEE as it focuses on the classroom level, but differentiations in people and working environment are actually related to student factors and school factors.

Table 2.7: **A model of differentiated teacher effectiveness (DTE)**

DIFFERENTIATED TEACHER EFFECTIVNESS: INSTRUCTIONAL ROLE			
Time stability	***Subject consistency***	***Differentiation***	
		Different people	*Working environment*
1. School year 2. Phase of implementation of an educational policy 3. Teaching periods 4. Periods in relation to the assessment of a teacher	1. Curriculum subjects 2. Areas within a subject 3. Difficulty of a teaching unit 4. Type of teaching objectives	1. Group of students (sex, age, SES, learning needs) 2. Colleagues 3. Parents	1. School type 2. Availability of resources/support 3. School culture 4. Community
DIFFERENTIATED TEACHER EFFECTIVNESS: ACROSS VARIOUS ROLES			

Note: This table is adapted from Campbell et al. (2004, p.84).

The first four dimensions are hypothesised related to teachers' instructional role. In the present context, classroom observation conducted at different phase of the school year, teaching periods, and school year can be considered as differentiation related to time stability. Although only one subject was chosen in this research, differentiation related to subject consistency can include areas within a subject, difficulty of a teaching unit.

Examples of differentiation of people are groups of students by sex, age, SES or learning needs. For differentiation in working environment, examples are school type, school culture, availability of resources, and/or community.

The term 'differentiation' of the DTE highlights its primary concerns on variation in teacher behaviours, rather than effectiveness factors seen as generic in nature. The DTE is attractive in two ways in the present context. First, it takes into account of the problems of identifying consistency and stability of teacher effects on student achievement and contextual factors like differentiation in people and working environment. For example, Campbell et al. (2004) argue that teacher effects may vary across subjects and, within a subject, across different subject areas and teaching objectives. Teacher effects are less likely to be stable but fluctuate over the school year, across different phases of implementation of an educational policy, across different teaching periods, and across lessons in which observation/assessment is taken place. This means that certain sampling procedures and controls are required for studies focusing on generic effectiveness factors. *In particular, the present research seeks to address the problems of subject consistency and time stability.*

Second, the DTE is a model that its proponents have put so much emphasis on the argument that "effective teaching is underpinned by moral values" to an extent that they insisted that (Campbell et al., 2004, p.113):

> ... any model of teacher effectiveness must include an analysis of values of the schools and teachers involved in teacher effectiveness research or appraisal.

Although most studies of teacher effectiveness adopt a goal-oriented model of measuring effectiveness (Stufflebeam & Shinkfield, 1995) in which it is often assumed that "a teacher is effective if she/he can accomplish *the planned goals and assigned tasks in compliance with school goals*" (Campbell et al. 2004, p.61). The value-driven nature of school goals and the diverse values of the stakeholders are seldom acknowledged in TER. As argued in Section 2.3, effectiveness is a value-laden concept that reflects the stakeholders' values. Accordingly, Campbell and his colleagues questioned the approach to treat teacher effectiveness as value-free and discussed the value of independent leaning in the context of how the concept of learning has been conceived (i.e., what counts as worthwhile achievement or learning) and the value of inclusiveness in the context of what kind of classroom climate or teacher-pupil relationship is desirable. In other words, teacher effectiveness reflects the value-driven choice and priorities of the agents and the extent to which the agents can exercise their powers in their own spheres. Similarly, class composition (or grouping) and medium of instruction policies reflect the conflicting values between stakeholders of the Hong Kong education system that affect teacher effectiveness to different degrees. *These contextual factors are considered crucial in understanding variation of classroom practices of the teachers in the Hong Kong case study*.

2.6.3 Marzano's (2003) probability model of educational effectiveness (PEE)

Unlike the last two models discussed, Marzano's (2003) PEE model focuses on the impacts of differential teacher effectiveness on students over various stages of their schoolings. Recently, Sammons and Luyten (2009) have discussed different approaches to study the absolute effects of schools and the differential teacher effects of the individual teachers who teach the students in their periods of schoolings, if longitudinal data are collected. The existence of differential teacher effects in a department would highlight the challenge to maintain consistent performance among teachers. Strong instructional leadership and a focus on improving teaching and learning are expected to achieve a high level of consistency in teaching effectiveness in a department. By showing that there may be effective departments in an ineffective school and ineffective departments in an effective school, Sammons et al. (1997) argue that it is important to look at individual departments using value-added approaches as this can show trends over time, a department's relative effectiveness for different groups of student, and variations between different subject departments. *This research seeks to investigate differential teacher effectiveness in a single department and school and thus, Marzano's model is considered particularly illuminating*.

The PEE illustrates why consistency in teaching effectiveness is important to individual teachers as well for a department. As a rule, higher quality work from more teachers in a department is a key factor that influences the value-added effectiveness of a department (Sammons, et al., 1997). Variation in teacher effectiveness among the teachers in a department affects not only the teaching quality of the lessons, but is also likely to affect student outcomes. For example, Table 2.8 illustrates six hypothetical scenarios of academic outcomes of a student in the first three years of schooling in a secondary school in Hong Kong under the current academic structure, if the PEE is applied. These scenarios are based on the assumption that the student is of 50^{th} percentile when s/he is first admitted to a school and taught by teachers who vary in teaching effectiveness in the junior form years. The six scenarios describe the joint effects of teacher and school on student achievement of students entering school at the 50^{th} percentile.

Table 2.8: **Effects on student achievement of school and teacher effectiveness with student entering school at the 50th Percentile**

School and Teacher Scenario	Achievement Percentile After Three Years
Average School and Average Teacher	50^{th}
Least Effective School and Least Effective Teacher	***3^{rd}***
Most Effective School and Least Effective Teacher	37^{th}
Least Effective School and Most Effective Teacher	***63^{th}***
Most Effective School and Most Effective Teacher	96^{th}
Most Effective School and Average Effective Teacher	78^{th}

Note: Adapted from Marzano (2003, p.74).

In the PEE, the school effect between schools is estimated to be about three-fourths of a standard deviation between the most effective schools and the average effective schools as well as between the average effective schools and the least effective schools. Table 2.8 above shows that a very effective teacher may not be easily noticeable in the least effective school because there is only a gain of 13^{th} percentile in student achievement, but the negative impact of a very ineffective teacher can lead to a huge decline of 47^{th} percentile in student achievement. For an average effective teacher working in a least effective school, her/his student performance is more likely to drop below the average to the 34^{th} percentile.

2.6.4 Applying the theoretical frameworks to account for consistency and variation

Rather than treating the DEE and the DTE as two opposing rival models, a more fruitful approach is to see them as alternative models that emphasise either variables or contexts more, but neither of which can exclude the other. From the above accounts, it is suggested that advocates for the DTE would disagree with the following statements:

- An effective teacher is effective in all dimensions and effective across all contexts.
- An ineffective teacher is ineffective in all dimensions and ineffective across all contexts.

However, they probably would not object to the weaker forms of these statements:

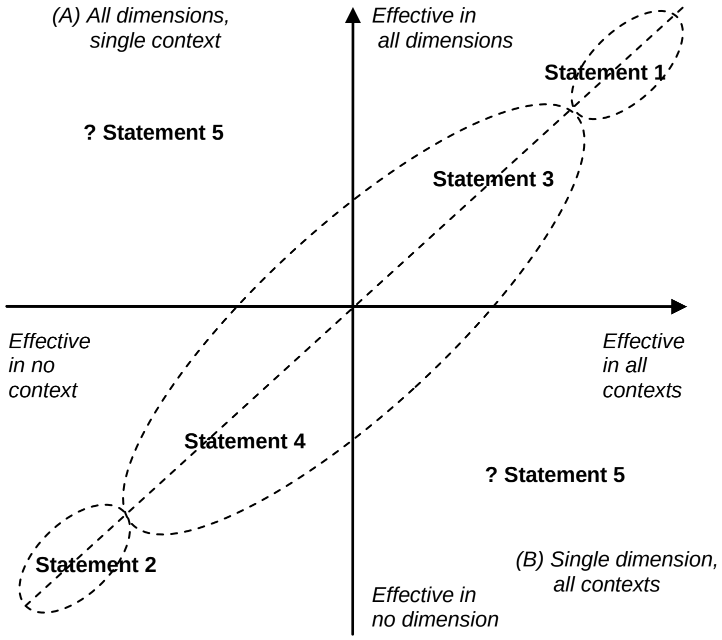
- An effective teacher is effective in most dimensions and effective in most contexts.
- An ineffective teacher is ineffective in most dimensions and ineffective in most contexts.

Yet, these statements still conform to the generic theorist's prediction that effectiveness is a generic feature of an effective teacher. In contrast, the differentiated theorist might agree with the following:

- A less effective teacher is ineffective in some dimensions and in some contexts only, but effective in some other dimensions and effective in some other contexts.

Consistency can be operating along two continuums that indicate the variability in effectiveness across different dimensions of teaching behaviours and the variability in effectiveness across different contexts. Figure 2.11 depicts different scenarios of consistency defined in the two continuums and indicated the scenarios where Statements (1) – (4) above would most likely describe. A generic theory of teacher effectiveness tends to predict that most teachers would fall along the dash line in Figure 2.11. A strong version predicts heavier weights at the two ends (like Statements 1 and 2), while a moderate version predicts a heavy middle (like Statements 3 and 4):

Figure 2.11: **Two continuums defining consistency**



In contrast, a differentiated theory emphasises contexts more and allows possibilities in other areas, it can offer better account for regions beyond the three ellipses, probably extending towards, regions where teachers may be effective in a single dimension in most contexts (i.e., Region A) or effective in most dimensions but in few contexts (i.e., Region B). In these cases, moderate generic theorists may insist that these teachers are all less effective as "*the concept of differential teacher effectiveness ought not to be polarized against a generic concept*" (Creemers & Kyriakides, 2008, p.82). However, like the differentiated theorists, they probably would show more interest in the interplay between dimensions and contexts than an extreme generic theorist does, because there may be strong relationship between certain dimensions and certain contexts. Though it may be debatable whether these teachers should be labeled as ineffective, their teaching behaviours are undeniably inconsistent. In this study, the interest is to specify the following:

- A teacher's teaching effectiveness is inconsistent if s/he is effective in some dimensions (A_i) in some contexts (C_j).

Thus, Statement 6 may occur in Region A or B as in Figure 2.11. It is regarded that characterising their inconsistencies is more constructive than labeling them as effective or ineffective teachers.

2.7 Implications of the literature reviewed and proposed research questions

Based on the literature reviewed above, four implications can be drawn. First, the contextual analysis in Section 2.2 suggests system-wide contextual factors in Hong Kong may affect the instrumental goal of teaching practices in relation to student attainment in examinations, while other goals such as equity, learner autonomy, or citizenship, may be marginalized. This would be explored in more depth in case studies. Second, literature indicates that although there is strong evidence for multidimensionality of teaching practices, the dimensions found seem to be subject to cultural influences and classroom observation instruments employed. This motivated the present employment of different instruments intended for better contextual validity to explore generalisability of dimensions. Third, the significance of contextual factors such as school leadership, collegial support and student feedback seemed to be illuminated in the teachers self-reports and case studies (e.g., Cheung et al., 2008; Day et al., 2006) and many school improvement studies (e.g., see Reynolds, Bollen, Creemers, Hopkins, Stoll, & Lagerweij, 1996; Harris, Day, Hopkins, Hatfield, Hargreaves, & Chapman, 2003; Hopkins, 2001). Thus, this would justify a mixed methods (MM) approach that incorporates qualitative data to explore contextual factors in a single department and a school (more details in Chapter 3). Fourth, the various theories concerning teacher effectiveness indicate that consistency and variation in teaching practices may affect individual teacher effectiveness and collective teacher effectiveness, but have not received enough regard.

Accordingly, there are strong links between the present research with previous TER and SER, but its theoretical contributions may be dependent on its methodology. Theoretically speaking, the present research places teacher effectiveness and classroom processes as the central foci. This approach is in accord with the traditional TER and those SER studies see the teaching and learning process as a major influence on student progress in school. In terms of methodology, the present research differs from mainstream SER but shares with the tradition of TER in its employment of systematic classroom observation instruments as one of its major data collection methods. These instruments contain items that describe some predefined teacher behaviours that have been found predict positive student outcomes in the literature. These instruments do not just look at teacher behaviours related to classroom effectiveness factors like quality of instruction as in Creemers' model (1994) but also include items concerning those teacher behaviours that promote a positive classroom climate.

Teachers' relative effectiveness is thus operationally defined in terms of the frequency or the strength of the observed teacher behaviours as specified scale items and profiles of effective lessons can thus be established. The advantage of this approach is that its objectivity is achieved through reliability and validity of the scale (or instrument) used and student outcome measure is not always necessary in such a study. This is a rather different approach in comparison with SER, in which teachers' relative effectiveness is often operationally defined in terms of their students' progress in some standardised or attainment tests. While assuming the presence of teacher effects, the limitation of this retrospective approach is that researchers often cannot specify what dimensions of teacher behaviours are more crucial than other, because it often does not involve any classroom observations. These two approaches are not mutually exclusive however, as some research has examined their links (see Muijs & Reynolds, 2000 and Antoniou, 2009).

A further step to look at teacher effectiveness is to use a purposive sample of effective teachers whose effectiveness is independently determined in other study or other methods. This was applied in the ECP study (Day et al., 2008), which also enjoyed the advantage of using different classroom observation instruments, not for determining the relative effectiveness of teachers, but for identifying *generic* characteristics of effective teaching practices. These generic characteristics were thus operationally defined as those observed teaching behaviours that were more consistently found in the effective teachers.

The present research takes another step to compare the teacher behaviours of a sample whose relative teacher effectiveness (as defined in Section 1.2.1) has not been measured. The purpose for this research is not to establish the relative teacher effectiveness of this sample but to explore the consistency and variation of effective teaching practices across lessons of the teachers. The underlying dimensions of observed teaching behaviours are to be considered as distinctive generic characteristics of effective teaching behaviours of the current sample. The amount of variability of these dimensions thus would inform consistency and variation of teachers' classroom practices*. In order to do this, a focus on a single department and one school is justified because this would minimise possible variation attributable to contextual differences.* Instead of employing a multilevel modelling approach to investigate variation between departments in different schools as in studies employing, *the present research adopted a case study approach to study factors that may affect within-department variation (i.e., variation among teachers) in more depth*.

Based on the above discussion, a set of seven research questions are formulated below and addressed in different relevant chapters:

- What are the characteristics found in the observed classroom practices across a large number of lessons? How do they vary with student backgrounds and class compositions? (*Chapters 4-8*)
- To what extent are these characteristics comparable to those identified in the English study by Day et al. (2008), despite the sample and contextual differences? (Chapters 4-6)
- To what extent are the characteristics identified using different quantitative observations instruments comparable? (Chapter 6)
- To what extent do these characteristics contribute to the quality of teaching in the lessons observed? (Chapter 6)
- To what extent do these characteristics contribute to the individual involvement by the students in the lessons observed? (Chapter 6)
- To what extent do these characteristics vary among individual teachers and vary across the lessons of each teacher? (Chapters 7 and 8)
- What are the teachers' views and perceptions about their teaching practices, their students' learning and the contextual factors that may affect teaching and learning in the school? In what way are they affected? (Chapters 7 and 8)

The first two research questions are intended to compare the different claims made by the generic theories of teacher effectiveness (GTE) and the differentiated theory of teacher effectiveness (DTE) regarding consistency and variation in classroom practices. According to GTE, teachers would *consistently show similar ratings in different dimensions of their classroom practices across different contexts*. In contrast, DTE would hold the view that *the ratings of teachers in different dimensions of their classroom practices tend to vary across different lessons and contexts* (e.g., in terms of different age groups or different ability groups of students). An advocate of a generic theory is also more likely to believe that different dimensions of teaching practices would be similar for individual teachers (i.e., generally effective in most aspects or generally typical or generally less effective) and in different cultural contexts.

The third research question addresses both theoretical and methodological issues. It will contribute to the theoretical debate between GTE and DTE because a proponent of a generic theory is more likely to believe that different classroom observation instruments could measure similar, rather than different, underlying dimensions of teaching behaviours. Clearly, the findings related to this question are very much dependent on the particular instruments selected. Thus, adopting a methodology that allows instrument comparison in this research will contribute to an issue rarely fully addressed in the existing TER literature, particularly since the instruments used differ in that are derived from an evaluation (inspection) perspective while the other is based on a lower inference approach.

Given that no student level quantitative data was collected, the fourth and fifth questions have to be addressed by associating the distinctive characteristics, or underlying dimensions of observed teachers' behaviours identified, with two global indicators of overall teacher effectiveness. This method was used by van de Grift (2007), but the statistical method he used for making the association was only Pearson correlation. This is considered an undesirable statistical method for its strength of interpretation, so multiple regression was employed instead. Multiple regression results are expected to show the relative contributions of these characteristics to overall teacher effectiveness as judged in terms of the instrument.

The answers to the last two questions are considered to be crucial for providing rich descriptions of an embedded case study. These descriptions include an account of the challenges and paradoxes in the difficult contexts of these teachers that may affect their teaching practices, the unique characteristics as identified in their patterns of the ratings as identified in the various underlying dimensions of their observed teaching behaviours, and the similarities and differences of their lessons.

Finally, before moving to the next chapter on the methodology of this study, it is important to reiterate the characteristics of the context of the present study. The four system-wide challenges in the education system of Hong Kong, namely, the secondary school places allocation system, streaming and setting, medium of instruction policy (MOI), and examination-oriented culture, are important contextual variables that affect the school, the department, the teachers and the students in the study directly or indirectly.

CHAPTER 3 : RESEARCH METHODOLOGIES

3.1 Introduction

This chapter on the research methodology of the thesis research is divided into five sections. Section 3.2 addresses the philosophical foundations of the present thesis research. By adopting an MM research design in the CVCP study, it has taken a stance that places the research questions as the drives of the study and a pragmatic philosophical position that does not view qualitative and quantitative as incompatible and fits its purpose to bring the best out of the quantitative and qualitative paradigms. It is argued that the success of this attempt is dependent on the extent of integration that MM research can make and the extent to which the research can show the distinctive qualities of MM.

Section 3.3 outlines the basic research design. First, this includes descriptions of the procedures and products of the various phases of data collection and analyses and interpretations of results. Then, research questions proposed at the end of last chapter are rephrased in operationalised terms and methods of analyses. Finally, the rationale for employing an MM embedded case study design is discussed.

Section 3.4 briefly describes the pilot study conducted prior to the main CVCP study. The implementation of this pilot study highlighted some difficulties, but it also contributed to the adoption of a case study methods to achieve the aim to understand the influences on teaching English in a challenging context in a Hong Kong secondary school.

Section 3.5 reports the sample, the instruments, and the data collection procedures of the CVCP study. The rationale behind the sample and instrument selections is also discussed. This section provides a background for understanding the results presented in the next four chapters.

In Section 3.6, some ethical issues are addressed. These include anonymity and confidentiality, reciprocity and teacher-research relationship and intrusion of the researcher as an observer. These issues highlighted the dynamic tensions between the researcher and the research participants.

3.2 Philosophical foundations of mixed methods and the extent of integration

3.2.1 Philosophical assumptions and paradigmatic issues

CVCP is intended to contribute to MM research. MM research is seen as an emerging alternative to the rivalry between qualitative and quantitative traditions discussed by various methodologists (e.g., Brannen, 1992, 2005; Bryman, 1988, 2006a; Creswell, 2002, 2007, 2009; Creswell & Plano Clark, 2007; Johnson & Christensen, 2004; Tashakkori & Teddlie, 1998, 2003b). For example, in the preface of their edited handbook on MM, Tashakkori and Teddlie (2003a) regarded "the ongoing emergence of mixed methods as the third methodological movement in the social and behavioural sciences" (p. xi). Allowing a flexible methodology that would integrate qualitative and quantitative techniques to simultaneously address multiple and diverse research questions, MM are considered superior to single approach designs in three ways (Teddlie & Tashakkori, , 2003, pp.14-15):

- MM research can answer research questions that the other methodologies cannot;
- MM research provides better (stronger) inferences;
- MM provide the opportunity for presenting a greater diversity of divergent views.

However, Greene and Caracelli (1997, p.5) noted that "using multiple and diverse methods is a good idea, but is not automatically good science". Perceiving and internalising conflicted epistemologies, some researchers see the advocates of quantitative and qualitative methods as tribes fighting over incompatible issues in the so-called "paradigm wars" (Gage, 1989). In contrast, Howe (1988) argued against the incompatibility of quantitative-qualitative paradigms from a pragmatic perspective, emphasising 'what works' methodologically the best is more important than the epistemological incompatibility of the competing positivistic and interpretivist paradigms. That is, "[p]aradigmatic philosophical assumptions are less important than the myriad 'practical demands' of the particular research problem when making choices about data collection and interpretation" (Rocco, Bliss, Gallagher et al., 2003, p.596; see also Bryman, 2006b).

In terms of philosophical inclinations, many researchers (e.g., Howe, 1988; Maxcy, 2003; Teddlie & Sammons, 2010; Teddlie & Tashakkori, 2003b) argued that MM is closer to pragmatism[26]than post-positivism[27]. Tashakkori and Teddlie (2003) argued that pragmatism is popular among MM research because in applied settings, like education, health sciences, and evaluation, where MM meets the needs to employ multiple data sources required for investigating complex social phenomena or making practical decisions, researchers tend to place research questions higher priority over the epistemological or paradigmatic issues. However, they also noted that there are criticisms of the pragmatic stance by some who believe it might overlook some irreconcilable divergences between qualitative and quantitative approaches in their ontological and epistemological assumptions.

In addition to the *pragmatic position*, Greene and Garacelli (1997 p.10) also called for a *dialectal position,* based on which a "synergistic" use of methods is found deliberately "shaped by both interpretivist and post-positivist paradigms in an integrative manner". Gorard and Taylor (2004) further claim that qualitative and quantitative methods can be complementary and combined to provide a better understanding of the object of study which cannot be gleaned by using either method alone. Using their empirical research on teacher effectiveness in different stages of teachers' professional life cycles as an example, Day, Sammons and Gu (2008, p.331) showed that "*synergistic* understandings that enabled the discovery and delineation of key findings that were both more enlightening and more robust than would have been the case if one method or another had dominated". In other words, the end product of MM should be more than the sum of the individual quantitative and qualitative parts.

Besides the pragmatic and dialectic stances, they initially proposed (Greene & Caracelli, 1997), Greene and Caracelli (2003) argued that there are two other stances in MM research. A *new paradigm* stance in MM research holds that new or emerged paradigms are "superior to older

26 Pragmatism has been associated with the philosophies by Charles Pierce, William James, John Dewey, Arthur Bentley, Abraham Kaplan, Richard Porty, Richard Bernstein, and Cleo Cherryholmes (Maxcy, 2003)

27 Postpositivism has been associated with the works of Karl Popper, Norwood Hanson, Thomas Kuhn, Imre Lakatos, Paul Feyerabend, Stephen Toulmin, Larry Laudan and William Newton-Smith (Phillips, 1990)

historical paradigms because they invite multiplism in methods and perspectives" (Greene & Caracelli 2003; p.96). In contrast, like the pragmatic stance, a *concept-driven* stance would regard paradigms issues rather unimportant in inquiry decisions. Instead, conceptual or theoretical congruence is considered mattering more.

Recently, Teddlie and Sammons (2010) argued that the dichotomy between quantitative and qualitative is actually false and damaging as it "only reflects the legacy of the so called 'paradigm wars' in social research evident during the last thirty years". They found that in educational effectiveness research (hereafter EER), the dichotomy has led the qualitative and quantitative camps rely on single method in their research. The quantitative camp has increasingly engaged in large-scale investigations that seek to identify and measure differential school and teacher effectiveness in promoting student's educational outcomes by the statistical prediction and explanation of variance in these outcomes. In contrast, the qualitative camp has tended to engage more in promoting effective school improvement initiatives and teacher development programs through action research and case studies in which 'thick' descriptions, rather than statistical predictions, are used to enhance understanding of school and classroom processes and participants' perspectives.

However, Sammons (2010) has highlighted the limitations of EER's traditional reliance on studies largely conducted within a single research paradigm, either quantitative or qualitative. Instead, citing the latest research by her and her colleagues (Sammons, Siraj-Blatchford, Sylva, Melhuish, Taggart, & Elliot, 2005; Sammons, Day, Kington, Gu, Stobart, & Smees, 2007; and Day et al., 2008), Sammons argues that the potential contributions of MM studies to EER lie in its enabling a "*dialectical" dialogue* between quantitative and qualitative researchers which increases "interplay in the interpretation of findings to create *synergistic understanding.*" It seems that Sammons regarded a dialectic stance in MM would eventually develop into a new paradigm stance.

MM is more common that often assumed (see Teddlie & Tashakkori, 2003b for a historical analysis). Using the work of some eminent psychologists like Leon Festinger (Festinger, Riecken & Schachter, 1956) and Stanley Millgram (Milgram, 1974), Maxwell and Loomis (2003, p.242) pointed out that MM research were practised in both natural and social sciences with a longer history than its explicit discussion, “when methods were less specialised and compartmentalised and the paradigm wars were less heated”. According to Maxwell and Loomis, the qualitative and quantitative distinction only highlighted two contrasting approaches to explanation in scientific discourse: variance theory, which deals with variables and their correlations, and process theory, which deals with events and the process that connect them. Thus, mixing quantitative and qualitative components in a study would mean, for example, the relative emphasis of variance theories and process theories in the researcher’s conceptual framework and the relative proportion of variance questions and process questions (e.g., on how and why, meaning and context, physical causality) in the research questions of a study. Similarly, citing Niglas (1999), Greene and Caracelli (2003, p.106) argued that although paradigms may theoretically matter in MM inquiry, there is strong evidence indicating that “research practice is as likely to commonly blend or mix features of different paradigmatic traditions”.

3.2.2 The challenges regarding the extent of integration and the distinctive qualities of MM research

Whether the attempts to bring the best out of the once-dichotomized quantitative and qualitative research methodologies may signal just an optimistic hope of some academies or represent “a new era in the conceptualization and utilization of integrated approaches” (Tashakkori & Creswell, 2007, p.3) is dependent upon two factors. First, it is imperative for researchers to demonstrate the extent to which MM research can integrate the qualitative and quantitative approach to provide meta-inferences and new understandings that cannot be otherwise obtained by reliance on only one methodological worldview. Second, there needs to be an accumulation of research studies that can show a genuine integration of the two approaches through which extra evidence is generated (Tashakkori & Teddlie, 1998;

2003a).

Regarding the first factor, Bryman (2007) recognised, there is still not enough attention paid to the extent to which MM researchers have genuinely integrated their findings. Bryman (2007, p.8) noted that we have to cast doubt on the extent to which MM has been genuinely integrated in a single piece of research: that is, “whether the components of a mixed methods investigation are related to each other or whether they are either totally or largely independent of each other”. Even when the two components are related, there is a question of whether the two may be impeded in their conception and findings may be overstated. For example, Porter and Gamoranm (2002, see also LeTendre, 2002) noted that *Trends in International Mathematics and Science Study* (TIMSS), a highly regarded international study, had also fallen short of integrating its case study and video study (i.e., qualitative data) with its achievement surveys (i.e., quantitative data).

However, there are some researchers like Brewer and Hunter (1998) and Morse (2003), who accepted the quantitative and qualitative are complementary to each other but rejected any form of integration, seeing mixing of methods as a serious threat to the validity of the MM research. If the qualitative and quantitative components are kept separate with no genuine integration, it is unlikely that this type of MM research would produce new understandings based on the integration of results and meta-inferences but two sets of unrelated findings. Moreover, the value of MM will be diminished if they are just a strategy to justify researchers’ methodological eclecticism (in Hammersley’s, 1996, term) or pragmatic considerations.

A definition of MM that simply requires a qualitative and a quantitative component in a study may be too loose, because inconsistencies and disagreements often arise as the possibilities of mixing the two components seem to be infinite (Johnson & Christensen, 2004). Many researchers (e.g., Bryman, 2006a; Creswell & Plano Clark, 2007; Creswell, Plano Clark, Gutmann, & Hanson, 2003; Newman, Ridenour, Newman, & DeMarco, Jr, 2003; Maxwell & Loomis, 2003; Morse, 2003; Teddlie & Tashakkori, 2006, 2009) have outlined many tentative typologies of MM research.

However, there lack some clear criteria for MM design than simply a typology of research design. For example, in a critical examination of the various MM designs identified by Creswell and Plano Clark (2007), it can be found that only the *triangulation* design will generate new evidence beyond essential quantitative or qualitative findings. It may be correct to classify the other three types (i.e., *embedded*, *explanatory*, and *exploratory*) as "*non-integrative"* MM designs, because researchers who adopt these designs often do not need to specify clearly how they are mixing the methods or mixing the data. Instead, it may be more appropriate to consider these research designs may involve *multiple data* or *multiple methods*, but not *MM*. Thus, it is arguable whether some rigid criteria are required for an MM design.

Following the arguments for synergistic understandings in MM research by Day, Sammons, and Gu (2008), '*an intent to integrate'* and '*the extent of integration'* can be regarded as the two key criteria for evaluating how much evidence, findings, or explanations are really the products of MM. It may not be appropriate to regard a researcher is adopting a true MM design, if s/he fails to show a clear and strong intention to generate some additional findings or interpretations beyond the scope of findings and interpretations originally raised by using either quantitative or qualitative approach. In other words, the four stances in MM research put forwarded by Greene and Caracelli (2003) would suggest the different levels and extent of integration may be adopted.

In her review of MM research in education, Sammons (2010) has documented studies that showed four distinctive qualities: synergistic by findings, non-linear by research process, richness-driven by choice of evidence, and not dependent on research domain. The first and the last qualities reflect the consequences or the extent of integration in an MM study, while the second and the third qualities reflect the researcher's intent to integrate and the extent of integration in an MM study. As quantitative and qualitative data often carry equal weight in MM research, a researcher often have to adopt an inductive/deductive logic which requires him/her constantly move "to and fro" in a non-linear research cycle in collecting, evaluating and interpreting evidence. An MM researcher's choice of evidence is not guided by paradigmatic concern, but driven by maximising the relative richness of the evidence. Certainly, a researcher cannot always be sure that such a research process and choice of evidence would necessarily lend to synergistic findings, but Sammons has reported some successes in applying MM in EER and witnessed the findings often have strong implications in various domains of EER (i.e., TER, SER). Accordingly, MM is not just another methodology, but also a methodology that can integrate different domains in EER that has been dominated by research largely in single paradigms.

The later sections will outline the research design and clarify the implementation of the research project. The intention to integrate and the extent of integration are clarified in these discussions. The synergistic findings and their implications will be addressed in the concluding chapter.

3.3 The Research Design

3.3.1 Phases, procedures and products

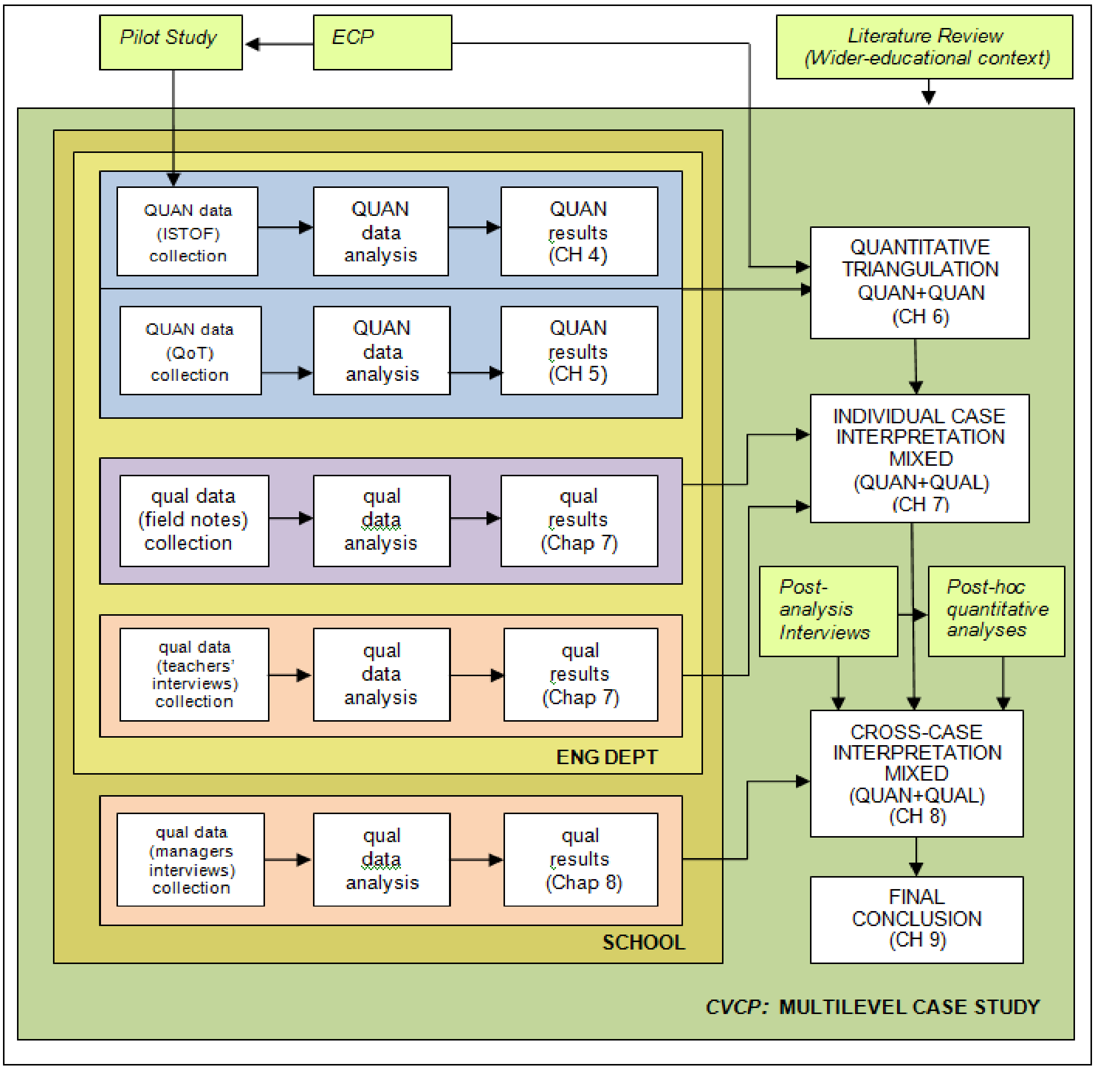
Both the experience and findings in the earlier ECP study and the pilot study adopted in this research study were found useful in informing the research strategies and research design of the main CVCP study. Intended to test the instruments and procedures in the ECP study in a secondary school, the pilot study informed the infeasibility of a large-scale replication. The details are addressed in Section 3.4. The CVCP was originally intended to be a single-phase study. However, an additional follow-up phase gradually emerged after the initial quantitative analysis. This follow-up phase involved three brief focus groups interviews and some further quantitative analyses. Thus, the research consists of two temporally distinct but conceptually overlapped phases as depicted in Figure 3.1[28] This figure illustrates the process, product and extent of data integration with an emphasis on the weighting[29] and triangulating of the quantitative and qualitative data and findings. The research design and process as shown in Figure 3.1 indicate the intention to integrate and the approach to generate meta-inferences in

28 The notations in the diagram follow the examples illustrated in Morse (2003) and Creswell and Piano Clark (2007).

29 Following the convention used by Creswell (2007) and Creswell and Plano Clark (2007), heavier weight is denoted by capital letters.

the CVCP study. The multilevel study was built on layers of different types and sources of data and different methods of analyses.

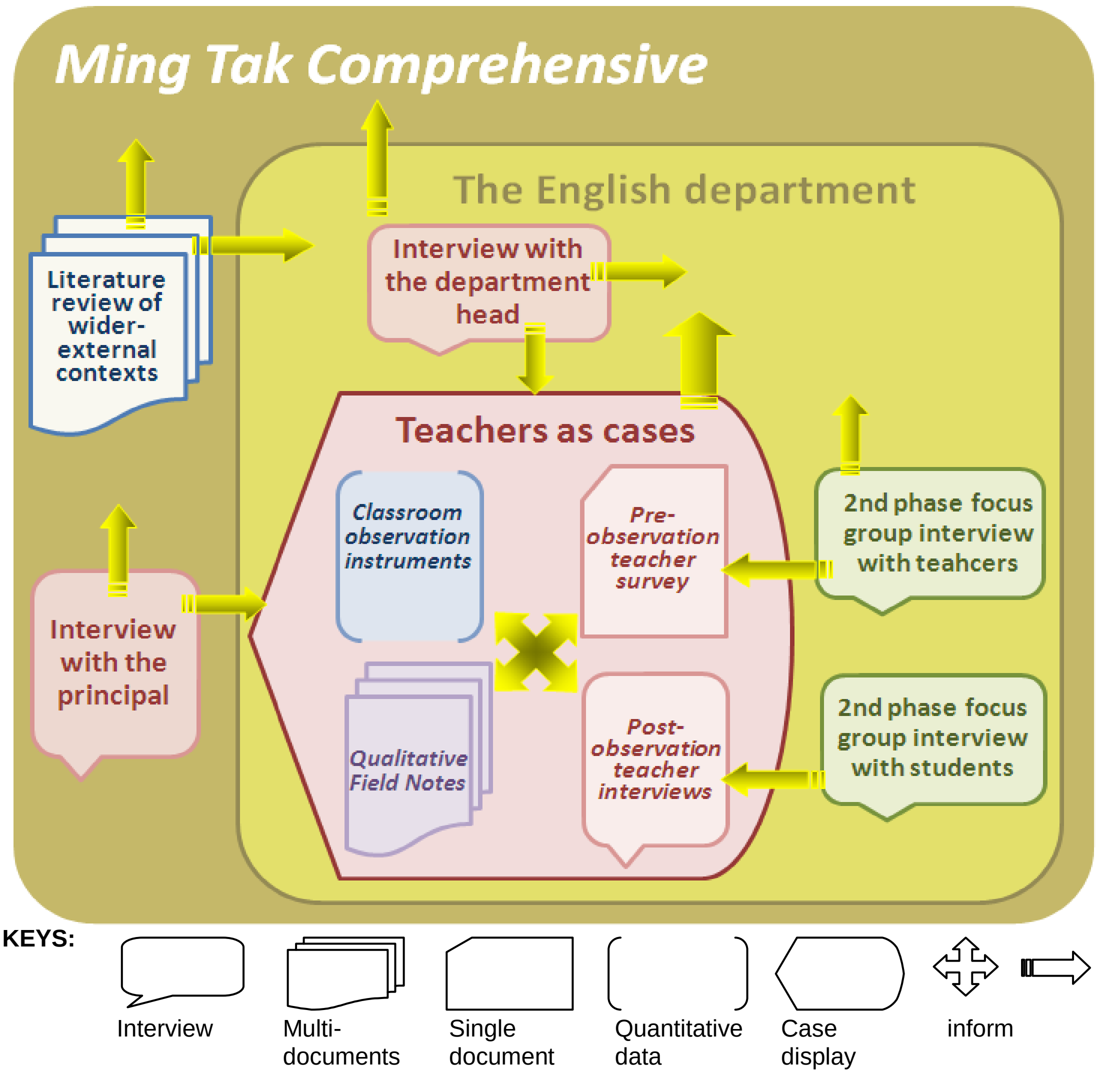
Figure 3.1: **Phases, procedures and products of CVCP**



At the bottom teacher-level layer, quantitative findings (*blue in Figure 3.1*) were triangulated with *the lesson* as the unit of analysis. The second layer involves validating the quantitative ratings of teaching behaviours using systematic observation protocols with qualitative field notes (*purple*) collected during classroom observation. In addition to the validation, interviews (*pink*) structured to elicit participants' views on classroom practices and factors affecting them were integrated to form holistic case descriptions for each teacher. At the last and top school layer, the individual case results were understood against a background of challenges faced by the English department and the school suggested in the results of the analysis of interviews (*pink*) with the English department head and the school principals.

A follow-up phase that explored some unanticipated issues brought up after the initial quantitative results led to further quantitative and qualitative analyses. These *new* results *were combined with the initial mixed individual case study results* and *the qualitative findings from the interviews.* These cross-case interpretations addressed factors found to affect the behaviours of different teachers of the English department and their relationships with the consistency and variation identified in observed teachers' behaviours across lessons. Based on the procedures outlined above, the triangulations between the qualitative and quantitative can be depicted as Figure 3.2.

Figure 3.2: **The MM design of the CVCP research: Illustration of sources of quantitative and qualitative data used to construct four case studies and inform the cross case study analysis**



This diagram extends the illustration in Figure 3.1 by showing how the different sources of quantitative and qualitative data were integrated to study consistency and variation within and between four teachers and used to create and underlie both teacher case studies and the cross case comparisons.

The multilevel design of the CVCP research does not conform to the two variants discussed in Creswell and Plano Clark (2007) (i.e., *embedded experimental* and *embedded correlational*)[30]. By structure, the present design shows layers of embedded data, but by function, it involves triangulation, exploration and explanation at various stages of the research process which cannot be classified as either *experimental* or *correlational*. This design allowed gradual data integration to build up a case about the impacts of contexts on teaching effectiveness of EFL teachers in a Chinese medium instructed school in a socioeconomically disadvantaged area in Hong Kong.

3.3.2 Research questions and their operationalisations

With respect to the gaps in the literature discussed in Chapter 2, the research design illustrated in Figure 3.1 was intended to address a set of research questions listed in Section 2.7. These questions are hereby rephrased in operationalised terms and analyses in Table 3.1:

Table 3.1: **Research questions and their corresponding operationalisations**

	Research Questions	**Operationalised in terms of analyses**
Chapters 4-8	What are the characteristics found in the observed classroom practices across a large number of lessons? How do they vary with student backgrounds and class compositions??	What are the underlying dimensions of observed teaching behaviours identified in the *confirmatory factor analyses*? To what extent do these underlying dimensions vary across lessons as in their *frequency distributions*?
Chapters 4-6	To what extent are these characteristics comparable to those identified in the English study by Day et al. (2008), despite the sample and contextual differences?	How far are these underlying dimensions comparable in the CVCP sample and the ECP sample in *terms of their related items/indicators*? How are these underlying dimensions *cross-validated* in the CVCP sample and the ECP sample?

30 This discrepancy, however, is not seen as a deficiency of the present design but rather a consequence of Creswell and Plano Clark's (2007) inconsistent classification. They seemed to use labels like *triangulation, explanatory*, and *exploratory* to refer to the different functions of different designs, but clearly the label embedded is better considered as referring to the structure, rather than the function of a design.

	Research Questions	Operationalised in terms of analyses
Chapter 6	To what extent are the characteristics identified using different quantitative observations instruments comparable?	How are the underlying dimensions identified in different instruments comparable in *correlation analyses*?
Chapters 6 and 8	To what extent do these characteristics contribute to the quality of teaching in the lessons observed?	How are the underlying dimensions associated with the overall indicator of teaching quality in *multiple regression*?
Chapter 6	To what extent do these characteristics contribute to the individual involvement by the students in the lessons observed?	How are the underlying dimensions associated with the indicator of positive involvement of the pupils in *multiple regression*?
Chapters 7 and 8	To what extent do these characteristics vary among individual teachers and vary across the lessons of each teacher?	What are the *characteristics* of the observed teacher behaviours noted in the *qualitative field notes of the observed lessons*? What are the *consistency and variation* found in the various underlying *dimensions* of observed teaching behaviours? How do the teachers vary in the underlying dimensions in ANOVA and *discriminant functions* using the underlying dimensions as predictors?
Chapters 7 and 8	What are the teachers' views and perceptions about their teaching practices, their students' learning and the contextual factors that may affect teaching and learning in the school? In what way are they affected?	What are the *main themes* identified in the teachers' views and perceptions regarding their teaching practices, their students' learning and the contextual factors that may affect teaching and learning in the school?

3.3.3 Rationale for case studies

Compatibility of MM and case studies

Elliot and Lukeš (2008, p.88) proposed to see educational case study as:

> a form of inquiry into a particular instance of a general class of things that can be given sufficiently detailed attention to illuminate its educationally significant features… Such a view of case study is methodologically open. Methods need to be justified pragmatically in terms of their 'fitness for purpose' rather than in terms of a priori principles derived from a theory of knowledge.

Their rejection of defining case study based on methodological terms was based on the fact that such an approach often led to unproductive paradigm wars. Thus, their definition seemed to be grounded on similar arguments for MM research discussed in Section 3.2 and formed the rationale for the adoption of both qualitative and quantitative methods in the present case study.

While the case studies were based on teachers and the quantitative sample on lessons in the present study, this marriage of MM and case study resulted in a quantitative component based on samples of lessons in both teacher case studies and cross-case study comparisons. Elliot and Lukeš (2008, p.96) argued that "[t]he role of educational case study research was therefore to complement rather than supplant the study of samples."

Unit of analysis and case selection strategy

The CVCP study was built on case studies of four EFL teachers teaching English as a foreign language to Cantonese speaking students in an underperforming CMI secondary school in which all other subjects are taught in Cantonese. This case selection strategy was motivated by four reasons. First, the lesson was chosen as the unit of the main quantitative analysis. The decision was motivated by maximising statistical power for an adequate but small sample size. In order to obtain perform a meaningful quantitative analysis like factor analysis with statistical power, a large sample size is always recommended (Brown, 2006; Hair, Black, Babin, Anderson, & Tatham, 2006; Tabachnick & Fidell, 2006). However, Marsh and Hau (1999) has noted that the determination of an appropriate and adequate small size is not a simple function between the sample size and the number of scale items, but varies with the complexity of the confirmatory factor analysis (hereafter CFA) models.

Based on their analysis on the ECP data, Ko and Sammons (2008b) found that a small sample size of 79 teachers could adequately produce a meaningful six-factor CFA model with 30 items from a scale of 45 items. This meant that a similar sample size as that of the ECP study was required if the current CVCP study was to employ the same instruments in that study. However, it was estimated that the researcher had to gain access to over twenty schools in order to maintain a similar sample size. This was an impossible task for a single doctoral researcher even before the pilot. This meant that using *the teacher* as the unit of quantitative analysis as in the ECP study was not feasible.

Second, choosing the lesson as the unit of the main quantitative analysis was also motivated by changing a disadvantage in sample selection to an advantage. In the literature, research evidence based on observation on one occasion suggested that teachers were quite consistent when viewed *as a group*, but quite inconsistent when considered *individually* (Emmer & Peck, 1971; Moon, 1971; Rosenshine, 1973). This meant that although the results of the ECP could be reliable and generalisable to effective teachers a group, they were *unreliable to characterise the behaviours of individual teachers across lessons*. As the researcher had no means to select a purposive sample of effective teachers as in the ECP study, a tentative research design for the current study could not allow for studying a large teacher group quantitatively.

However, by shifting the research focus from *the teacher* to *the lesson* as the unit of quantitative analysis, the current study could address the limitations of the ECP study by informing the consistency and variation of the observed teaching behaviours of *a few teachers*, rather than *a specific teacher group*. In order to do so, it was essential to observe a minimum number of lessons of a teacher such that the teaching behaviours could be representative of a larger sample of classroom practices of that teacher. According to Rosenshine (1973, p.221), "11 to 20 observations would be required to obtain a stable mean score for each teacher", but sometimes it may require up to 30 observations (or 16 hours of observation) to measure teaching behaviours related to some cognitive aspects of learning. Accordingly, it was decided that a five-day observation period of about 20-25 lessons of a teacher would be sufficient to produce a stable mean score to characterise the teaching behaviours of a teacher.

Third, it was hypothesised that quantitative findings could be easier to interpret if the contextual variables were controlled by focusing on the teachers of the same department of a single school. Following Stake's (2005) definition of a case as a choice of what to be studied from a special perspective and with a special interest, a focus on the lessons of a few teachers of the same department of a school could allow a depth of data that were more illuminating. Based on the literature review on some system-wide challenges that may affect the classroom practices as discussed in Section 2.6, it was hypothesised that the English department of a CMI school might have to face more system-wide challenges. This is because despite its importance as a license to further education and career prospect, English is taught as a content subject that is hardly appealing to students who rarely use it in their social lives. For example, teaching English as a content subject is a difficult task because the lack of enriching linguistic environment, poor student motivation, inadequate family support, and negative washback effects of public examinations are common hindrances that undermine learning outcomes and negatively moderate teacher effectiveness. In other words, the EFL teachers are bounded in not only a school, but also a system with all these unique contextual characteristics.

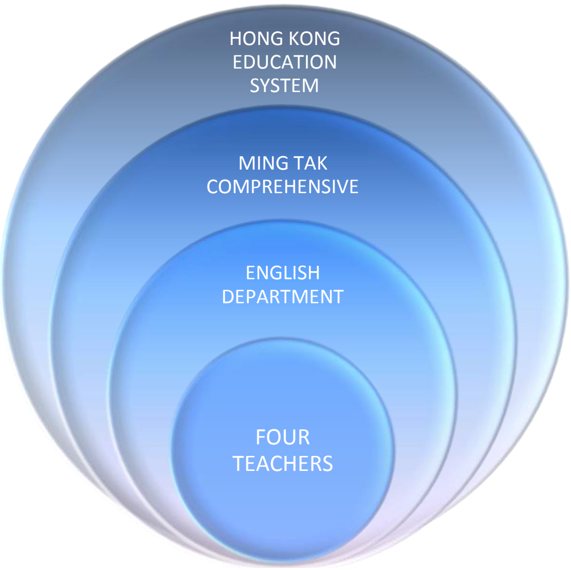
Finally, to conduct a study in an underperforming secondary school was motivated by the intent to choose a school from a socially disadvantage area in Hong Kong in which there were more schools with less favourable public examination results (i.e., a crude indicator of lower effectiveness in the eyes of the public). It was expected that the current results could be linked to the findings to the literature of schools in challenging contexts elsewhere as discussed in Section 2.3.3.

Multiple units of analysis and the multilevel design

The CVCP was a study containing more than one sub-unit of analysis which could be classified as an embedded case study (Yin, 2003). Although the unit of quantitative analysis was the lesson, there were several subunits in the current study, among which the basic qualitative unit was *the teacher*. Each of the participant teachers in the CVCP study was a case study that served a purpose to show the differential teacher effectiveness within a single department. It might be debatable whether these individual teachers might be multiple cases suitable for synthesis into a single case, or rather they should be considered as “multiple experiments -- that is, to follow a ‘replication’ logic” (Yin, 1994, p.45). Researchers like Yin (2003) tend to treat multiple cases as replications of an experiment, which aim at generalisable theoretical propositions, while these propositions are not necessary generalised to population.

As the identification of sub-units in an embedded or a multilevel study allows for a more detailed level of inquiry, each of the subunits not only is the constituent of different hierarchical levels, but also informing the same problem. As depicted in Figure 3.3, the teacher, the department and the school are not only the subunit of the various hierarchical levels of the Hong Kong education system, but they also inform the relationship between differential teacher effectiveness and the contextual factors at different hierarchical factor.

Figure 3.3: **The multilevel study design**



This design resembles Korthagen's (2004) 'onion' model of levels of change and Cheung et al.'s (2008) model of teacher success discussed in Section 2.5.4. Thus, it is assumed that cross-level interactions exist and they may be revealed in the case studies and cross-case analyses. That is, the consistency and variation in the classroom practices of the lessons of individual EFL teachers of the English department of a CMI school would inform what challenges these teachers might have to face and what strategies they chose to employ individually and collectively in the same school settings. However, it should be noted that because no data were collected other than the English department and only one school was observed, the present study did not attempt to generalise the findings to other departments and to other schools. Thus, data concerning the higher levels of the system in Figure 3.3 were limited. In other words, the multilevel study was still mainly an account of the consistency and variation in the classroom practices of the lessons of four EFL teachers of the English department of a CMI school.

Multilevel embeddedness, multiple data, and multiple methods

A multilevel study adopts a case study research methodology that relies on multiple sources of evidence to add breadth and depth to data collection, to assist in bringing a richness of data together in an apex of understanding through triangulation, and to contribute to the validity of the research (Scholz & Tietje, 2003; Yin, 2003). When a research inquiry moves upward from the teacher or classroom (or micro) level to the department and the school (or meso) level and from there to the education system (or macro) level, there is a tendency to rely more on qualitative data. This is partly because variables at the higher hierarchical level are more difficult to operationalise quantitatively, but mainly because the sample required for conducting a multi-level quantitative analysis will increase dramatically (at least 25 to 30 times for each level). In contrast, the case study methodology is not subject to statistical power. Instead, multi-level embeddedness may enhance the understanding of a case if the descriptions of the contexts of each level can enrich the overall descriptions and interpretations.

To enhance the depth of the case descriptions in the CVCP study, multiple data types were collected using different collection methods. As the main data were collected through lesson observation, the advantages and disadvantages of classroom observation protocols have been discussed more fully in Section 2.2.4. According to Croll (1985), the qualitative field notes are usually collected during the observation to supplement descriptions of the classroom process such that they would shed light on the reasons why the researcher rated a particular teaching behaviour. Accordingly, in Figure 3.2, findings of quantitative ratings by two different systematic observation schedules are not just compared with, but also validated, or not validated, by findings of qualitative field notes as the second type of data. According to Creswell and Plano Clark (2007, p.84), this is a “triangulation design-validating quantitative data model" involving "[using] qualitative information to validate the quantitative results".

The third type of data was transcripts of interviews with the teachers, the department head and the school principal. Interviews with multiple stakeholders in the school (i.e., the teachers, the department head, and the school principal) are often the most common method to elicit different perspectives (Frey & Fontana, 1991; Fontana & Frey, 2005). Semi-structured interviews enjoy the advantages of being flexible and spontaneous as in unstructured interviews and the advantages of being directive and phenomenological as in structured interviews. Therefore, semi-structured interviews with not only would add in extra voices, but also highlight how different stakeholders' roles may affect their perceptions of the same issues. These different data types were expected to form different layers of meaning in the interpretations of the four case studies.

3.4 The pilot study

3.4.1 Purposes

A pilot was conducted for three purposes. First, it was essential to trial the quantitative classroom observation instruments and to develop a framework for the qualitative field notes and the post-observation interviews. Second, it was expected that the pilot would inform a better research design and administrative procedures if the main study was to be carried out in a more extensive scope in different schools. Third, it was intended to explore the feasibility of studying a broad range of subjects and a broad spectrum of teacher characteristics and student groups.

3.4.2 Sample

A convenience sample was recruited in a substandard size secondary school[31] in Hong Kong with the assistance of the school principal. Ten participating teachers taught various subjects including Chinese Language, English Language, Geography, and Liberal Studies. These teachers varied in

31 A standard size school in Hong Kong generally has 5 classes of 40 students for each year level, but this one had only 3 classes in each year level.

age, gender, and teaching experience, but generally had less than 3 years of experience in the school.

3.4.3 Instruments

Both classroom observation protocols used previously in the ECP study (Day et al., 2008) were trialled. A framework for the qualitative field notes was developed and the teacher survey and the post-observation interview were modified to suit the Hong Kong contexts (for details see Section 3.5.3).

3.4.4 Data collection procedures

The pilot study was administered in a three-day visit in May, 2008 in Hong Kong. The same classroom observation protocols, ISTOF and QoT, as in the ECP study and the pilot study were used, except that these two instruments were completed on the same occasion to cover the same lesson rather than on two different occasions. This allowed for divergent findings more easily to be attributed to differences between two instruments rather than different occasions. Because the two instruments were used to rate the same lesson, it was possible to explore consistency in teachers' practices observed across instruments as there was no discrepancy due to a different time and context. Although all teachers were given a copy of the teacher questionnaire, only half of them were returned. Only one interview was conducted in the pilot because many teachers and the principal regarded it rather time-consuming given that they were observed once.

3.4.5 Results suggesting feasibility and difficulties for the main study

The feasibility of doing a replication of the ECP study in Hong Kong was found low and several difficulties were anticipated and called for modifications. First, regarding the classroom observation instruments, the results suggested that the ecological validity was generally evident, except that the component of QoT concerning *Effective classroom layout* was found not readily applicable in the school contexts of Hong Kong. Lee et al. (2003) also noted that Hong Kong teachers seemed to be particularly weak in adapting the physical characteristics of the classroom for instructional purposes. Although this component was deleted in the latest version of QoT (see van de Grift, 2007), this component was retained in the CVCP without any change because it was decided that it would enable comparison with the findings of the ECP research.

Second, although only half of teacher surveys were returned, teachers seemed to have no particular difficulties in filling the items in the questionnaire. For those teachers who did not return the survey, problems seemed to be related to their reluctance to spend extra time on filling out the questionnaire or disclose some of the biographical information and their attitudes to teaching. This suggested that the practical value of the teacher survey was going to be limited in the main study.

Third, given the unenthusiastic feedback of the teachers, the post-observation interview was not fully implemented. Teachers were rather reluctant to spend an hour to do the post-observation interview if only one or two observations were to be observed. It was also considered impractical to conduct a study with a large amount of qualitative data from the interviews given the sample size required due to the limited resources of a single researcher. From the only interview and other informal contacts with other teachers, the researcher learned that teachers might not be very comfortable about discussing their behaviours in the classroom and avoid emphases on issues related to aspects of teachers' lives and their well-being that were more personal. This suggested those parts of the interview referring to more personal matters might not be suitable for the cultural contexts of Hong Kong.

In contrast, in the interview and the informal contacts with the teachers, it seemed that teachers were more willing to talk about the impacts of the responses of their students in the classroom and the broader external contexts on their teaching. This suggested that the CVCP main study should place more emphases on the relative importance of teacher-student interactions and the broader external contexts. These results were consistent with the findings that teachers would be less supportive of observation for judgmental appraisals and performance oriented (Lee et al., 2003).

Finally, the pilot study was initially designed to investigate teacher effectiveness based on one lesson per teacher. This would require many participants. The difficulties in gaining access to schools suggested that it may be impossible for a PhD researcher to recruit enough schools and teachers to participate in the main study. Moreover, the results suggested that the subject taught and year level of the class could be major confounding variables if the sample selection in the main study could not systematically control them. Thus, the most practical research strategy was to maximise the number of classroom observations while keeping a manageable number of schools and teachers such that research questions in could still be addressed accordingly. This suggested a multilevel study design was more likely to fit the research purpose for the main study after the pilot study.

3.5 The CVCP study

3.5.1 Purpose

The CVCP study was intended to provide data to address the research questions outlined in Section 3.3.2 in Hong Kong.

3.5.2 Sample

The four EFL teachers selected for the in-depth case study in Chapter 7 belonged to a large department of 10 teachers as English was a major subject in the school. The fictitious names given to these teachers were Charlie, Lucy, Sally and Linus[32]. By choosing teachers from one subject department, it was possible to provide a better focus on teacher variation because possible differences that might occur between departments were avoided. Teachers were observed in all their lessons with a range of students age groups. However, it was originally intended that the participant teachers selected would teach at least one class of the target Form (i.e., Form 5) to allow for some specific comparisons of teaching practices for the same year group of students. This strategy became impossible when one teacher who withdrew from the study was replaced by Charlie, who only taught Forms 6 and 7. The inclusion of Charlie might pose a potential challenge to the validity of the data. First, he might act differently as he was the department head. Also, Charlie only taught Form 6 and Form 7 classes where students were post-16 students, about half to two-thirds of whom had not studied in the

32 A Chinese name was used for the school but English names for teachers as it is a common practice that EFL teachers usually use English names among themselves.

same school before. Because these students were in senior forms, they might have been motivated and better-behaved, making classroom management and student engagement easier. Nonetheless, having an additional teacher in the case studies was essential to the quantitative component to ensure a sufficient number of lesson observations for analysis. Also by including the head of department, it was possible to explore whether how his role might influence his observed classroom practice and teaching strategies. However, given the possibility that his role and the different age group of students might affect results in certain analyses, his data were excluded. This was done in Section 8.4.5, where the contrast between Junior Form and Senior Form was made.

Besides working in the same department, the four teachers also shared a lot in common. Their educational background (having a degree and teaching certificate majoring in English teaching and qualified for the benchmark test), age (all in their early thirties), and teaching experience (about 7-10 years) were similar. Detailed descriptions of each teacher are presented in the individual case studies in Chapter 7. Given that the school was a CMI school, the English proficiency of the students has not been high. As the English department was also known for their relatively lower value-added results, studying this department was expected to illuminate its many challenges internally and externally (for details see Section 8.3).

There were seventy-six 35-minute lessons observed, covering all the school years of secondary education. It should be noted that nearly all the observations were conducted in two consecutive lessons. Table *3.2* shows the distribution of the lessons observed tabulated by teacher and school year.

Table 3.2: **Distribution of the lessons observed tabulated by teacher and school year**

Teacher	**Secondary School Year**							
	Form 1	Form 2	Form 3	Form 4	Form 5	Form 6	Form 7	Total
Charles						7	8	15
Lucy	6	8	2		7			23
Sally		10		6	4			20
Linus	8		4		6			18
Total	14	18	6	6	17	7	8	76

Both the choice of the department and the choice of the school were intended to highlight the internal and external challenges. The school, *Ming*

Tak Comprehensive (a fictitious name), was an underperforming secondary school located at a socioeconomically disadvantaged residential area[33]. *Ming Tak Comprehensive* was considered underperforming because its attainment and value-added results of many subjects including English were lower than Hong Kong average and the average of schools with similar contextual backgrounds[34]. The district is known for its high percentage of low-income and single parent families. Though there is private housing around, most students are living in the nearby public housing estates or rural villages. Like most of the schools in the area, about one-third of the students' families are receiving social benefits. Most of their parents are housewives, unemployed, blue collar workers or low-income service providers like security guards, drivers, and salespersons. There were another nine secondary schools within 2 miles radius of *Ming Tak Comprehensive,* but over 30 secondary schools of the same catchment area. None of the ten schools nearby uses English as the medium of instruction, but competitions among schools remain intense as the popularity of a school is reflected in the proportion of the top, middle and low bands of students in the intake. *Ming Tak Comprehensive* has gradually lost its share of middle band students in the area, from previously two thirds to currently one-third of its intake. In terms of student intake, *Ming Tak Comprehensive* has shown a spiral downturn in attracting more students of higher ability in its catchment area. Selecting this school for study can inform some possible causes of its downturn.

Schools in the area tend to publicise themselves with banners hanging around its exterior walls, showing its attainment and sometimes value-added results. Although valued-added results are not published to the public, schools with less desirable results are more likely not to publicise it when they market themselves. Thus, parents of prospective students can still evaluate the academic performance of a school based on what is marketised and what is *not* marketised. The word-of-mouth among local residents and

33 The general background about the neighbourhood, the community, and students' families were informed by mainly by my knowledge and working experience there, but they had been verified with other teachers I knew.

34 I have not obtained the actual value-added results in the past four years, but the general picture was confirmed in the meetings with the school principal and the English department head.

the recommendations of the primary school teachers may also affect the image of a school and eventually its intake.

3.5.3 Instruments

Classroom observation instruments

For eliciting the quantitative data, two different instruments were adopted: the *International System for Teacher Observation and Feedback (ISTOF) Scale* (Teddlie et al., 2006) as in Appendix I and the *Lesson Observation Form for Evaluating the Quality of Teaching (QoT)* (van de Grift et al., 2004) as in Appendix II. Both instruments had been used in the earlier ECP project that had stimulated the research focus of this study in another context. Although there are many classroom observation instruments available (as discussed in Section 2.3.5), none has been developed specially for the Hong Kong context. The current two choices were considered more appropriate than other instruments as they had been developed for application in international contexts. In addition, by using international instruments in a specific context (i.e., EFL department of a disadvantaged school in Hong Kong), this study had the potential to make an additional contribution by establishing the applicability or otherwise of existing constructs related to teaching practices and how they may be interpreted in terms of teaching effectiveness and individual teacher effectiveness.

International System for Teacher Observation and Feedback (ISTOF) Scale

The ISTOF was intended to be an observation protocol for measuring generic characteristics of teacher effectiveness in lessons with a broad external validity for a variety of country and cultural settings. The scale was produced as part of a collaborative, cross-national research initiative by the Methodology of Research in Effectiveness (MORE) group of the International Congress for School Effectiveness and Improvement (ICSEI) involving 21 countries. The MORE group consisted of researchers, practitioners and education advisers/inspectors, whose opinions about what constitutes effective teaching were used to generate the various components in the instrument ISTOF.

By an iterative, multiple-step, and internet based modified Delphi technique, an original scale of 103 items in 11 components was initially obtained and reduced into the present form of 45 items in 2006[35]. While the seven theoretical components retained, the current 45 items became more administrable for data collection and more manageable for analysis. These 45 items were descriptive statements specifying a particular teacher behaviour (e.g., *Item 1: The teacher makes explicitly clear why an answer is correct or not)*. Of the 45 items, two to four items were grouped to represent an indicator that describes a certain dimension of teaching behaviours. As a result, there were 21 indicators (e.g., *Indicator 1.1: The teacher gives explicit, detailed and constructive feedback)*, two to four of which were further grouped under one of the seven theoretical components (e.g., *Component 1: Assessment and Evaluation*). Though counting the occurrence of the specific teaching behaviours is not required, the rating is expected to be based on the observed relative frequency of the behaviours.

The Lesson Observation Form for Evaluating the Quality of Teaching (QoT)

The inclusion of the QoT was intended to utilise professional judgments within a deliberate high inference evaluative framework. It was a product of the collaboration between Her Majesty Inspectorate and the Dutch Inspectorate, after their mutual agreement in 1996 that led to a series of comparative studies on the instruments used by the inspectors in England and the Netherlands. Thus, the framework was expected to conform to an inspection model that emphasised on what constituted effective teaching or good practices based on the professional judgment of the English and Dutch inspectors.

According to van de Grift (2007, p.128), “the standards and indicators [of QoT] must be observable in (almost) each lesson” such that the instrument could be used every time in classrooms an inspection visit. This means that the QoT would not be appropriate for measuring events that may not happen in every lesson such as “opportunities to learn, monitoring pupils’ results and special measures for struggling learners” (van de Grift, 2007,

35 This is different from the version reported in Teddlie and his colleagues (2006), which only had 43 items.

p.129). The QoT is an event sample instrument similar to the SSOS (Schaffer, Nesselrodt, & Stringfield, 1991) and the VTBI (Teddlie, Virgilio & Oescher, 1990; Virgilio, Teddlie, & Oescher, 1991). Accordingly, inspectors are expected to *rate teachers' behaviours in terms of their perceived effectiveness*, *rather than their frequency*. The internal consistency, inter-rater reliability and validity of the QoT and its application to identify the quality difference in the teacher strategies were confirmed in a study on England and the Netherlands (van de Grift, et al., 2004) and later in another study on two countries England, the Netherlands, Flanders (Belgium) and Lower Saxony (Germany) (van de Grift, 2007).

Like the original Dutch inspection instrument, the QoT Form comprised a detailed checklist of twenty-six indicators (e.g., *Indicator 1.1 [The teacher] ensures a relaxed atmosphere*), covering nine criteria (e.g., *Criterion 1: Safe and orderly school climate*) for evaluating the quality of teaching.[36] To facilitate making judgment, each indicator is supplemented with a few corresponding descriptive statements of teaching behaviours as good practice examples. Raters are instructed to give a score indicating more strengths than weaknesses only when all good practice examples (if applicable) are really observed. This 2004 version of the QoT differed from its Dutch predecessor in its inclusion of an *overall grade* for teaching to reflect an *overall judgment of the lesson quality*, which was a distinctive characteristic of the English instrument. According to van de Grift (2007), the advantage of such an inclusion would thus allow a correlation analysis be performed on the Dutch indicators and the overall grade. It was expected that the correlation analysis would indicate which teacher behaviours have the greatest association with the global judgment of teacher effectiveness and eventually a set of indicators suitable for an international comparative analysis of characteristics of effective teaching would be developed.

Scale comparisons between ISTOF and QoT

Despite its origin as a professional instrument for lesson observation

36 According to van de Grift, et al. (2004), the original Dutch instrument only had 23 indicators and 7 criteria. This older version is different from the one recently reported in van de Grift (2007), which has only 24 indicators. The two indicators deleted in the latest version concern the classroom layout (i.e., Indicator 9.1: [The teacher] ensures the classroom layout supports the pupil activities and Indicator 9.2: the teaching environment is educational and contemporary). This new version was not available prior the conduction of the ECP study.

used by inspectors, the QoT shares aims similar to those of the ISTOF in its development. The two scales are also similar in the dimensions of teaching behaviours they cover. For example, both concern classroom climate, classroom management, clarity of instruction, strategies to cater for individual differences, etc. These similarities suggest that the two scales are conceptually comparable. However, the two scales still differ much in their usage or administration. For example, a major distinction between the two lies in their scales for evaluation. In the ISTOF, a five-point Likert scale is used to indicate how *often* the teacher behaviour is seen, while in the QoT, a four-point Likert scale is used to indicate the *relative strength* or *relative weakness* in regard to the indicator in concern. Although relative strength (as used in the QoT) and relative frequency (as used in the ISTOF) of a particular teaching behaviour are related, they are conceptually and practically different types of measure. It may be easier for field researchers to judge on basis of their impression on the presence and strength of some related behaviour(s) when they evaluate on the QoT indicators. Finally, the ISTOF is distinctive for its Neutral and NA options, because a force-choice design in the QoT might encourage the field researchers to commit to a more definite option and consequently might reduce the number of possible missing values.

Teacher survey and interviews

The teacher survey (Appendix III) was adapted from that in the ECP study with some modifications. For example, the answers for two questions (i.e., *What year group(s) are you currently teaching?* and *What is your current post and responsibility?*) were converted such that they would be applicable to Hong Kong context. This teacher survey aimed to collect background details regarding teaching experience, personal commitment, working hours, additional responsibilities, etc. These details were expected to inform any of their associations that might bear on classroom practice.

Similar to the teacher survey, a semi-structured interview (Appendix IV) based on that for the ECP study was used, but modifications made to adapt to the Hong Kong context were extensive. First, after providing probes consisted of seven areas covering those in the ISTOF, more specific sub-questions were added for some questions, for example, in Question Four and its sub-questions: *What areas do you think effective teaching should cover? And which matters most to you? In which areas do you show more strength? And in which areas do you find more challenges/difficulties?*. Second, some questions had to be modified accordingly to suit the Hong Kong contexts. For example, the policy agendas in the UK such as Literacy *and Numeracy hour/KS3 Strategy* were changed: *Do external policy agendas, such as <u>Chinese as Medium of Instruction</u> and <u>The New Academic Structure for Senior Secondary Education and Higher Education</u>, affect your teaching practices?* Third, a new section (i.e., *Teacher's views on pupils*) was added to see whether their teaching practices were affected by their perceptions of their pupils. Fourth, new questions were added into a section; (e.g., *Teacher Efficacy*) for example: *In your opinion, have you achieved your goals in teaching?* And, *in your opinion, what factors have influenced your teaching most? And how these influences may affect your students?*

3.5.4 Data collection procedures

Lesson observations and ratings by two instruments

The field work elements of the main study was supposed to be conducted in 20 school days, but was unexpectedly interrupted and shortened one day by a typhoon. Given the tight schedule, no alternative date could be arranged to make up for the loss of four lesson observations. As in the ECP study and the pilot, the same classroom observation protocols, ISTOF and QoT, were used. As in the pilot study, these two instruments were immediately completed after the observations rather than on two occasions. Qualitative field notes were also made during the lessons to complement the value-laden ratings obtained using the two instruments.

Given that the applicability of both classroom observation schedules were generally established in the pilot study, no modifications had been made to these measuring instruments. Regarding the problem concerning the suitability of the component *Effective Classroom Layout* of the *QoT* in Hong Kong contexts, it was decided that modifications to enhance the relative ecological validity of individual instrument was not the main purpose of the current thesis. Rather, it would be of interest to compare the two instruments as they are in different cultural contexts. In particular, it was expected that the ISTOF may have some comparative advantages over the QoT as it was developed to be applicable in a broader range of cultural contexts in mind, while the QoT was developed specifically to be used in England and the Netherlands. This would allow a clearer contrast between the ISTOF and the QoT with the former developed for the purpose of applications in international contexts.

For each ISTOF item, teachers were rated on a five-point Likert scale with values ranging from 5 for the highest or 'strongly agree' to 1 for the lowest or 'strongly disagree'. Thus, a higher value would indicate more of the behaviour described by the item would have been observed by the field researcher. There was an 'NA' (i.e., not applicable, unable to observe) response option to indicate a condition in which the item might not be relevant or observable in some classroom settings. However, an 'NA' rating would have to be treated as missing data and mean something different from a neutral rating represented by a value of 3. In the ECP research, a considerable amount of missing data of this type was found, so it was decided that the NA option had to be eliminated instead. For the QoT indicator, ratings were based on a force-choice, four-point Likert scale with values indicating the perceived effectiveness of the teaching behaviours: 1 for 'predominantly weak'; 2 for 'more weaknesses than strengths'; 3 for 'more strengths than weaknesses'; and 4 for 'predominantly strong'.

Qualitative Field Notes

The qualitative field notes produced during classroom observations form a major part of the qualitative data of the Hong Kong study. On average, a 35-minute lesson would yield 200-400 words of notes after they were digitally transcribed into full English texts suitable for coding and analysing in relevant software. Limited by their length, these field notes were by no means comprehensive. As these field notes were snap shots of classroom events, lots of details regarding the classroom processes might have been left out. As the lessons were not filmed as in other studies like (Day, 1998; Tsui, 2003), the researcher had no means to review the accuracy of the observations. However, five principles guided the recording of the classroom processes observed. First, since there has been a continuous debate on quantitative amount of time spent on task, classroom events of teacher behaviour and students' behaviours were time-logged. There was no attempt to specifically record the relative time of teacher-initiated versus pupil-initiated talk, because it was assumed that this would be revealed in the time-logged classroom events.

Second, since previous research evidence (e.g., Biggs, 1988; Biggs & Collis, 1982, 1989; Bowden, 1988; Marton & Saljo, 1976) has showed that interactions between students and between the teacher and students often reflect the depth of learning and the strategies that the teacher may employ to cater for individual differences, they were recorded in detail. Third, as classroom management is generally regarded as the key factor for maintaining a smooth transition of teaching and learning activities, disruptive behaviours of students and teachers' reactions were recorded in detail. Fourth, although the specific content of the lesson was not evaluated in the lesson observations, it was recorded in detail because it would reveal whether the purposes of the lesson were fulfilled or not. This would also inform the *overall judgment of the lesson quality* (as measured by IND100 in QoT) of the observed teacher. Finally, since these notes were expected to supplement the quantitative data, attention was paid to teaching behaviours related to those described in the instruments.

Teacher survey and interviews

All four teachers were given a copy of the teacher questionnaire, but only three were returned. It was decided at the analysis stage that information collected by this questionnaire would not justify a quantitative analysis, but might contribute as data enriching the case studies of individual teachers.

A digitally recorded post-observation semi-structured interview was conducted to explore the in-depth probing of issues relating to the observed teaching session and its purposes, as well as following up factors that may affect their teaching practices such as teacher effectiveness, leadership issues, and teacher efficacy. The interview also provided teachers with an opportunity to reflect on their role which has been reported as a popular approach amongst teachers involved in the CVCP pilot. Except the department head, each teacher was interviewed once, but the length of interview varied because teachers differed in the extent to which they elaborated their answers. One teacher, who was also the department head, was interviewed again to explore his role and work in the department. The principal was also interviewed to find out about his leadership role and his views on issues related to teacher effectiveness. All teachers were allowed to use Chinese or English whenever they felt comfortable, but over 85% of the interviews were conducted in English. However, the interview with the principal was conducted in Cantonese, but was translated and transcribed into English later before the data analysis.

Post-analysis interviews in the follow-up phase

Three unstructured focus group interviews were conducted, one with the teachers and two with two classes of one participant teacher.[37] The main purpose of the interview with the teacher was to discuss some of the main quantitative findings with them. However, issues concerning junior and form teaching, integrating curricula across different forms and levels, what constituents an effective lesson and factors affecting the teaching practices and the *overall judgment of the lesson quality* in the classroom and in the department were brought up again. Some of these issues were formulated into emerging research questions and addressed in subsequent quantitative analyses. The group interviews with students were intended to explore the students' perspectives about English learning and teaching. Some of their views contrasted with the teachers' in the post-observation interviews.

3.5.5 A summary of quantitative and qualitative elements of the design components

Following Maxwell and Loomis' (2003) characterisation, the quantitative and qualitative elements of the design components of current research can be summarised in Table 3.3:

37 As the interviews were not planned in advance, the researcher did not specifically ask to interview the students of that participant teacher. The interviewees were in fact a convenient sample arranged by other teachers not related to the present study.

***Table 3.3:* Quantitative and qualitative elements of the design components in CVCP**

	Quantitative	Qualitative
Purposes	• Precise measurement and comparison of underlying dimensions of observed teaching behaviours; • Establishing relationships between underlying dimensions of observed teaching behaviours	• Meaning of effective teaching • Context of a CMI school • Process of English teaching • Discovering unanticipated classroom events, influences, and conditions in the department and the school • Understanding single cases and cross-case
Conceptual frameworks	*Variance theories* • Dynamic model of educational effectiveness; • Differentiated model of teacher effectiveness; • Probability model of educational effectiveness	*Process theories* • Dynamic model of educational effectiveness; • Differentiated model of teacher effectiveness; • Probability model of educational effectiveness
Research questions	• Variance questions • Presence or absence • Degree • Correlation	• Process questions • Context (holistic) – English lessons; English department; an underachieving CMI school;
Research methods		
Relationship	• Objectivity (researcher as extraneous variable)	
Sampling		• Purposeful sampling – EFL teachers
Data collection	• Prior testing of instruments; • Measurement by two classroom observation schedules;	• Adapting to beginning of the school term • Collection of field notes on classroom events • Interviews with teachers, the department head and the school principal to identify contextual and process variables
Data analysis	• Numerical descriptive analysis; • Inferential analysis: confirmatory factor analysis, multiple regression, ANOVA, discriminant function analysis	• Grounded analysis of qualitative data
Validity		
Internal validity	• Statistical conclusion validity; • Construct validity;	• Descriptive validity; • Interpretative validity; • Causal validity; • Transferability
Generalisability	• External validity (comparability)	

3.6 Ethical considerations

3.6.1 Rapport, researcher bias, reciprocity, and critical friendship in insider research

The advantages and disadvantages of insider research have been well discussed in Elliott (1984), Hockey (1993), Mercer (2007), and Gallais (2008) and these advantages generally have strong ethical implications. The dual roles of the present researcher as an insider (i.e., a previous teacher and the head of two functional departments for ten years of the school) and an outsider (i.e., a researcher from a foreign university conducting a study unfamiliar to the teachers) created a tension between the polarity of “familiarity” and “strangeness” (Gadamer, 1975, p.125). Doing insider research in a school in which the researcher previously worked certainly contributed to the convenience in data collection because it was based on pre-existing rapport. However, my past role in the school and the fact that teachers were recruited by the head of the department added an extra power dimension onto the researcher or experimenter bias (Rosenthal & Fode, 1963) that might have affected the participant’s behaviours and responses in the observation and the interview.

Classroom observation can be a sensitive topic as it is often a crucial part of teacher evaluation and appraisal process and thus generally an unwelcome experience among teachers, if the focus is not developmental and student-focused (Lee at al., 2003). Yet, teachers and schools nowadays need critical friendship (Day, 1998, 1999; James, 2006; James et al., 2006; Stoll & Sammons, 2007). On the one hand, as teachers in the setting had some experience with peer observation, this might have made them more readily open to critical friends and the fact that I was not a peer might make teachers less anxious to behave differently as in peer observation as they did not need to worry about losing face before their colleagues. On the other hand, teachers may not fully comprehend the meaning of a “critical friend” and expect untimely reciprocity. Just as Tsui (2003) also noted that, teachers might sometimes solicit information or suggestions or expect the researcher to give advice during the observation period or in the interview. Thus, the researcher had to avoid suggesting his own perspectives on effective practice to the observed teachers as this would affect the perceptions of the participants. The researcher also tried to keep a low-profile in the school and did not discuss the lessons or his observations with the participants or any other teachers at any time except in the interviews.

3.6.2 Methodology, epistemology and the ethics of interpretation

Another major disadvantage of insider research is concerned with the expected familiarity by the researchers and/or the participants. There are the dangers of “the potential loss of the nurturing and socializing guidance of informants” (Hockey, 1993, p. 203), the overlook of familiar language and key terms, and the taken-for-granted tacit patterns, regularities and assumptions resulting in superficial descriptions in the data analysis (Spradley, 1979; Le Gallais, 2008). As three observed teachers were unknown to the present researcher and there had been many changes in the school since he left, these teachers did not treat him as a fully informed insider and showed willingness to help him to get familiarise with the contexts. Just as Hockey (1993) noted, the boundary between the insider and the outsider is not clear and static, but exists in gradient and negotiation.

Le Gallais (2008) reported her failure to appreciate fully the implications of her managerial status for the research relationship in her study on a mentoring programme with which she was familiar. Familiarity may lead to ‘restricted vision’ and ‘overrapport’ (Le Gallasis, 2008, p.148). As the present researcher was once involved in some of the current practices of the observed school such as setting and co-teaching, an impartial assessment of the impacts of these departmental policies based on previous research findings was particularly important and demanding. Given an uneasy relationship with the principal and a close collegial relationship with the department head in the past, the researcher considered that it was important not to let his past role in the school affect the objectivity of the analyses and interpretations. While Elliot and Likeš (2008, p.115) argued positively about the case study methodology can turn epistemology as ethics because it was based on practical rationality to present the situated reasoning and judgement, rather than “an attempt to provide an epistemologically transcendent account of the representativeness of sampled data.” This situated reasoning may be undermined if it is biased by the pre-existing judgements of the researcher. To facilitate his objectivity, the researcher relied on reflexivity, as Le Galliasis (2008) also recommended, and tried to maintain an outsider stance as a researcher to present an "as is" analysis, rather than a "should be" or "to be" analysis.

3.6.3 Anonymity and confidentiality

To maintain the anonymity and confidentiality of the data collected in the CVCP study was even more crucial as the school might be easily identified if too much contextual background information was to be revealed. In the last meeting with the school principal, he expressed the view that a full description of the school and the teachers would be inappropriate because the school was chosen for its low attainment and disadvantaged context, he was concerned about that the identities of the school and the teachers would be more easily recognised in detail. Thus, at the expense of giving richer descriptions of the cases, full disclosure of the qualitative data and analyses was withheld in order to minimize the risk that the participants might withdraw their consent for that fear.

3.6.4 The myth of unobtrusive observer

The classroom is a very special cultural setting in which the teacher and the students are expected to be the only actors in most of the circumstances in the Chinese culture. The norm is that these actors interact most of the time by themselves without any outsiders. Thus, classroom observations are always "intrusive" in the sense that both the teachers and the students know that they are being silently watched. Classroom observation may be subject to the Hawthorne effect, that is, a form of reactivity in which participants improve a dimension of their behaviour being simply because they are being observed. Both the teacher and students are subject to the Pygmalion effect in classroom observation if they alter their behaviours to meet the observer's expectations. However, just as the observer is watching, s/he is also being watched by the main actors of the scene. It is naive to assume that the behaviours of these actors would not change accordingly in the presence of an outsider, whether s/he is a parent, a teacher of the same school, an inspector, the principal or just a researcher. There is no adequate measure to avoid reactivity in classroom observation, but by observing more lessons of a teacher, a more stable and consistent pattern of observed teaching behaviours is expected to emerge (Rosenshine, 1973).

Although the teachers being observed were expected to understand the purpose of the present research, the researcher could not expect the same from the students even after they were briefed in advance of the observations. Occasionally, he could recognise from the eyes of some students who were disturbing the lessons verbally or physically were soliciting his reactions. These students were seeking attention, but he could not determine to what extent they were attempting more extreme behaviours in order to seek his attention. Disruptive behaviours of students always create tensions that can be intensified when they are deliberately aiming to seek attention from both the teacher and the observer as an outsider. In this case, a researcher is not seen by the teacher and the students as a silent observer at the back of the scene, but an insider and an actor that is capable of acting in response to their actions. An unobtrusive observer is more likely to be a myth in an insider research or classroom observation research.

3.7 Summary

In this chapter, the epistemological and methodological assumptions of the research have been reviewed. Such assumptions can be linked to pragmatism for its adoption of an MM approach, which has been shown in previous research a constructive strategy to integrate qualitative and quantitative methods for exploring complex issues that may not be adequately addressed if only either method is attempted. The multilevel case study design reflects the hierarchical nature of education system and the recognition of a Chinese model of teacher effectiveness that emphasises interactions of personal and professional qualities of the teacher and the contextual variables existed at the various levels of the education system.

CHAPTER 4 : DIMENSIONS OF AND CONSISTENCY AND VARIATION IN CLASSROOM PRACTICES USING ISTOF AS THE OBSERVATION INSTRUMENT

4.1 Introduction

The results of systematic classroom observation presented in this chapter are based on ISTOF, a high-inference instrument developed by international academies (Teddlie et al., 2006; see Section 3.5.3 for detail). Following the ECP study discussed in last chapter, two instruments were also employed, but unlike that study, both instruments were used in *every lesson* observed. The two instruments were designed for different purposes, ISTOF as an international instrument suitable for studying teacher and school effectiveness in a variety of contexts, while QoT as an instrument for classroom observation for professional inspectors of different countries. It was assumed that despite these different intentions, the two instruments are comparable and using them concurrently would allow us to explore *whether similar dimensions of teaching behaviours can be identified as effective classroom practices*. In addition, a comparison of these two instruments would contribute to our knowledge on systematic classroom observation because, as discussed in Chapter 2, such instruments have rarely been compared directly and studied in the past.

However, this chapter is solely devoted to results obtained using ISTOF as the observation instrument. The results of the second instrument, QoT, are to be presented in the next chapter. Separating discussions for these two instruments is justified as this would permit better understanding of the unique strengths and constraints of each instrument. While instrument comparison and variation will be addressed more fully later in Chapter 6, results presented here do highlight their similarities as the discussion proceeds. Studying extensively 76 lessons of four teachers means that the unit of analysis is the *lesson*, rather than the *teacher*. This also means that variation found in the results presented here did not necessarily occur across teachers, but only reflected variation *across lessons*. Within-teacher and between-teacher variations are to be explored further in the case analyses in Chapters7 and the cross-case analyses in Chapter 8.

The following sections summarise five sets of results obtained using ISTOF as the instrument for systematic observation. First, descriptive statistics of the ISTOF items are discussed in Section 4.2 with a focus on the general patterns emerged in those items with highest means, lowest means and their frequency distributions. Comparisons of these patterns with those in the ECP study discussed in the last chapter seem to reflect differences between a convenient sample of teachers and a purposive sample of *more effective* teacher. In particular, English teachers in the ECP sample demonstrated more effective teaching behaviours than the Hong Kong teachers in most lessons (as defined by the ISTOF theoretical components intended to tap effective practices). Alternatively speaking, most of the effective classroom practices were *less often* observed or found with *less strength* in the lessons of the Hong Kong sample than in the English sample.

Second, exploratory and subsequent confirmatory analyses generated underlying factors found in the ratings obtained using ISTOF as the observation instrument. These results are presented respectively in Sections 4.3 and 4.4. As the main aim of the current study was to explore whether there were any generic characteristics of effective teaching practices in these two samples, exploratory and confirmatory analyses would allow us to explore the underlying factors that can characterise the effective classroom practices observed in the lessons of the EFL teachers in Hong Kong. These characteristics would shed light on variables that are "basic, generic and replicable" in a variety of settings that can serve as the basis for cross-country comparisons (Teddlie et al., 2006, p.565) and provide a crucial support to a generic model of teaching or educational effectiveness. The results revealed six underlying factors for the ISTOF item-based model.

Third, the frequency distributions of the underlying factors of the CFA model found for the Hong Kong sample on ISTOF data provide evidence for differentiated teacher effectiveness. These results presented in Section 4.5 showed in individual lessons how the four Hong Kong teachers' behaviours varied in different dimensions of classroom practices, as they were rated better and more often in some but not in the others and thus might reflect where their strengths and weaknesses lied. The magnitudes of variation in these underlying factors across lessons contribute to our understanding of differential teaching effectiveness of the Hong Kong sample. In general, Hong Kong teachers seemed to show strengths in classroom management, presentation, and maintenance of lesson focus more often, but less strength in meta-cognitive skills teaching.

Fourth, Section 4.6 presents and discusses the extent to which the resulted six-factor CFA model could be comparable to the hypothetical factor structure with seven theoretical components. A CFA model using all theoretical components is put forward as a reference for comparison. It is argued that the six-factor and seven-factor models seemed to reveal some classroom practices are fundamental, effective practices and some of the discrepancies between the two models might not be explained purely by idiosyncratic characteristics of the Hong Kong sample.

Finally, a factor by factor comparison between CFA models in the CVCP and the ECP studies in Section 4.7 was performed. Similarities between the factors of the two models suggested that there were some generic characteristics of effective teaching practices in these two samples like emphasis of learning objective and focus, student engagement, classroom management, and supportive teaching strategies. In particular, the first two features seemed to emerge as important dimensions of effective classroom practices which were not anticipated as independent and distinctive constructs in the original scale.

4.2 Features of strengths and weaknesses in Hong Kong teachers' observed classroom practices and comparisons with the ECP results

Features of strengths and weaknesses in the lessons of the four Hong Kong teachers' observed classroom practices can be see n in an excerpt of the descriptive statistics in Table 4.1 (for a full table, see Appendix V).

Table 4.1: **Features and ranking (descending) of strengths and weaknesses in Hong Kong teachers' observed classroom practices in the CVCP study (N=76 lessons) using the ISTOF instrument**

Item No.	Item Description (T = The teacher; S= student)	Mean	Std. Deviation	Skew-ness	z-skewness	Kurtosis	z-kurtosis
The first ten Items with the highest means in the CVCP study							
Item 10	*T communicates in a clear and understandable manner.*	*4.33*	.89	-1.06	*3.85**	.04	0.08
Item 1	T makes explicitly clear why an answer is correct or not.	*4.24*	.81	-.92	*3.34**	.41	0.74
Item 40	Teacher makes sure that Ss are involved in learning activities until the end of the lesson.	*4.24*	1.00	-1.23	*4.45**	.73	1.35
Item 3	*Assignments given by T are clearly related to what Ss learned.*	*4.21*	.84	-.98	*3.55**	.55	1.01
Item 2	T provides appropriate feedback to the answers given by the Ss.	*4.13*	.84	-.53	1.94	-.65	1.19
Item 9	*T regularly checks for understanding.*	*4.07*	.90	-1.04	*3.76**	1.16	*2.12**
Item 42	There is clarity about when and how Ss can get help to do their work in class.	*4.03*	1.23	-.93	*3.37**	-.38	0.69
Item 39	*Teacher starts lesson on time.*	*4.00*	1.06	-.97	*3.52**	.61	1.11
Item 41	Actions are taken to minimize disruption.	3.96	1.24	-.92	*3.34**	-.40	0.74
Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	3.91	1.31	-1.00	*3.63**	-.21	0.39
tems with a mean below 3.0 in the CVCP study							
Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	2.97	1.18	.15	0.55	-.89	1.63
Item 24	T encourages Ss to ask one another questions and to explain their understanding of topics to one other.	2.96	1.32	.11	0.40	-1.19	*2.18**
Item 23	T explicitly provides instruction in problem-solving strategies.	2.95	.86	-.15	0.56	-.47	0.87
Item 30	Ss are invited to give their own examples.	2.91	1.19	.13	0.48	-.89	1.63
Item 7	T makes a distinction in the scope of the assignments for different groups of Ss.	2.84	1.07	.19	0.69	-.34	0.63
Item 21	T invites Ss to use strategies which can help them solve different types of problems.	2.83	1.04	-.38	1.39	-.74	1.36
Item 20	*T uses different, appropriate instructional strategies for different groups of Ss.*	2.82	.96	.11	0.38	-.24	0.45
Item 27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	2.80	1.14	.01	0.05	-1.14	*2.08**
Item 26	T motivates the Ss to think about the advantages and disadvantages of certain approaches.	2.79	1.10	-.06	0.23	-1.06	1.95
Item 22	T invites Ss to explain the different steps of the problem solving strategy which they are using.	2.63	.91	-.18	0.67	-.70	1.28
Item 28	T invites the Ss to give their personal opinion on certain issues.	2.62	1.13	.29	1.07	-.71	1.31

Of the ten items with the highest means between 3.91 and 4.33, only 8 items (out of 45, about 17.8%) have a mean over 4.0. High scores indicate a greater incidence of effective behaviour observed of the items. These are items of *Indicator 3.1: The teacher show good communication skills* and of two components, *Assessment and Evaluation* and *Classroom Management*. Items with higher means also tended to be negatively skewed. Nine out of these ten items have a negative skew that reaches a significant level[38]. A negative skew in these items indicates that the effective behaviours as

38 The value z-skewness was obtained by dividing the skewness value by its standard error. When this z-score of the skewness value (|ske| /s.e.) is above 1.96 it is statistically significant, as indicated by an asterisk in Table 4.1.

described in these items were frequently observed in most lessons. Significantly positive kurtosis (or leptokurtic distribution) of Item 9 (i.e., the z-score of its kurtosis value (|*β2*| /s.e.) reaches 1.96 or above, indicated by an asterisk in Table 4.1) also indicates *less* variability in ratings across lessons on this item. By contrast, there are 11 items (or 24.4%) with a mean below 3.0, which are mainly items of *Promoting active learning and developing meta-cognitive skills* (Component Five). This indicates that overall these dimensions were less commonly observed than many other features in lessons. Two of these items showed variability greater than usual as indicated by their statistically significant negative kurtosis or platykurtic distributions.

Characteristics of these descriptive statistics shown in Table 4.2 reveal four interesting patterns when they are compared with those found for the ISTOF ratings in the ECP study.

Table 4.2: **Features and ranking (descending) of strengths and weaknesses in English teachers' observed classroom practices in the CVCP study using the ISTOF instrument in the ECP study (N=79 lessons/teachers[39])**

The first ten Items with a mean be above 4.0 in the ECP Study								
Item No.	**Item Description (T = The teacher; S= student)**	**Mean**	**Std. Deviation**	**Skew-ness**	**z-skew-ness**	**Kurtosis**	**z-kurtosis**
Item 10	*T communicates in a clear and understandable manner*	4.86	0.38	-2.80	*0.35**	7.74	*14.47**
Item 3	*Assignments given are clearly related to what students learned*	4.78	0.47	-2.11	*7.77**	3.89	*7.22**
Item 32	T shows respect for the students in both his/her behaviour and use of language	4.77	0.60	-3.94	*4.55**	20.4	*38.03**
Item 34	T's instruction is interactive	4.74	0.59	-2.97	*0.93**	10.47	*19.45**
Item 39	*T starts lesson on time*	4.71	0.56	-2.27	*8.19**	6.54	*11.94**
Item 31	T demonstrates genuine warmth and empathy toward all students in the classroom	4.70	0.56	-1.72	*6.34**	2.04	*3.81**
Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts	4.62	0.59	-1.69	*6.13**	3.94	*7.22**
Item 14	T implements the lesson smoothly moving from one stage to another with well-managed transition points	4.57	0.75	-1.96	*7.24**	3.73	*6.97**
Item 17	T poses questions which encourage thinking and elicit feedback	4.57	0.71	-1.79	*6.63**	3.15	*5.90**
Item 9	*T regularly checks for understanding*	4.56	0.57	-1.28	*4.75**	3.05	*5.70**
Items with a mean below 3.0 in the ECP Study							
Item 12	T asks students to identify the reasons why specific activities take place in the lesson	2.75	1.26	0.48	1.58	-1.06	1.75
Item 20	*T uses different, appropriate instructional strategies for different groups of students*	2.73	1.42	0.50	1.61	-1.20	*1.96**

First, there was a clear inflation of ratings in most items in the ECP study. For example, there were 33 items (about 73.3%) in the ECP study in

39 It should be noted that the unit of analysis in the ECP study was one lesson per observation instrument for every teacher in the sample, but the unit of analysis here is the lesson.

England with a mean above 4.0, while only 2 items (4.4%) had a mean below 3.0. This means that the English lessons in that study were *rated highly on more teaching behaviours* and *more often* than the Hong Kong lessons in the CVCP study. The partial statistics of ISTOF items in the ECP study in Table 4.2 below shows that the ten items with the highest means had a mean ranging from 4.56 to 4.86, which were about on average 12.2% to 16.6% higher than the top ten items with the highest means in the CVCP. The number of items with a mean below 3 in the Hong Kong sample was also much larger than that in the English sample (11 vs 2, or 24.4% vs 4.4%). These results suggest that the Hong Kong lessons in the current study did not show effective teaching behaviours as many as and as often as the English lessons in the ECP study.

Second, there is an interesting contrast between the emphasis on classroom management in the Hong Kong sample as shown in the presence of Items 40, 41, 42, and 44 in the ten highest means list and the emphasis on classroom climate in the English sample as shown in the presence of Items 31, 32, and 34. Classroom management is likely to be especially relevant for effective teaching in Hong Kong because the number of students in a class, which may go up to 42, is relatively larger than that in England. English lessons in the ECP sample also showed better in lesson structure and promoting thinking and eliciting feedback (Items 13, 14, and 17), while Hong Kong lessons seemed to be stronger at explicit, detailed and constructive feedback (Items 1 and 2).

Third, the striking difference, however, lies at the degree of *negative* skewness and the degree of *positive* kurtosis or leptokurtic distribution between them. As statistically significant negative skewness and positive kurtosis (marked with an asterisk in Table 4.2 above) indicate teacher behaviours frequently received positive and high ratings, consistent with the fact that the English sample were chosen to represent lessons of more effective teachers in England, while the Hong Kong sample were not selected to be lessons of more effective, but reflected a naturally occurring group of teachers of a less effective department of an underperforming school (see Section 8.3 for details). Highly negative skewness and positive kurtosis often mean stronger consistency in desirable teaching behaviours. Smaller variations in teaching behaviours with higher ratings among effective English lessons are noted in the smaller standard deviations in the top ten items in comparison with those of the Hong Kong lessons.

Fourth, despite the differences noted above, the current results also suggest that some effective teaching practices were consistently appeared more often across two samples. Items appeared in both studies in the top ten highest means or with a mean below 3 are highlighted in *italics* in Table 4.1 and Table 4.2 above. Four items which appeared in the top ten list of highest means in both studies include *Item 10: The teacher communicates in a clear and understandable manner*, *Item 3: Assignments are given are clearly related to what students learned, Item 39: Students start on time,* and *Item 9: The teacher regularly checks for understanding*. *Item 20: The teacher uses different, appropriate instructional strategies for different groups of students* also appeared in both sample with a mean lower than 3.0. Although only Item 20 was found with a mean below in both studies, similarities of teaching practices that received lowest ratings in both studies might be closer than it appeared. If those items received a non-applicable rating in the ECP study were to be recoded differently and negatively, then more items of Component Five would have a lower mean as it was the case in the CVCP study.

The partial frequency table in Table 4.3 below (for a full table, see Appendix VI) illustrates the variations of individual effective classroom practices across lessons. In general, the frequency distribution patterns of ISTOF items in this table show that the four Hong Kong teachers were rated *more often higher* for their teaching behaviours specified in the ten items of highest means (i.e., a higher percentage of ratings in the *Moderately Agree* and *Strongly Agree* categories). These items are negatively skewed and nine of them have negative skewness at a significant level (see Table 4.1). Again, this would suggest in most lessons, teachers consistently showed more strength in these areas. In contrast, items with lowest means tended to have a higher percentage of ratings in the *Moderately Disagree* and *Strongly Disagree* categories.

Table 4.3: **Variability identified in the distribution of strengths and weaknesses in Hong Kong teachers' ranked observed classroom practices in the CVCP study (N=76) using the ISTOF instrument**

Item No.	Item Description (T = The teacher; S= student)	Strongly Disagree	Moderately Disagree	Neutral	Moderately Agree	Strongly Agree
Items with a mean above 4.0 in the CVCP study						
Item 10	T communicates in a clear and understandable manner.	0.00%	3.90%	15.80%	*23.70%*	***56.60%***
Item 1	T makes explicitly clear why an answer is correct or not.	0.00%	3.90%	11.80%	*40.80%*	***43.40%***
Item 40	Teacher makes sure that Ss are involved in learning activities until the end of the lesson.	1.30%	6.60%	13.20%	*25.00%*	***53.90%***
Item 3	Assignments given by T are clearly related to what Ss learned.	0.00%	5.30%	10.50%	*42.10%*	***42.10%***
Item 2	T provides appropriate feedback to the answers given by the Ss.	0.00%	2.60%	21.10%	*36.80%*	***39.50%***
Item 9	T regularly checks for understanding.	1.30%	5.30%	13.20%	***46.10%***	*34.20%*
Item 42	There is clarity about when and how Ss can get help to do their work in class.	3.90%	10.50%	18.40%	*13.20%*	***53.90%***
Item 39	Teacher starts lesson on time.	3.90%	2.60%	23.70%	*28.90%*	***40.80%***
Item 41	Actions are taken to minimize disruption.	3.90%	14.50%	10.50%	*23.70%*	***47.40%***
Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	7.90%	10.50%	10.50%	*25.00%*	***46.10%***
Items with a mean below 3.0 in the CVCP study						
Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	9.20%	***30.30%***	*26.30%*	22.40%	11.80%
Item 24	T encourages Ss to ask one another questions and to explain their understanding of topics to one other.	14.50%	***28.90%***	*18.40%*	22.40%	15.80%
Item 23	T explicitly provides instruction in problem-solving strategies.	3.90%	***26.30%***	*42.10%*	26.30%	1.30%
Item 30	Ss are invited to give their own examples.	11.80%	***28.90%***	*26.30%*	22.40%	10.50%
Item 7	T makes a distinction in the scope of the assignments for different groups of Ss.	10.50%	***26.30%***	*39.50%*	15.80%	7.90%
Item 21	T invites Ss to use strategies which can help them solve different types of problems.	14.50%	***18.40%***	*38.20%*	27.60%	1.30%
Item 20	T uses different, appropriate instructional strategies for different groups of Ss.	7.90%	***28.90%***	*40.80%*	18.40%	3.90%
Item 27	T asks the Ss to reflect on the solutions/ answers they gave to problems or questions.	13.20%	***32.90%***	*18.40%*	31.60%	3.90%
Item 26	T motivates the Ss to think about the advantages and disadvantages of certain approaches.	13.20%	***30.30%***	*23.70%*	30.30%	2.60%
Item 22	T invites Ss to explain the different steps of the problem solving strategy which they are using.	11.80%	***30.30%***	*40.80%*	17.10%	0.00%
Item 28	T invites the Ss to give their personal opinion on certain issues.	17.10%	***32.90%***	*26.30%*	18.40%	5.30%

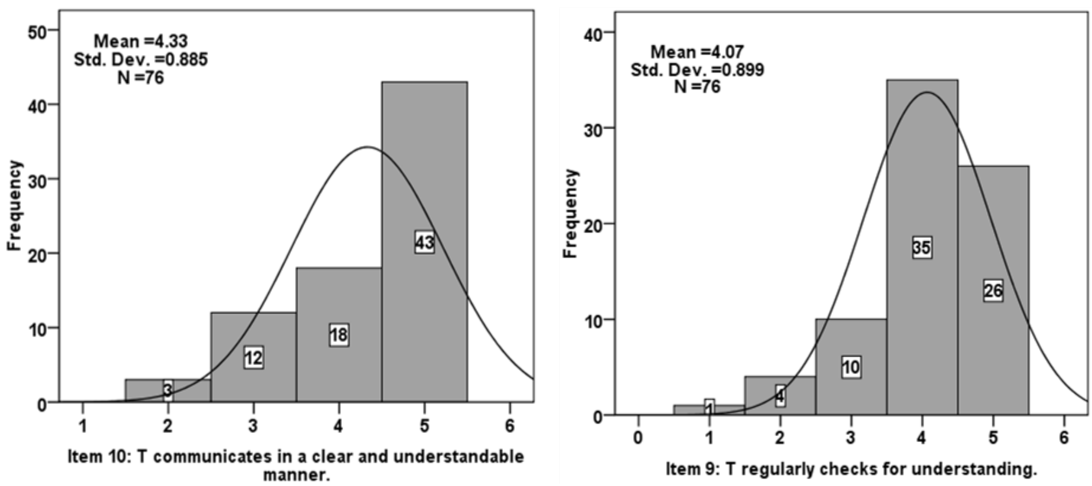
By contrast, Table 4.4 clearly shows that the strengths of the English sample are relatively stronger than the Hong Kong sample as there were a higher percentage of lessons receiving highest ratings in the ten items with the highest means. However, the score distributions Item 12 and Item 20 also shows that more teachers were rated negatively in the English sample than in the Hong Kong sample for the items with a mean below 3.0, suggesting that the English sample were not necessarily rated better than the Hong Kong sample in areas where their weaknesses lied.

***Table 4.4*: Variability identified in the distribution of strengths and weaknesses in English teachers' ranked observed classroom practices in the ECP study (N=79) using the ISTOF instrument**

Items with a mean above 4.0 in the ECP study						
Item No.	**Item Description (T = The teacher; S= student)**	**Strongly Disagree**	**Moderately Disagree**	**Neutral**	**Moderately Agree**	**Strongly Agree**
Item 10	T communicates in a clear and understandable manner	0.0%	0.0%	1.3%	*11.4%*	***87.3%***
Item 3	Assignments given are clearly related to what students learned	0.0%	0.0%	2.6%	*16.7%*	***80.8%***
Item 32	T shows respect for the students in both his/her behaviour and use of language	1.3%	0.0%	1.3%	*15.2%*	***82.3%***
Item 34	T's instruction is interactive	0.0%	2.6%	0.0%	*17.9%*	***79.5%***
Item 39	T starts lesson on time	0.0%	1.3%	1.3%	*22.7%*	***74.7%***
Item 31	T demonstrates genuine warmth and empathy toward all students in the classroom	0.0%	0.0%	5.1%	*20.3%*	***74.7%***
Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts	0.0%	1.3%	1.3%	*31.6%*	***65.8%***
Item 14	T implements the lesson smoothly from one stage to another with well-managed transition points	0.0%	3.8%	3.8%	*24.1%*	***68.4%***
Item 17	T poses questions which encourage thinking and elicit feedback	0.0%	2.5%	5.1%	*25.3%*	***67.1%***
Item 9	T regularly checks for understanding	0.0%	1.3%	0.0%	*40.5%*	***58.2%***
Items with a mean below 3.0 in the ECP study						
Item 12	T asks students to identify the reasons why specific activities take place in the lesson	11.5%	***47.5%***	6.6%	23.0%	11.5%
Item 20	*T uses different, appropriate instructional strategies for different groups of students*	18.6%	***42.4%***	5.1%	15.3%	18.6%

Frequency distributions of individual items often reveal great variation in variability in different observed teachers' teaching behaviour. The frequency distributions shown below highlight some typical examples. For example, Figure 4.1 below shows the frequency distributions of Item 10, which has statistically significant negative skewness, and of Item 9, whose negative skewness and positive kurtosis are *both* statistically significant.

***Figure 4.1:* Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of two negatively skewed items (Items 10 & 9)**

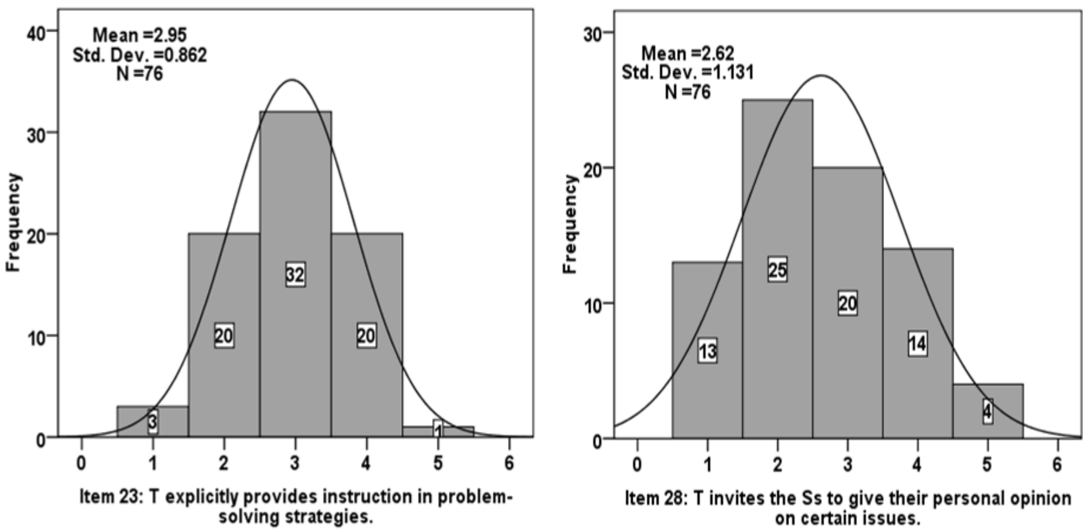


Skewness = -1.06; Kurtosis = 0.04 **Skewness = -1.04; Kurtosis = 1.16**

In contrast, although the eleven items with the lowest means suggest that teachers tended to have lower ratings more often and higher ratings less often, the degrees of skewness and kurtosis were not always predictable in the same pattern across items. For example, while seven of those eleven items are slightly positive skewed, four items are negatively skewed.

For example, Figure 4.2 below shows Item 23 with a nearly normal distribution and Item 28 with a positively skewed distribution. Generally, a positive skewed distribution means that lessons were rated negatively on that item more often, while negative skewed distribution suggests the opposite.

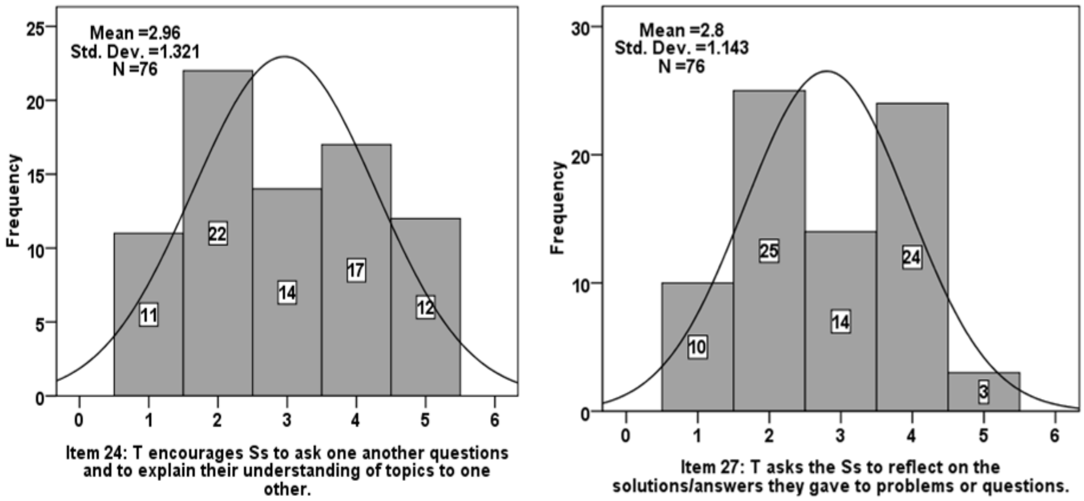
Figure 4.2: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of two items with lowest means (Items 23 & 28)**



Skewness = -0.15; Kurtosis = -0.47   Skewness = 0.29; Kurtosis = -0.71

Although there were eleven items with bimodal distributions in the ECP study, they were rarer in the current study: only occurred in Items 24, 26, 27, 32 and 43. Figure 4.3 below shows items with a mean below 3 and a bimodal distribution. Item 24 has a marginal bimodal distribution with a less prominent peak, while Item 27 has two clear peaks.

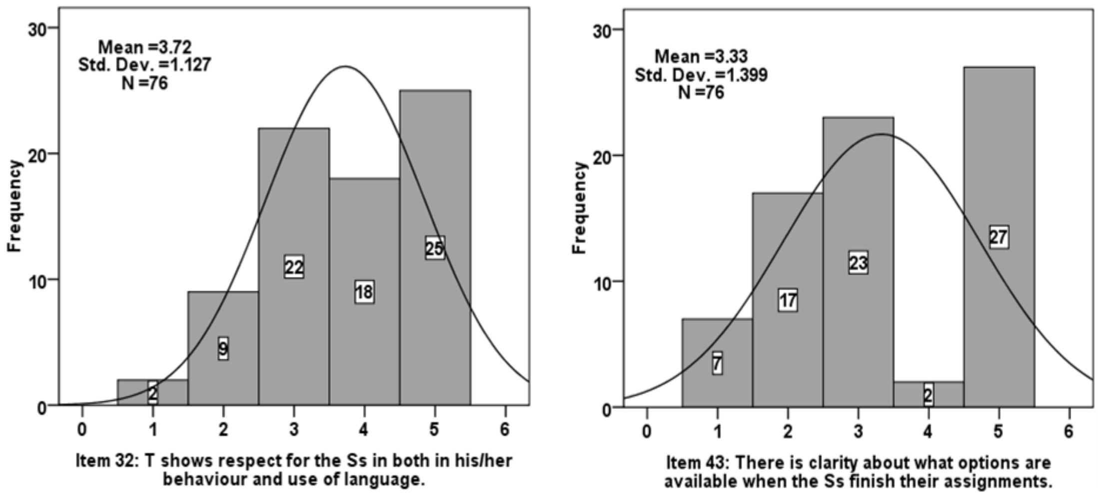
Figure 4.3: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of two items with means below 3 and bimodal distributions (Items 24 & 27)**



Skewness = 0.11; Kurtosis = -1.19 Skewness = 0.01; Kurtosis = -1.14

Figure 4.4 below shows items with a mean above 3 and a bimodal distribution. Item 32 has a slightly negative skewed bimodal distribution, while Item 43 shows two clear separate distributions, indicating lessons either receiving the highest ratings or neutral or lower ratings.

Figure 4.4: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of two items with means above 3 and bimodal distributions (Items 32 & 43)**



Skewness = -0.41; Kurtosis = -0.78 Skewness = -0.01; Kurtosis = -1.35

4.3 Preliminary dimensions of effective classroom practices identified in a seven-factor item-based model: the results of the exploratory factor analysis

All the 45 items originally in the ISTOF scale were used in the initial exploratory factor analysis (hereafter EFA). As there were no missing data in the current data set, as it was happened in the ECP study (Ko & Sammons, 2008a), no imputation was required. The EFA performed using the principal component analysis extraction with varimax rotation specified in DATA REDUCTION in SPSS16 yielded seven factors. As shown in Table 4.5, the total variances explained by the factors are high at 82.7%, probably because there were no missing values in the data and no significantly skewed or kurtotic frequency distribution found in any item.

Table 4.5: **Relative importance of EFA factors of ISTOF items as identified in the variances they accounted for in the CVCP study (N=76)**

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	26.97	59.94	59.94	26.97	59.94	59.94	7.56	16.80	16.80
2	2.65	5.88	65.82	2.65	5.88	65.82	7.35	16.33	33.14
3	2.39	5.32	71.14	2.39	5.32	71.14	6.40	14.21	47.35
4	1.58	3.50	74.64	1.58	3.50	74.64	5.52	12.27	59.61
5	1.33	2.97	77.60	1.33	2.97	77.60	5.27	11.71	71.32
6	1.18	2.62	80.22	1.18	2.62	80.22	3.28	7.29	***78.61***
7	1.12	2.50	82.72	1.12	2.50	82.72	1.85	4.11	***82.72***

The item-based EFA factors in Table 4.6 below are well defined by the items in the factors. The loadings of items of factors and their communalities are high. The overall factor loadings of the 45 items are mostly high. Only 4 items have satisfactory loadings (λ between 0.40 and 0.5), while almost three-fourth of the loadings range from moderate to high (λ between 0.60 and 0.85). All the communalities are high (with h^2 0.7 between 0.72 and 0.92).[40]

Although the last factor has only two items (i.e., Item 11 and Item 12), both have a high communality (h^2 = 0.72 and 0.87, respectively) indicating its distinctiveness. Certainly, deleting this seventh factor would not weaken the total explanatory power as the cumulative percentage of variances explained for the sixth factor is still high above 78%. However, combining the sixth and seventh EFA factors as one in the rotated component matrix in Table 4.6 was supported as they were similar in meanings and there were too few items for the seventh factor to be reliably explored further in a confirmatory factor analysis. This also helped to reduce the number of parameters needed to

40 High communalities are particularly important for the present data as (Preacher, 2002). Preacher and MacCallum (2002) found that high communalities are crucial for models built on small sample size. See also the general discussion in later section. Mundfrom, Shaw, & Ke (005) found similar results but also noted that when the variables-to-factors ratio exceeds 6, the minimum sample size begins to stabilize regardless of the number of factors or the level of communality.

perform confirmatory factor analysis. Items that were retained after the CFA are highlighted in colours in Table 4.6.

Table 4.6: **Preliminary dimensions of effective classroom practices identified in the EFA factors of ISTOF items and their loadings (above 0.4) in the CVCP study (N=76)**

Factor Name	tem No	Item Description (T = The teacher; S= student)	Component 1	2	3	4	5	6	7
Meta-Cognitive Skills Teaching (MetaCogn)	**Item 22**	T invites Ss to explain the different steps of the problem solving strategy which they are using.	.85						
	Item 21	T invites Ss to use strategies which can help them solve different types of problems.	.83						
	Item 23	T explicitly provides instruction in problem-solving strategies.	.81						
	Item 28	T invites the Ss to give their personal opinion on certain issues.	.76						
	Item 27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	.74						
	Item 26	T motivates the Ss to think about the advantages and disadvantages of certain approaches.	.73			.41			
	Item 24	T encourages Ss to ask one another questions and to explain their understanding of topics to one other.	.52				.43		
	Item 20	T uses different, appropriate instructional strategies for different groups of Ss.	.52		.52				
	Item 17	T poses questions which encourage thinking and elicit feedback.	.49			.47			
Classroom Management and Climate (MgtClima)	**Item 41**	Actions are taken to minimize disruption.		.79					
	Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.		.78					
	Item 45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	.41	.73					
	Item 40	Teacher makes sure that Ss are involved in learning activities until the end of the lesson.		.73					
	Item 42	There is clarity about when and how Ss can get help to do their work in class.		.69	.41				
	Item 39	Teacher starts lesson on time.		.68					
	Item 31	T demonstrates genuine warmth and empathy toward all Ss in the classroom.		.65		.			
	Item 32	T shows respect for the Ss in both in his/her behaviour and use of language.		.61		.46			
	Item 38	T makes clear that all Ss know that he/she expects their best efforts in the classroom.		.59				.37	
Differentiation and Support (DiffSupp)	**Item 8**	T gives additional opportunities for practice to Ss who need them.			.78				
	Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.			.73				
	Item 7	T makes a distinction in the scope of the assignments for different groups of Ss.			.70				
	Item 43	There is clarity about what options are available when the Ss finish their assignments.			.68				
	Item 15	T provides sufficient wait time and response strategies to involve all types of learners.			.57				
	Item 25	T gives Ss the opportunity to correct their own work			.51				
	Item 19	T uses a variety of instructional strategies during the class period			.44		.42	.43	
Clarity and Logic of Presentation (PrntClar)	**Item 14**	T implements the lesson smoothly moving from one stage to another with well-managed transition points.				.68			
	Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts.				.67			
	Item 1	T makes explicitly clear why an answer is correct or not.				.66			
	Item 2	T provides appropriate feedback to the answers given by the Ss.				.65			
	Item 10	T communicates in a clear and understandable manner.		.56		.56			
	Item 9	T regularly checks for understanding.			.45	.51			
	Item 18	The length of the pause following questions varies according to the difficulty level of questions	.41			.46		.42	

Factor Name	tem No	Item Description (T = The teacher; S= student)	Component						
			1	2	3	4	5	6	7
Student Engagement (EngagStd)	**Item 5**	Ss communicate frequently with one another on task-oriented issues.					.70		
	Item 36	T seeks to engage all Ss in classroom activities.			.51		.68		
	Item 30	Ss are invited to give their own examples.	.50				.64		
	Item 29	T systematically uses material and examples from the Ss' daily life to illustrate the course content.	.43				.64		
	Item 33	T creates purposeful activities that engage every S in productive work.			.42		.57		
	Item 6	All Ss are actively engaged in learning.		.42	.52		.56		
	Item 16	T gives assignments that stimulate all Ss to active involvement.			.41	.46	.54		
	Item 34	T's instruction is interactive.			.43		.48		
Strategies to Enhance Learning and Lesson Focus (LrnGoal)	**Item 4**	T explains how assignments are aligned to the learning goals of the lesson.						.72	
	Item 37	T praises children for effort towards realizing their potential.		.43				.64	
	Item 3	Assignments given by T are clearly related to what Ss learned.				.42		.59	
	Item 12	T asks Ss to identify the reasons why specific activities take place in the lesson.							.79
	Item 11	T clarifies the lesson objectives at the start of the lesson.						.54	.60

4.4 Six dimensions of effective classroom practices identified in an item-based model in the results of the confirmatory factor analysis

Based on the EFA model above, a confirmatory factor analysis was conducted using the Maximum Likelihood estimation in LISREL 8.72. As shown in Figure 4.5 below, a CFA model of six latent variables (hereafter CVCP (ISTOF) Model) with 30 items was obtained[41]. The fifteen deleted items generally have lower factor EFA loadings or are cross-loaded in two or more factors in the original EFA model in Table 4.1. Deletions of these items were supported by modification indices and the general requirement to keep the number of parameters not exceeding the sample size[42].

---

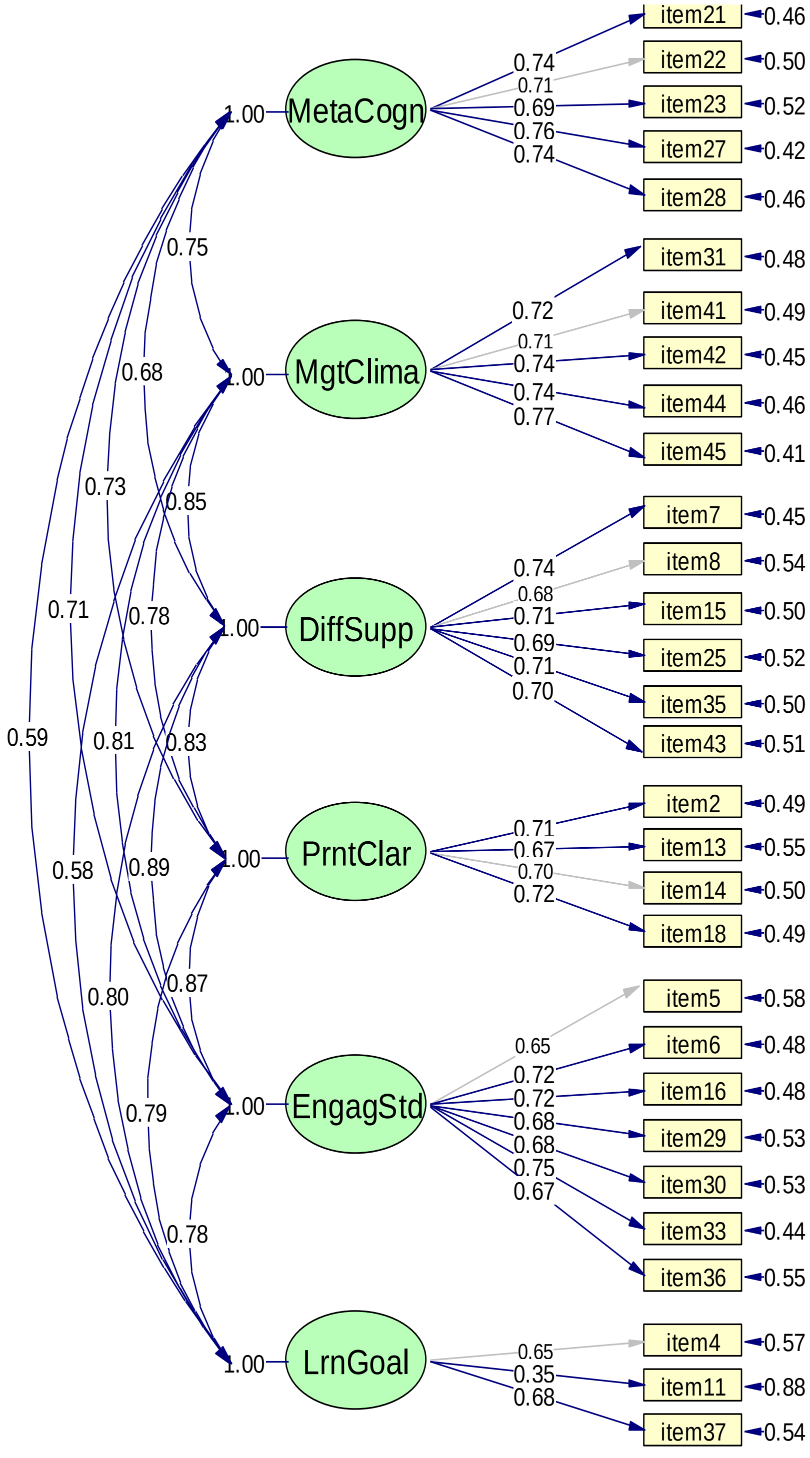
41 The solution reported here does not represent the "best" model that can be obtained from the data. The "best" solution would be data-driven by running CFA for individual EFA factors one at a time and then building a full model based on those CFA results. The results are presented in Appendix VII. Like many other researchers, Brown (2006) argued that model revisions purely driven by data were not always desirable as the" best" model might be too rigid and represent idiosyncratic characteristics of the sample rather than the population. It was not the intent here to arrive at the "best" solution, nor a model that can approximate the population because the sample size was too small to justify any strong claim about the population.

42 A hypothesized model based on the seven EFA factors and 45 items would result in biased estimation and statistic because such a model has more parameters (i.e., t=111; 45 factor loadings, 45 measurement error variances, and 21 correlations) than the sample cases (N=76). MacCallum, Browne & Sugawara (1996) showed that although the minimum requirement for parameter estimation was not invariant across studies, sample size should be larger than or at least equal to the number of observed variables ($N \geq t$). Thus, in order to perform CFA with adequate parameter estimation, my hypothesized model had to compromise on the number of items as well as the number of factors.

However, Marsh and Hau (1999) reported that breaking this critical barrier might not have a devastating effect as they had a model with t=75 and N=50 as the number of indicators per factor ratio (p/f) was 12. They suggested that high p/f had compensated the biased effects of small sample size on parameter estimates and the success rate of model convergence. The present results seemed to confirm their observation.

Preacher and MacCallum (2002) found that a higher number of factors was crucial for successful factor recovery for EFA, but MacCallum et al. (1996) also found that the minimum sample size required for achieving a given level of power for any test of fit was inversely related to the sizes of degree of freedom. This means that to a certain extent there may be a trade-off between maximizing parameter estimation and maximizing power for test by increasing the higher number of factors or a number of observed variables.

Figure 4.5: **The underlying dimensions of effective classroom practices as indentified in a six-factor CVCP (ISTOF) Model (N=76) with standardised coefficients**



KEYS:

MetaCogn:
Meta-cognitive skills teaching
(Average variance extracted = 53.06%; Composite reliability= 0.85; Cronbach's alpha=0.94)

MgtClima: Classroom management and climate
(Average variance extracted = 54.21%; Composite reliability= 0.86; Cronbach's alpha=0.95)

DiffSupp:
Differentiation and support
(Average variance extracted = 49.74%; Composite reliability= 0.86; Cronbach's alpha=0.94)

PrntClar:
Clarity and logic of presentation
(Average variance extracted = 49.04%; Composite reliability= 0.79; Cronbach's alpha=0.91)

EngagStd:
Student engagement
(Average variance extracted = 48.51%; Composite reliability= 0.87; Cronbach's alpha=0.94)

LrnGoal:
Strategies to Enhance Learning and Lesson Focus
(Average variance extracted = 33.58%; Composite reliability= 0.59; Cronbach's alpha=0.70)

The whole scale:
(Average variance extracted = 49.04%; Composite reliability= 0.97; Cronbach's alpha=0.97)

Chi-Square=150.30, df=390, P-value=1.000,RMSEA=0.000

Goodness-of-fit was evaluated using multiple fit indices including chi square (χ^2), root mean square error of approximation (RMSEA), normed fit index (NFI), comparative fit index(CFI), incremental fit index (IFI), relative fit index (RFI), standardised root mean square residual (SRMR) and its 90%

confidence interval (90% CI) and test of close fit (CFit), and Hoelter's critical N. Each of the overall goodness-of-fit indices suggested that the six-factor model fit the data well[43] (χ^2 =150.30, df=390, p = 1.0; RMSEA=0.0 with 90%CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.97; CFI=1.00; IFI=1.05; RFI=0.97; SRMR=0.051; Critical N = 250.61; a full list of fit indices available in Appendix VIII).

An inspection of standardised residuals and modification indices indicated no localised points of ill fit in the solution (e.g., the largest modification index= 6.64, largest standardised residual = 2.58). All freely estimated unstandardised parameters were statistically significant (p<0.01). Standardised parameter estimates from this solution are presented in Figure 5.5. Except Item 11, factor loading estimates ranged between 0.65 and 0.77, suggesting that all items were moderately strong in their relations with their purported latent factors (except Item 11, square multiple correlation or R^2 ranged from 0.42 to 0.59). Except for the factor *Strategies to enhance learning and lesson focus,* the average variance extracted for each factor (i.e., the average squared factor loading, Hair et al., 2006, p.777[44]) as well as that of the whole scale was close to or above 50%. This indicates that acceptable convergent validity was found for most factors and for the whole model. Also, the composite reliability computed from the squared sum of factor loadings for each construct and the sum of error variance terms for a construct (see Hair et al., 2006) was generally high at 0.79 or above. Reliability tests based on Cronbach's alpha ranged from 0.70 to 0.97[45] for each factor as well as for the whole scale were also acceptable to good. These results suggest each factor and the scale were good in terms of internal consistency.

43 It should be noted that the current results were obtained through setting the ridge option to 0.06 as the matrix was initially not positive definite. The matrix became positive definite when ridge value reached 0.6. Several other ridge values were tried, but this produced the best results. The ridge option has been a standard option since LISREL 7 to adjust regression models with near-multicollinearity (see Jöreskog & Sörbom, 1996, pp.24; 167; 169; 322). With the high reliability found for ISTOF factors and the whole scale, it was very likely that this kind of adjustment might be required.

44 The average variance extracted and composite reliability below were calculated using the equations discussed in Hair et al. (2006) in Excel, but an online calculator developed by Md-Basir et al. (2010) is also available for obtaining the same values at http://www.hishammb.net/cvc2/

45 There are different methods of estimating reliability. There are discrepancies between the results of different estimations. For example, the estimations based on congeneric model or factor analysis tended to be lower than those based on Cronbach's alpha. Widhiarso (2007) demonstrated that congeneric models tended to bias more than alpha estimates for multidimensional measures.

A split-half test was run for the whole scale with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.95. Spearman-Brown Equal Length and Unequal Length are both 0.93 and Guttman Split-Half coefficient is 0.92.

Nevertheless, it must be acknowledged that the various factors though distinct were themselves fairly strongly associated. An inspection of high intercorrelations between factors in Figure 4.5 suggested that the underlying dimensions might be weak in discriminant validity. Accordingly, a test comparing the average variance extracted and shared variances was performed to measure the discriminant validity of the factors (see Hair et al., 2006; Farrell, 2010) and the results are shown in Table 4.7.

Table 4.7: **Average variance extracted and shared variance estimates of the underlying dimensions of the six-factor CVCP (ISTOF) Model**

Factor Name	*Meta-Cognitive Skills Teaching*	*Classroom Manage-ment & Climate*	*Differentiation & Support*	*Clarity & Logic of Presentation*	*Student Engage-ment*	*Strategies to enhance learning and lesson focus*
Meta-cognitive skills teaching	**0.53**	0.56	0.46	0.53	0.50	0.35
Classroom management and climate	0.75	**0.54**	0.72	0.61	0.66	0.34
Differentiation and support	0.68	0.85	**0.50**	0.69	0.79	0.64
Clarity and logic of presentation	*0.73*	*0.78*	*0.83*	**0.49**	0.76	0.62
Student engagement	*0.71*	*0.81*	***0.89***	*0.87*	**0.49**	0.61
Strategies to enhance learning and lesson focus	0.59	0.58	*0.80*	0.79	0.78	**0.34**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

Among all factors, *Student engagement* and *Clarity and logic of presentation* are the two factors that show more high intercorrelations with other factors. As shown in Table 4.6, the factor *Student engagement* also includes several items that had high cross-loadings, which might lead to its high correlations with other factors. In addition, recognising that student engagement may be seen as a likely outcome under the influence of teaching (van de Grift, 2007), further analyses were conducted (see Sections 6.3 to 6.5), where it was treated as a dependent variable and predicted by different factors.

Since the average variance explained estimates on the diagonal are generally lower than the shared variance estimates above the diagonal in Table 4.7, which are the squared correlations below the diagonal, the model showed insufficient discriminant validity. Thus, these results indicated a common problem about establishing the multidimensionality of teaching because different dimensions tended to be highly correlated (see also Muijs

& Reynolds, 2000, who preferred to use a composite score to indicate overall effective teaching instead). However, given that adequate convergent validity and composite reliability were established in general, further analysis to establish any second order variables, and thus fewer variables, can be attempted in future. Alternatively, it is argued that it is helpful to consider different distinctive features of classroom practice, booth from a theoretical and a practical view point. The concepts of generic and differentiated teacher effectiveness will be discussed in more detail in Chapter 7 and 8.

4.5 Variation across lessons of the six dimensions of effective classroom practices in the CVCP (ISTOF) Model

As defined in Chapter One, variation in classroom practices is operationally defined as variability found in the teacher's behaviours across lessons. Variability across teachers and variability across lessons are considered as indicators of differentiated teacher effectiveness in Chapter Two. Characteristics of the six dimensions underlying the CVCP (ISTOF) Model showed in Table 4.8 illustrate the latter and seem to confirm two dimensions of the nature of differentiated teacher effectiveness: teachers vary in their strengths and weaknesses in different dimensions of effective classroom practices and these dimensions vary across lessons.

Table 4.8: **Characteristics of underlying dimensions of effective classroom practices identified in the CVCP (ISTOF) Model (N=76)**

Factor Name	Aggregate Mean	Median	Mode	Std. De-viation	Skewness	z-skewness	Kurtosis	z-kurtosis
Meta-cognitive skills teaching	2.01	2.04	***1.46***	0.68	-0.21	0.78	-0.90	**1.66**
Classroom management and climate	2.87	3.23	3.68	0.85	-0.79	***2.86****	-0.55	1.01
Differentiation and support	2.22	2.17	3.05	0.72	-0.10	0.36	-0.84	**1.53**
Clarity and logic of presentation	2.52	2.63	2.98	0.59	-0.60	***2.19****	-0.36	0.66
Student engagement	2.26	2.29	***1.39***[46]	0.68	-0.04	0.13	-0.62	1.14
Strategies to enhance learning and lesson focus	2.08	2.02	***2.68***	0.45	-0.08	0.30	-0.88	**1.62**

An examination of individual factors in the following section reveals the magnitude of variation for each factor and indicates the characteristics of the lessons of the four Hong Kong teachers in the sample. In general, negative skewness and kurtosis were noted. On the one hand, negative skewness

46 Multiple modes exist. The smallest value given by SPSS16 is shown.

means that the majority of the lessons were rated above the mean. This was evident in factors *Classroom management and climate* and *Clarity and logic of presentation* as both median and mode were above the mean. On the other hand, the negative kurtosis or platykurtic distribution[47] in every factor paints a more worrying picture for school administrators as it indicates that the number of lessons around the mean was smaller than normal due to large variations across the lessons observed. Although none of the z-scores for the kurtosis of the factors reaches a significant level (i.e., at or high than 1.96), a high probability for extreme ratings might still be found, especially for three factors: *Meta-cognitive skills teaching, Classroom management and climate*, and *Strategies to enhance learning and lesson focus.*

4.5.1 Factor *Meta-cognitive skills teaching*

This first factor of the current CVCP (ISTOF) model indicates an empirical support for the two indicators of Component Five (i.e., IND51: *The teacher helps pupils develop problem-solving and meta-cognitive strategies* and IND53: *The teacher fosters critical thinking in Ss*). Four items with the lowest factor loadings in the EFA factor were deleted, resulting five items shown in Table 4.9 with moderate high factor loadings ranged from 0.69 to 0.74. Their similar square multiple correlations suggest that these items are of similar importance in their contributions to the factor. High reliability of this factor (with Cronbach's alpha = 0.94) indicates its high internal consistency.

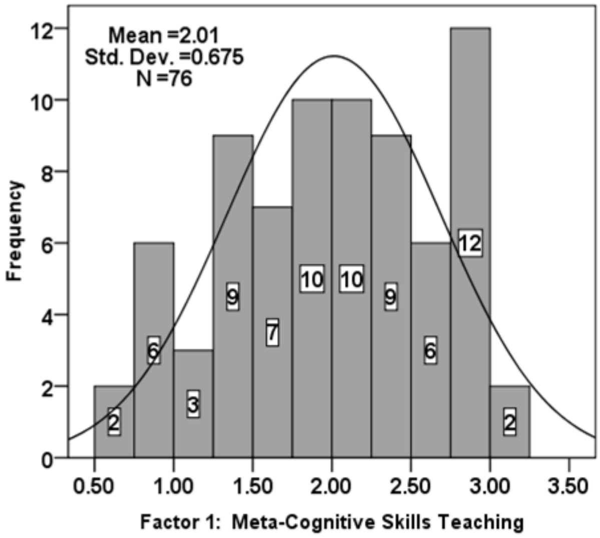
Table 4.9: **Coherence and relative importance of items as identified in the factor estimates of Factor *Meta-Cognitive Skills Teaching* in the CVCP (ISTOF) Model (N=76)**

Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Meta-cognitive skills teaching (Cronbach's alpha=0.94)	IND51	**Item 21**	T invites Ss to use strategies which can help them solve different types of problems.	0.74	1.66	0.55
	IND51	**Item 22**	T invites Ss to explain the different steps of the problem solving strategy which they are using.	0.71	fixed at 1	0.50
	IND51	**Item 23**	T explicitly provides instruction in problem-solving strategies.	0.69	0.78	0.48
	IND53	**Item 27**	T asks the Ss to reflect on the solutions/ answers they gave to problems or questions.	0.76	1.04	0.58
	IND53	**Item 28**	T invites the Ss to give their personal opinion on certain issues.	0.74	1.08	0.55

47 A platykurtic distribution is characterised by a high degree of flatness as scores in such a distribution tend to be clustered away from the mean but much more closely to the two ends than they would be in normal distributions. The tails of a leptokurtic distribution are thus shorter and thicker than those of a normal distribution.

Figure 4.6 below shows a slightly multimodal platykurtic distribution with several peaks. Though not all the peaks were noted as modes in SPSS, those at either tail might have led to variability in ratings for this factor across lessons because some lessons were distinctively rated high or low for teaching behaviours specified by this factor. The negative kurtosis (= -0.9) and its z-score (=1.66) are the highest among the factors. This is also the only factor with a mode (=1.46) below the mean (=2.01) and the median (=2.04) by almost one standard deviation (=0.68). This mode at 1.46 would indicate a high number of lessons were on average the rater disagreed on most of the five items of the factor.

Figure 4.6: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of *Factor Meta-cognitive skills teaching***



The **lowest** possible value for this factor is 0.73, meaning that the rater **strongly disagreed** on all 5 items for a particular lesson.

A value of 1.46 may mean that on average the rater **disagreed** on most of the 5 items for a particular lesson.

A value of 2.18 may mean that the ratings on the 5 items were average out to a **neutral** position for a particular lesson.

A value of 2.91 may mean that on average the rater **agreed** on most of 5 items for a particular lesson.

The **highest** possible value for this factor is 3.64, meaning that the pupils **strongly agreed** on all 5 items.

Skewness = -0.21; Kurtosis = -0.9

4.5.2 Factor *Classroom management and climate*

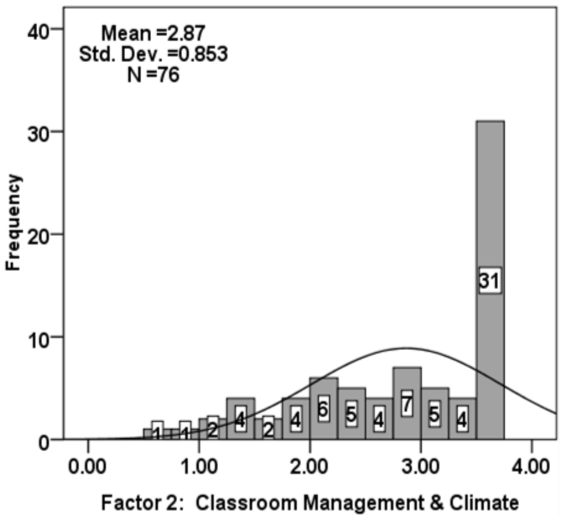
Table 4.10 shows that this factor was made up of items from Component Six (*Classroom Climate*) and Component Seven (*Classroom Management*) in the EFA model, but only one item of Component Six was retained in this factor in the CFA model. Thus, the factor was dominant by the four items concerning classroom management. Yet, the inclusion of Item 31 suggests that a positive classroom climate is compatible and crucial to the proactive and positive classroom management. Again, in Table 4.10 both factor loadings and square multiple correlations of all items were moderately high and similar, suggesting their similar relative importance. This factor received the highest reliability (Cronbach's alpha= 0.95) among the six factors, strongly confirming the coherence of the items.

Table 4.10: **Coherence and relative importance of items as identified in the factor estimates of Factor *Classroom management and climate* in the CVCP (ISTOF) Model (N=76)**

Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Classroom management and climate (Cronbach's alpha=0.95)	IND61	**Item 31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	0.72	1.36	0.52
	IND71	**Item 41**	Actions are taken to minimize disruption.	0.71	fixed at 1	0.50
	IND72	**Item 42**	There is clarity about when and how Ss can get help to do their work in class.	0.74	1.28	0.55
	IND73	**Item 44**	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	0.74	1.73	0.55
	IND73	**Item 45**	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	0.77	1.30	0.59

Figure 4.7 shows a statistically significant (p<0.005) negatively skewed distribution with a dispersed long tail and a clear peak in the histogram. About half of the lessons received the highest possible ratings for all or most of the items in this factor, but in some lessons (about 22% with a factor score below 2.21) low ratings were found for most of the items of the factor indicating the lessons were poorly managed.

Figure 4.7: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Classroom management and climate* (N=76)**



Skewness = -0.79; Kurtosis = -0.55

The **lowest** possible value for this factor is 0.74, meaning that the rater **strongly disagreed** on all 5 items for a particular lesson.

A value of 1.47 may mean that on average the rater **disagreed** on most of the 5 items for a particular lesson.

A value of 2.21 may mean that the ratings on the 5 items were average out to a **neutral** position for a particular lesson.

A value of 2.94 may mean that on average the rater **agreed** on most of 5 items for a particular lesson.

The **highest** possible value for this factor is 3.68, meaning that the pupils **strongly agreed** on all 5 items.

4.5.3 Factor *Differentiation and support*

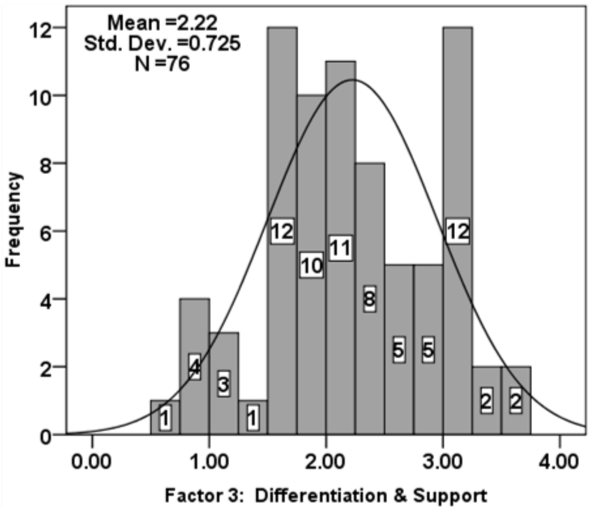
Of the seven items of this factor in the EFA, six items were retained in the CFA. Though more items were included, their factor loadings and square multiple correlations as shown in Table 4.11 remained moderately high and similar, indicating their similar relative importance. The internal consistency of this factor was indicated by a high reliability score (Cronbach's alpha=0.94) as well as by an inspection of the meanings of the items. The highlighted italic words of each item in Table 4.11 clearly indicate the kind of differentiation and support strategies that the teacher might have adopted in the lesson. It would range from scope of assignments, additional practices, longer wait time and response strategies, opportunity to correct one's own work, teacher-initiated participation, and optional due date for assignments.

Table 4.11: **Coherence and relative importance of items as identified in the factor estimates of Factor *Differentiation and support* in the CVCP (ISTOF) Model (N=76)**

Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Differentiation and support (Cronbach's alpha=0.94)	IND22	**Item 7**	T makes a distinction in the ***scope of the assignments for different groups*** of Ss.	0.74	1.27	0.55
	IND22	**Item 8**	T gives ***additional opportunities for practice*** to Ss who need them.	0.68	fixed at 1	0.46
	IND41	**Item 15**	T provides ***sufficient wait time and response strategies to involve all types*** of learners.	0.71	0.71	0.50
	IND52	**Item 25**	T gives Ss the ***opportunity to correct*** their own work.	0.69	0.69	0.48
	IND63	**Item 35**	T gives turns to and/or ***involves those Ss who do not voluntarily participate*** in classroom activities.	0.71	1.05	0.50
	IND72	**Item 43**	There is clarity about what ***options are available when the Ss finish their assignments.***	0.70	1.30	0.49

Figure 4.8 below shows the multiple peaks of this factor. These peaks may have contributed to a likely multimodal platykurtic distribution with a high z-score (=1.53) for its negative kurtosis (= -0.84). These results suggest though there was moderately high variability in ratings for this factor across lessons. There were about more lessons receiving similar range of ratings at the high and middle intervals as indicated by the two prominent peaks in the histogram.

Figure 4.8: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Differentiation and support* (N=76)**



Skewness = -0.10; Kurtosis = -0.84

The **lowest** possible value for this factor is 0.71, meaning that the rater **strongly disagreed** on all 6 items for a particular lesson.

A value of 1.41 may mean that on average the rater **disagreed** on most of the 6 items for a particular lesson.

A value of 2.12 may mean that the ratings on the 6 items were average out to a **neutral** position for a particular lesson.

A value of 2.82 may mean that on average the rater **agreed** on most of 6 items for a particular lesson.

The **highest** possible value for this factor is 3.53, meaning that the pupils **strongly agreed** on all 6 items.

4.5.4 Factor *Clarity and logic of presentation*

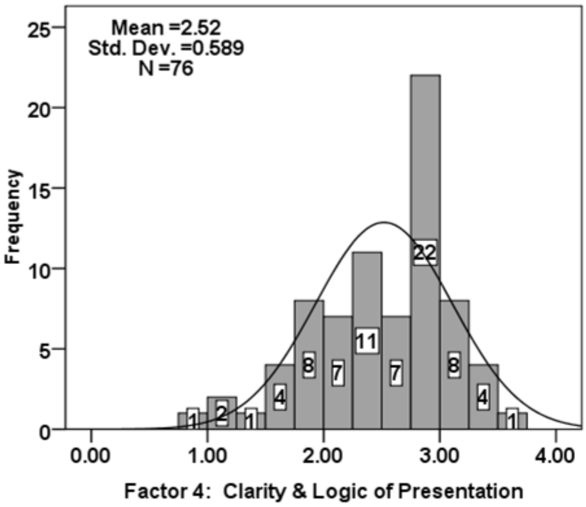
After deletion of three items in the EFA factor, the CFA solution for this factor in Table 4.12 conveys an emphasis on the organizational structure of the lesson (Items 13 and 14). Feedbacks to student answers (Item 2) and appropriate question skills (Item 18) enhance the clarity of presentation and reflect the communication skills of the teacher. Moderately high and similar factor loadings and square multiple correlations again indicate the similar weighting of these items in the factor. The reliability test also indicates high internal consistency for these items (Cronbach's alpha=0.91).

Table 4.12: **Coherence and relative importance of items as identified in the factor estimates of Factor *Clarity and logic of presentation* in the CVCP(ISTOF) Model (N=76)**

Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Clarity and logic of presentation (Cronbach's alpha=0.91)	IND11	**Item 2**	T provides appropriate feedback to the answers given by the Ss.	0.71	0.71	0.50
	IND33	**Item 13**	T presents the lesson with a logical flow that moves from simple to more complex concepts.	0.67	1.28	0.45
	IND33	**Item 14**	T implements the lesson smoothly moving from one stage to another with well-managed transition points.	0.70	fixed at 1	0.49
	IND42	**Item 18**	The length of the pause following questions varies according to the difficulty level of questions	0.72	0.92	0.52

Like the factor *Classroom management and climate*, Figure 4.9 shows that this is another factor that has a statistically significant (p<0.05) negatively skewed distribution. In about 70% of the lessons, the teacher was rated at neutral or better on average for the items of this factor. This would suggest that in most lessons, lessons in the sample were rated favourably for most items of the factor.

Figure 4.9: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Clarity and logic of presentation* (N=76)**



Skewness = -0.60; Kurtosis = -0.36

The **lowest** possible value for this factor is 0.7, meaning that the rater **strongly disagreed** on all 4 items for a particular lesson.

A value of 1.4 may mean that on average the rater **disagreed** on most of the 4 items for a particular lesson.

A value of 2.1 may mean that the ratings on the 4 items were average out to a **neutral** position for a particular lesson.

A value of 2.8 may mean that on average the rater **agreed** on most of 4 items for a particular lesson.

The **highest** possible value for this factor is 3.5, meaning that the pupils **strongly agreed** on all 4 items for a particular lesson.

4.5.5 Factor *Student engagement*

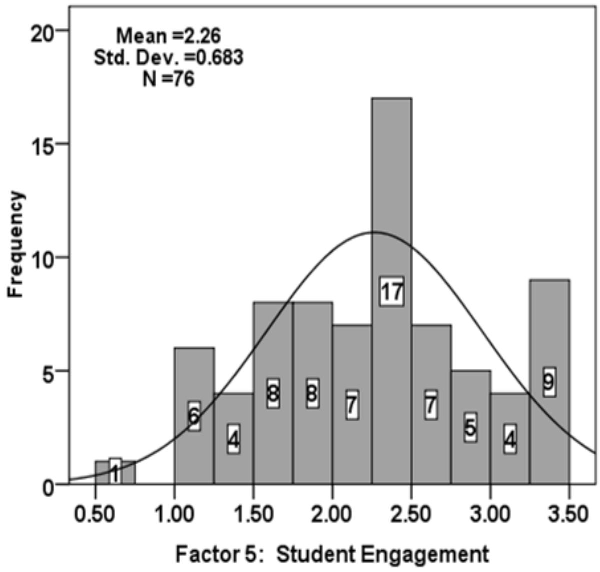
Like the factor *Differentiation and support*, this factor shown in Table 4.13 retained most EFA items in the CFA solution. Again the item with the lowest factor loading was excluded. As indicated in Table 4.6, many of the items in this factor were found cross-loaded in other factors, probably led to its high intercorrelations with other factors as shown in Table 4.7.

Table 4.13: **Coherence and relative importance of items as identified in the factor estimates of Factor *Student engagement* in the CVCP (ISTOF) Model (N=76)**

Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Student engagement (Cronbach's alpha=0.94)	IND21	**Item 5**	***Ss communicate frequently*** with one another ***on task-oriented issues***.	0.65	fixed at 1	0.42
	IND21	**Item 6**	All Ss are actively ***engaged in learning***	0.72	1.41	0.52
	IND41	**Item 16**	T gives assignments that ***stimulate all Ss to active involvement***.	0.72	1.19	0.52
	IND54	**Item 29**	T systematically uses material and examples from the Ss' ***daily life to illustrate the course content.***	0.68	1.16	0.46
	IND54	**Item 30**	Ss are invited to ***give their own examples***.	0.68	1.34	0.46
	IND62	**Item 33**	T creates purposeful activities that ***engage every S*** in productive work.	0.75	2.72	0.56
	IND63	**Item 36**	T seeks to ***engage all Ss*** in classroom activities.	0.67	0.81	0.45

This factor included many items originated in different components of the theoretical scale, but they highlighted common and active engagement in diverse ways: in communication on task-oriented issues, assignment, learning, familiar daily life content, own examples, productive work and classroom activities assigned (see the highlighted words in the above table). Not surprisingly, the factor loadings and square multiple correlations suggest their similar importance and the reliability score indicates its high internal consistency (Cronbach's alpha=0.94). Though the negative kurtosis (= -0.62) of this factor was not statistically significant (i.e., z-kurtosis=1.14), the histogram in Figure 4.10 shows that a multimodal platykurtic distribution with a short/thick tail at the right end. Yet, a high variability of ratings for this factor was evident because multiple modes were noted by SPSS and there were some outlier lessons with highest and lowest possible ratings.

Figure 4.10: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Student engagement***



Skewness = -0.04; Kurtosis = -0.62

The **lowest** possible value for this factor is 0.7, meaning that the rater **strongly disagreed** on all 7 items for a particular lesson.

A value of 1.39 may mean that on average the rater **disagreed** on most of the 7 items for a particular lesson.

A value of 2.09 may mean that the ratings on the 7 items were average out to a **neutral** position for a particular lesson.

A value of 2.78 may mean that on average the rater **agreed** on most of 7 items for a particular lesson.

The **highest** possible value for this factor is 3.48, meaning that the pupils **strongly agreed** on all 7 items for a particular lesson.

4.5.6 Factor *Strategies to enhance learning and lesson focus*

As shown in Table 4.14, this factor was formed by combining two EFA factors relating to learning goal and lesson objectives, Item 11 were found with high cross-loadings on two factors. As a result, Item 11 seemed to hold on together with Item 4 and Item 37, though the factor loading of Item 11 was low at 0.35 and its square multiple correlation was only 0.12. Deleting Item

11 might improve some selected goodness-of-fit indices[48] but would not justify the cost of a model with a factor with only two items. Moreover, both the reliability test (Cronbach's alpha=0.7) and inspection of the meanings of items indicate that the three items show reasonable internal consistency.

Table 4.14: **Coherence and relative importance of items as identified in the factor estimates of Factor *Strategies to enhance learning and lesson focus* in the CVCP (ISTOF) Model (N=76)**

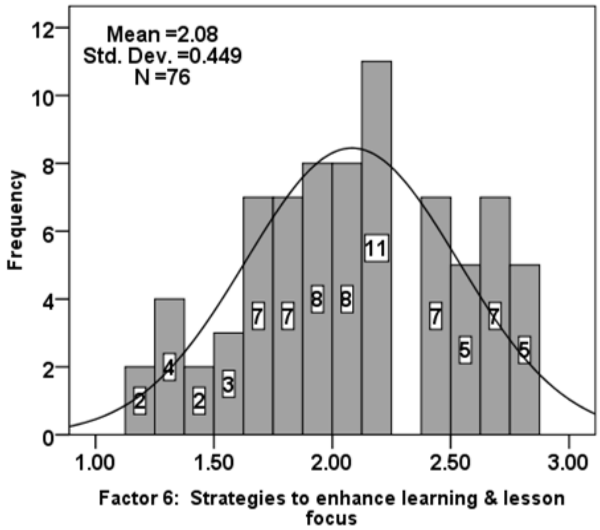
Factor Name	Indi-cator No.	Item No	Item Description (T = The teacher; S= student)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Strategies to enhance learning and lesson focus (Cronbach's alpha=0.70)	IND22	**Item 4**	T explains how assignments are aligned to the learning goals of the lesson.	0.65	fixed at 1	0.42
	IND32	**Item 11**	T clarifies the lesson objectives at the start of the lesson.	0.35	0.60	0.12
	IND64	**Item 37**	T praises children for effort towards realizing their potential.	0.68	1.23	0.46

In contrast to the factor *Meta-cognitive skills teaching*, the mode (=2.68) of this factor was *higher* than its mean (=2.08) and median (=2.02) by more than one standard deviation (=0.45). However, Table 4.8 shows that like the factors *Meta-cognitive skills Teaching* and *Differentiation and support*, the negative kurtosis (= -0.88) of this factor was high though its z-score (=1.62) was not statistically significant.

The platykurtic distribution of this factor in Figure 4.11 reveals a greater variability in ratings because in a group of lessons the teacher were rated very positively for most or all items of this factor, but a slightly higher number of lessons were also rated on average somewhat negatively or neutral on most of the items.

48 A nested model with Item 11 deleted from the six-factor CFA model presented in Section 5.4 would improve very little some of the goodness-of-fit indices while other indices remained constant: NFI improved by 0.01; SRMR by 0.006; Critical N by 10.65, but $\chi 2$ only insignificantly improved by 16.9 after degree of freedom decreased by 28.

Figure 4.11: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Strategies to enhance learning and lesson focus***



Skewness = -0.08; Kurtosis = -0.88

The **lowest** possible value for this factor is 0.56, meaning that the rater **strongly disagreed** on all 3 items for a particular lesson.

A value of 1.12 may mean that on average the rater **disagreed** on most of the 3 items for a particular lesson.

A value of 1.68 may mean that the ratings on the 3 items were average out to a **neutral** position for a particular lesson.

A value of 2.24 may mean that on average the rater **agreed** on most of 3 items for a particular lesson.

The **highest** possible value for this factor is 2.8, meaning that the rater **strongly agreed** on all 3 items for a particular lesson.

4.6 The extent of similarities and discrepancies between CVCP factors and theoretical factors

4.6.1 Patterns of differences in the shared items of CVCP factors and theoretical factors

Although excellent fit indices were obtained for the current model, it should be emphasised that it was deviated from the seven-factor theoretical model, as shown in Table 4.15 below.

Table 4.15: **Shared items in the CVCP (ISTOF) factors and the theoretical components of ISTOF (N=76)**

Current CVCP Factor	Item No	Item Description (T = The teacher; S= student)	Original Theoretical Component	Item No. in concern
Meta-cognitive skills teaching	**Item 21**	T invites Ss to use strategies which can help them solve different types of problems.	**Promoting active learning and developing metacognitive skills**	**21, 22, 23**, 24 **25**, 26, **27, 28** 29, 30
	Item 22	T invites Ss to explain the different steps of the problem solving strategy which they are using.		
	Item 23	T explicitly provides instruction in problem-solving strategies.		
	Item 27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.		
	Item 28	T invites the Ss to give their personal opinion on certain issues.		
Classroom management and climate	**Item 31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	**Classroom climate**	**31**, 32, 33, 34 35, 36, 37, 38
	Item 41	Actions are taken to minimize disruption.	**Classroom management**	39, 40, **41, 42** 43, **44, 45**
	Item 42	There is clarity about when and how Ss can get help to do their work in class.		
	Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.		
	Item 45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.		

Current CVCP Factor	Item No	Item Description (T = The teacher; S= student)	Original Theoretical Component	Item No. in concern
Differentiation and support	**Item 7**	T makes a distinction in the scope of the assignments for different groups of Ss.	**Differentiation and inclusion**	5, 6, **7, 8**
	Item 8	T gives additional opportunities for practice to Ss who need them.		
	Item 15	T provides sufficient wait time and response strategies to involve all types of learners.	**Instructional skills**	**15,** 16, 17, 18 19, 20
	Item 25	T gives Ss the opportunity to correct their own work.	**Promoting active learning and developing meta-cognitive skills**	21, 22, 23, 24 **25**, 26, 27, 28 29, 30
	Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	**Classroom climate**	31, 32, 33, 34 **35**, 36, 37, 38
	Item 43	There is clarity about what options are available when the Ss finish their assignments.	**Classroom management**	39, 40, 41, 42 **43**, 44, 45
Clarity and logic of presentation	**Item 2**	T provides appropriate feedback to the answers given by the Ss.	**Assessment and evaluation**	1, **2**, 3, 4
	Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts.	**Clarity of instruction**	9, 10, 11, 12, **13, 14**
	Item 14	T implements the lesson smoothly moving from one stage to another with well-managed transition points.		
	Item 18	The length of the pause following questions varies according to the difficulty level of questions	**Instructional skills**	15, 16, 17, **18** 19, 20
Student engagement	**Item 5**	Ss communicate frequently with one another on task-oriented issues.	**Differentiation and inclusion**	**5, 6**, 7, 8
	Item 6	All Ss are actively engaged in learning.		
	Item 16	T gives assignments that stimulate all Ss to active involvement.	**Instructional skills**	15, **16,** 17, 18 19, 20
	Item 29	T systematically uses material and examples from the Ss' daily life to illustrate the course content.	**Promoting active learning and developing metacognitive skills**	**21, 22, 23, 24 25, 26, 27, 28 29, 30**
	Item 30	Ss are invited to give their own examples.		
	Item 33	T creates purposeful activities that engage every S in productive work.	**Classroom climate**	**31, 32, 33, 34 35, 36, 37, 38**
	Item 36	T seeks to engage all Ss in classroom activities.		
Strategies to enhance learning and lesson focus	**Item 4**	T explains how assignments are aligned to the learning goals of the lesson.	**Assessment and evaluation**	**1, 2, 3, 4**
	Item 11	T clarifies the lesson objectives at the start of the lesson.	**Clarity of instruction**	**9, 10, 11, 12, 13, 14**
	Item 37	T praises children for effort towards realizing their potential.	**Classroom climate**	**31, 32, 33, 34 35, 36, 37, 38**

Part of deviations was a result of the fact that one factor and several items were deleted to reduce the number of parameters for estimation. However, it was also likely that some of the components of the theoretical factor might need revision, though it has been modified many times using Delphi method (Teddlie et al., 2006). Like the ECP study, the CVCP study may be considered as one of the many ongoing empirical studies[49] that piloted the instrument for scale development. There are four patterns in the

49 According to Teddlie et al. (2006), there were about twenty-five participating countries in developing ISTOF piloting on the instrument. Some of these countries are piloting the instrument.

shared items between CVCP factors and the theoretical components as shown in Table 4.15.

First, there is strong support for retaining the theoretical components as unique for meta-cognitive skills teaching and classroom management. Unlike the ECP study, which had a significant number of missing data for the component, the CVCP study allowed for exploration of this factor as indeterminate rating was avoided. Consequently, the current model showed that Component 5 (*Promoting active learning and developing metacognitive skills*) and Component 7 (*Classroom Management*) could be retained after deletions of some items.

Second, some components were retained but showed different combinations of items. For example, those items originally present in other components but suggesting teachers' feedback and logical presentation flow enriched the original component *Clarity of instruction.* Similarly, Component *Differentiation and inclusion* was largely expanded with items indicating supportive and inclusive strategies for students like Items 15, 35 and 43.

Third, new factors like *Student Engagement* and *Strategies to enhance learning and lesson focus* seemed to emerge in the current model. Containing items of four theoretical components, *Student engagement* suggests a mixture of strategies to be inclusive (i.e., Items 5 and 6), to engage students with real life experiences (i.e., Items 15 and 16), and to enhance participation and involvement (e.g., Items 16, 30, and 36). Items of factor *Strategies to enhance learning and lesson focus* also reflect its composite nature. Explanation of purpose (Item 11) and assignment aligning to the goal of the lesson (Item 4) seem to be linked together as these two items also formed a distinctive factor in the ECP model. Additional praise to help realise the potentials of the students (Item 37) may be crucial to create a sense of purpose in students when they have to learn a compulsory foreign language. In general, appropriate praise can be a strong external motivator for learners who lack confidence.

Finally, the current solution seemed to show poor support for *Assessment and Evaluation, Instructional Skills* and *Classroom Climate* elements of the original scale as distinctive components, though many items of these components were retained in different CVCP factors. It may be theoretically possible to categorise classroom practices of these components as distinctive, but they might not be hold on together empirically as distinctive dimensions of effective classroom practices. Alternatively, these constructs may better be considered as multi-dimensional, but the instrument was not constructed in such a way to allow this to be tested in the present study, nor it was intended to focus on any single dimension of effective classroom practices.

4.6.2 Exploring the convergent and discriminant validity of the seven-component theoretical model of ISTOF

To explore the convergent and discriminant validity of the theoretical model, several confirmatory factor analyses (CFA) were conducted to establish a good-fitted model for each theoretical component and then proceeded to build a full model because starting off analyses with an initial full model including all items would result in unreliable estimates when the total number of parameters exceeded the sample size. Five out of seven theoretical components produced reasonably goodness-of-fit indices after rigorous data-driven procedures and the results are summarised in Table 4.16. The *fourth* column of Table 4.16 indicates the factor loadings of the items of the CFA model for each individual component.

Table 4.16: **Factor loadings, reliability, and selected goodness-of-fit indices of the CFA model of each ISTOF theoretical component in the CVCP study (N=76)**

Component Model	**Item No.**	**Item Description (T = The teacher; S= student)**	**Factor Loading**	**Selected Goodness-of-fit Indices**
Assessment and Evaluation (Cronbach's alpha=0.88)	**1**	T makes explicitly clear why an answer is correct or not.	0.97	$\chi2$ =55.53, df=2, p = 0.0; RMSEA=0.60; 90%CI=0.47, 0.74; p-value for CFit (RMSEA<0.05) = 0.0; NFI= 0.61; CFI=0.61; IFI=0.61; RFI=-0.18; SRMR=0.13
	2	T provides appropriate feedback to the answers given by the Ss.	1.01	
	3	Assignments given by T are clearly related to what Ss learned.	0.78	
	4	T explains how assignments are aligned to the learning goals of the lesson.	0.63	
Differentiation and Inclusion (Cronbach's alpha=0.90)	**5**	Ss communicate frequently with one another on task-oriented issues.	0.76	$\chi2$ =20.08, df=2, p = 0.00004; RMSEA=0.35; 90%CI=0.22, 0.49; p-value for CFit (RMSEA<0.05) = 0.00016; NFI= 0.92; CFI=0.93; IFI=0.93; RFI=-0.77; SRMR=0.048
	6	All Ss are actively engaged in learning.	0.88	
	7	T makes a distinction in the scope of the assignments for different groups of Ss.	0.96	
	8	T gives additional opportunities for practice to Ss who need them.	0.85	

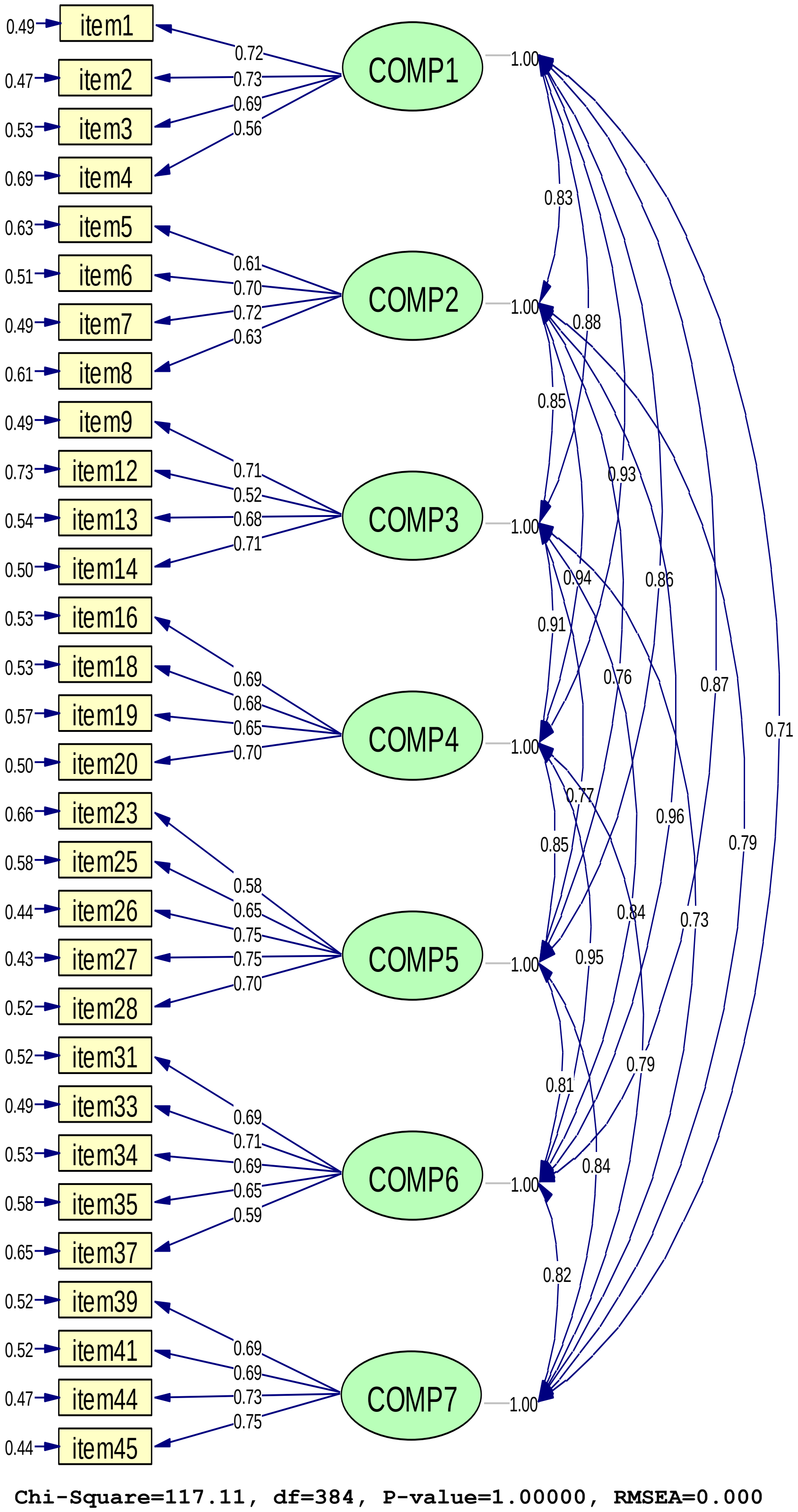
Component Model	Item No.	Item Description (T = The teacher; S= student)	Factor Loading	Selected Goodness-of-fit Indices
Clarity of Instruction (Cronbach's alpha=0.88)	**9**	T regularly checks for understanding.	0.77	χ2 =1.11, df=2, p = 0.57; RMSEA=0.00; 90%CI=0.0, 0.19; p-value for CFit (RMSEA<0.05) = 0.63; NFI= 1.00; CFI=1.00; IFI=1.00; RFI=-0.99; SRMR=0.0098
	12	T asks Ss to identify the reasons why specific activities take place in the lesson.	0.68	
	13	T presents the lesson with a logical flow that moves from simple to more complex concepts.	0.97	
	14	T implements the lesson smoothly moving from one stage to another with well-managed transition points.	0.99	
Instructional Skills (Cronbach's alpha=0.91)	**16**	T gives assignments that stimulate all Ss to active involvement.	0.80	χ2 =0.33, df=2, p = 0.85; RMSEA=0.00; 90%CI=0.00, 0.12; p-value for CFit (RMSEA<.05) = 0.87; NFI= 1.00; CFI=1.00; IFI=1.01; RFI=-1.00; SRMR=0.0046
	18	The length of the pause following questions varies according to the difficulty level of questions.	0.91	
	19	T uses a variety of instructional strategies during the class period	0.90	
	20	T explicitly provides instruction in problem-solving strategies.	0.95	
Promoting active learning and developing metacognitive skills (Cronbach's alpha=0.93)	**23**	T explicitly provides instruction in problem-solving strategies.	0.74	χ2 =2.09, df=5, p = 0.84; RMSEA=0.00; 90%CI=0.00, 0.093; p-value for CFit (RMSEA<.05) = 0.88; NFI= 0.99; CFI=1.00; IFI=1.01; RFI=-0.99; SRMR=0.015
	25	T gives Ss the opportunity to correct their own work.	0.71	
	26	T motivates the Ss to think about the advantages and disadvantages of certain approaches	0.94	
	27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	0.97	
	28	T invites the Ss to give their personal opinion on certain issues.	0.91	
Classroom Climate (Cronbach's alpha=0.92)	**31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	0.86	χ2 =5.59, df=5, p = 0.35; RMSEA=0.60; 90%CI=0.00, 0.17; p-value for CFit (RMSEA<.05) = 0.46; NFI= 0.99; CFI=1.00; IFI=1.00; RFI=-0.97; SRMR=0.018
	33	T creates purposeful activities that engage every S in productive work.	0.94	
	34	T's instruction is interactive (lots of questions and answers).	0.89	
	35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	0.87	
	37	T praises children for effort towards realizing their potential.	0.80	
Classroom Management (Cronbach's alpha=0.94)	**39**	T starts lesson on time.	0.82	χ2 =2.25, df=2, p = 0.32; RMSEA=0.041; 90%CI=0.00, 0.24; p-value for CFit (RMSEA<.05) = 0.39; NFI= 0.99; CFI=1.00; IFI=1.00; RFI=-0.98; SRMR=0.010
	41	Actions are taken to minimize disruption.	0.93	
	44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	0.99	
	45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	0.98	

As shown in Table 4.16., models for Components *Assessment and Evaluation* and *Differentiation and Inclusion* retained all items in the components because there are four items originally in each of these two components, any further reduction of items would yield a model that lack degree of freedom necessary for estimating goodness-of-fit. Thus, without a reduction of items, the CFA analyses could only arrive at a poorly-fitted model for these components. In contrast, a reasonably well-fitted model with four or five items retained was obtained for each of the other components. These results seemed to contradict with what is suggested in the discussion above because they showed some clear support for *Clarity of Instruction*, *Instructional Skills* and *Classroom Climate* of the original scale as distinctive components. However, the fact that twenty out of the thirty items (with their item number in bold and underlined in Table 4.16) in these models were found in the six factors of the CFA item-based model for ISTOF indicated that CVCP factors were still generally comparable to the theoretical components. Regardless of the goodness-of-fit of these models, reliability tests showed that they all achieve high internal consistency, suggesting high coherence of items of each theoretical component.

A full CFA model utilising all individual theoretical components in Table 4.16 with excellent goodness-of-fit indices is shown in Figure 4.12. Each of the overall goodness-of-fit indices suggested that a CFA model with all seven components might also fit the data well (χ^2 =117.11, df=384, *p* = 1.0; RMSEA=0.0 with 90%CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.98; CFI=1.00; IFI=1.06; RFI=0.98; SRMR=0.042; Critical N = 322.94). These indices were slightly better than those obtained for the six-factor empirical model discussed in the last three sections. No large standardised residuals and modification indices (e.g., the largest modification index = 7.99, largest standardised residual = 2.20) were found, indicating the absence of localised points of ill fit in the solution.

Nevertheless, it should be noted that this solution was considered less preferable because a warning message was flagged by LISREL indicating that the parameter estimates were unreliable as the total sample size (N=76) was smaller than the number of parameters (p=81). The number of parameters of this solution was higher than the previous one because this solution has one factor more.

Figure 4.12: **A CFA model of 30 ISTOF items built on seven theoretical components with standardised coefficients (N=76)**



KEYS:

Comp1: Assessment and Evaluation
(Average variance extracted = 44.06%; Composite reliability= 0.80; Cronbach's alpha=0.88)

Comp2: Differentiation and Inclusion
(Average variance extracted = 45.53%; Composite reliability= 0.81; Cronbach's alpha=0.90)

Comp3: Clarity of Instruction
(Average variance extracted = 44.28%; Composite reliability= 0.80; Cronbach's alpha=0.88)

Comp4: Instructional Skills
(Average variance extracted = 43.95%; Composite reliability= 0.80; Cronbach's alpha=0.91)

Comp5: Promoting active learning and developing metacognitive skills
(Average variance extracted = 47.44%; Composite reliability= 0.82; Cronbach's alpha=0.93)

Comp6: Classroom Climate
(Average variance extracted = 44.57%; Composite reliability= 0.80; Cronbach's alpha=0.92)

Comp7: Classroom Management
(Average variance extracted = 51.22%; Composite reliability= 0.81; Cronbach's alpha=0.94)

The whole scale:
(Average variance extracted = 46.17%; Composite reliability= 0.96; Cronbach's alpha=0.98)

Standardised parameter estimates from this solution are presented in Figure 4.12 above and Table 4.17 below.

Table 4.17: **Factor loadings, reliability, and selected goodness-of-fit indices of a full theoretical model consisting of 30 items for comparing with the CVCP (ISTOF) Model**

Component	**Item No.**	**Item Description (T = The teacher; S= student)**	**Factor Loading**	**Unstand-ardised Estimate**	**Square multiple correlation**
Assessment and Evaluation (Cronbach's alpha=0.88)	**1**	T makes explicitly clear why an answer is correct or not.	0.72	fixed at 1	0.51
	2	T provides appropriate feedback to the answers given by the Ss.	0.73	0.63	0.53
	3	Assignments given by T are clearly related to what Ss learned.	0.69	1.17	0.47
	4	T explains how assignments are aligned to the learning goals of the lesson.	0.56	0.53	0.31
Differentiation and Inclusion (Cronbach's alpha=0.90)	**5**	Ss communicate frequently with one another on task-oriented issues.	0.61	fixed at 1	0.37
	6	All Ss are actively engaged in learning.	0.70	1.46	0.49
	7	T makes a distinction in the scope of the assignments for different groups of Ss.	0.72	1.56	0.51
	8	T gives additional opportunities for practice to Ss who need them.	0.63	1.17	0.39
Clarity of Instruction (Cronbach's alpha=0.88)	**9**	T regularly checks for understanding.	0.71	fixed at 1	0.51
	12	T asks Ss to identify the reasons why specific activities take place in the lesson.	0.52	0.36	0.27
	13	T presents the lesson with a logical flow that moves from simple to more complex concepts.	0.68	1.07	0.46
	14	T implements the lesson smoothly moving from one stage to another with well-managed transition points.	0.71	0.83	0.50
Instructional Skills (Cronbach's alpha=0.91)	**16**	T gives assignments that stimulate all Ss to active involvement.	0.69	fixed at 1	0.47
	18	The length of the pause following questions varies according to the difficulty level of questions.	0.68	1.19	0.47
	19	T uses a variety of instructional strategies during the class period	0.65	1.18	0.43
	20	T explicitly provides instruction in problem-solving strategies.	0.70	1.08	0.50
Promoting active learning and developing metacognitive skills (Cronbach's alpha=0.93)	**23**	T explicitly provides instruction in problem-solving strategies.	0.58	fixed at 1	0.34
	25	T gives Ss the opportunity to correct their own work.	0.65	1.88	0.42
	26	T motivates the Ss to think about the advantages and disadvantages of certain approaches	0.75	1.68	0.56
	27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	0.75	1.58	0.57
	28	T invites the Ss to give their personal opinion on certain issues.	0.70	1.56	0.48
Classroom Climate (Cronbach's alpha=0.92)	**31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	0.69	fixed at 1	0.48
	33	T creates purposeful activities that engage every S in productive work.	0.71	1.28	0.51
	34	T's instruction is interactive (lots of questions and answers).	0.69	0.64	0.47
	35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	0.65	0.56	0.42
	37	T praises children for effort towards realizing their potential.	0.59	0.58	0.35
Classroom Management (Cronbach's alpha=0.94)	**39**	T starts lesson on time.	0.69	fixed at 1	0.48
	41	Actions are taken to minimize disruption.	0.69	0.29	0.48
	44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	**0.73**	**0.51**	**0.53**
	45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	**0.75**	**0.44**	**0.56**

All freely estimated unstandardised parameters were statistically significant ($p < 0.01$). Factor loading estimates between 0.52 and 0.75 suggested that all items were moderately strong in their relations with their purported latent factors (square multiple correlation or R^2 ranged from 0.27 to 0.56). Except Component *Classroom Management,* average variance extracted of each factor as well as that of the whole scale was below 50%, indicating *insufficient* convergent validity was found for most factors and the whole model. That is, contrary to the case of the CVCP empirical model presented in Figure 4.5. However, the composite reliability for each component and the whole model was generally high above 0.8. Reliability test based on Cronbach's alpha for each component ranged from 0.88 to 0.94 and that for the whole scale yielded a very high figure at 0.98[50], suggesting the scale were good in terms of internal consistency.

Like the CVCP model, this theoretical model also shows a lot of high intercorrelation values between components in Figure 4.12, suggesting these components might also be weak in discriminant validity. A similar test performed to compare the average variances extracted and shared variances confirmed the insufficient discriminant validity of components as indicated in Table 4.18.

Table 4.18: **Average variance extracted and shared variance estimates of the CFA model based on seven theoretical components of ISTOF**

Factor Name	Assess-ment and Evaluation	Different-iation and Inclusion	Clarity of Instruction	Instruction-al Skills	Promoting active learning & developing meta-cogn-itive skills	Classroom Climate	Classroom Manage-ment
Assessment and Evaluation	**0.44**	0.69	0.77	0.86	0.74	0.76	0.50
Differentiation and Inclusion	0.83	**0.46**	0.72	0.88	0.58	0.92	0.62
Clarity of Instruction	0.88	0.85	**0.44**	0.83	0.59	0.71	0.53
Instructional Skills	0.93	0.94	0.91	**0.44**	0.72	0.90	0.62
Promoting active learning and developing metacognitive skills	0.86	0.76	0.77	0.85	**0.47**	0.66	0.71
Classroom Climate	0.87	0.96	0.84	0.95	0.81	**0.45**	0.67
Classroom Management	0.71	0.79	0.73	0.79	0.84	0.82	**0.51**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

For all components, the average variance explained estimates on the diagonal are lower than the shared variance estimates above the diagonal in Table 4.18, indicating the problem of insufficient discriminant validity is more serious than the CVCP model. Thus, despite a good fit for the data and high internal consistency, the seven-component theoretical model lacked both convergent and discriminant validity in this data set.

---

50 A split-half test was run for the whole scale with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.96. Spearman-Brown Equal Length and Unequal Length are both 0.94 and Guttman Split-Half coefficient is 0.93.

Given the current results, it seemed that both the empirical and theoretical solutions have good internal reliability, though it is acknowledge that their convergent and discriminant validity are weak. The results suggest that all the factors developed are likely to contribute to overall effective teaching behaviour. The factors *Clarity of instruction, Differentiation and support, Meta-cognitive skills teaching*, and *Classroom management* are potentially important dimensions of effective classroom practices, but the CVCP model also suggested that the factors *Student engagement* and *Strategies to enhance lesson focus* might be important as well. Accordingly, there is some merit in treating the dimensions as separate but establishing their patterns of association and teachers' variations in their observed practices in more detail as is done in the four teacher cases in Chapters 7 and 8. Nevertheless, further research with large samples in other contexts is still required to establish their importance in predicting student progress and other outcomes.

4.7 The extent of similarities and discrepancies between CVCP and ECP factors

A close examination of the present factors shows that they are comparable with the original theoretical components of the instrument as well as the factors of the six-factor item-based ECP model. These results provide support to ISTOF as a valid instrument for measuring the classroom practices of teachers in different cultural contexts. Similarities between the CVCP and the ECP models suggest that some characteristics found in the samples of the two studies are comparable, despite their obvious contextual variations. The following paragraphs summarise the findings that address the second sub-question of the first research question: *to what extents are these characteristics comparable between England and Hong Kong, despite their obvious contextual differences*?

Factor Meta-cognitive skills teaching

The first factor, *Meta-cognitive skills teaching*, of the current model indicates an empirical support for the Component Five of the original scale, but it consists only of the five items of two indicators of Component Five (i.e., IND51: *The teacher helps pupils develop problem-solving and meta-cognitive strategies* and IND53: *The teacher fosters critical thinking in Ss*). Items of the remaining indicators of Component Five are still important, but linked with other items to form other factors in the current model. In the Hong Kong context of EFL teaching, active learning and linking real life experiences with learning may be less likely to be considered as falling into the domain of meta-cognitive skills. Items of these two indicators might be less cognitively oriented and thus more likely to be related to items of other components. As all items were excluded in the analysis for deriving the item-based ECP model, there is no factor in that model that is directly comparable to the present factor. Thus, the present result provided a unique empirical evidence for the importance of the theoretical component.

Factor Classroom management and climate

Table 4.19 below shows that the second factor, *Classroom management and climate*, is almost identical to the original Component Seven (*Classroom Management*) of the ISTOF instrument and the second factor of the six-factor item-based ECP model (see Sammons & Ko, 2008).

Table 4.19: **Shared items on Factor *Classroom management and climate* in CVCP and ECP factors of CFA models for ISTOF items**

CVCP Factor (N=76)	Indicator No.	Item No.	Item Description (T = The teacher; S = student)	ECP Factor (N=79)	Indicator No.	Item No.	Item Description (T = The teacher; S= student)
Classroom management and climate (Cronbach's alpha = 0.95)	IND61	**Item 31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	**Positive classroom management (Cronbach's alpha = 0.82)**			
	IND71	***Item 41***	*Actions are taken to minimize disruption.*		*IND71*	***Item 41***	*Actions are taken to minimize disruption.*
	IND72	**Item 42**	There is clarity about when and how Ss can get help to do their work in class.				
	IND73	***Item 44***	*T corrects misbehaviour with measures that fit the seriousness of the misconduct.*		*IND73*	***Item 44***	*T corrects misbehaviour with measures that fit the seriousness of the misconduct.*
	IND73	***Item 45***	*T deals with misbe-haviour and disrup-tions by referring to the established rules of the classroom.*		*IND73*	***Item 45***	*T deals with misbe-haviour and disrup-tions by referring to the established rules of the classroom.*

The inclusion of Item 31 seems to suggest that for the lessons of Hong Kong case study sample, classroom management is more closely related to classroom climate, though this factor is dominant by three items concerning classroom management. The inclusion of Item 42 also suggests that classroom management is not just purely concerning misconduct or disruption in class, because it is likely that students are less prone to disruptive behaviours if they can seek help and feel the genuine warmth and empathy from their teachers. This kind of classroom management is not just positive but proactive as it acts or takes effect before troubles begin. In the lessons observed, it was rare that the teachers had to deal with misbehaviours and disruptions frequently in a lesson, especially in a lesson in which classroom climate was warm and supportive. High ratings were also awarded to teachers who had managed to have no disruptions or misbehaviours in their students. Interpretations of high ratings are better understood together with qualitative notes as will be discussed in Chapter 7. This factor received the highest reliability (Cronbach's alpha= 0.95) among the six factors, strongly confirming the coherence of the items.

Factor Differentiation and support

As shown in Table 4.20 below, the third factor, *Differentiation and support*, seems to be most similar to the fourth factor of the ECP item-based model, *Teacher strategies with respect to teacher expectations*, as they share two items (i.e., Item 7 and Item 43).

Table 4.20: **Shared items on Factor *Differentiation and support* in CVCP and ECP factors of CFA models for ISTOF items**

CVCP Factor N= 76	**Indicator No.**	**Item No.**	**Item Description** (T = The teacher; S= student)	**ECP Factor** (N=79)	**Indicator No.**	**Item No.**	**Item Description** (T = The teacher; S= student)
Differentiation and support (Cronbach's alpha = 0.94)	*IND22*	***Item 7***	*T makes a distinction in the scope of the assignments for different groups of Ss.*	**Teacher strategies with respect to teacher expectations (Cronbach's alpha = 0.74)**	*IND22*	***Item 7***	*T makes a distinction in the scope of the assignments for different groups of Ss.*
	IND22	**Item 8**	T gives additional opportunities for practice to Ss who need them.				
	IND41	**Item 15**	T provides sufficient wait time & response strategies to involve all types of learners.		IND43	**Item 20**	T uses different, appropriate instruc-tional strategies for different groups of Ss.
	IND52	**Item 25**	T gives Ss the opportunity to correct their own work.				
	IND63	**Item 35**	T gives turns to and/ or involves those Ss who do not voluntarily participate in classroom activities.		IND64	**Item 37**	T praises children for effort towards realizing their potential.
	IND72	***Item 43***	*There is clarity about what options are available when the Ss finish their assignments.*		*IND72*	***Item 43***	*There is clarity about what options are available when the Ss finish their assignments.*

Although items of this factor originally belong to different components in the original scale, their meanings are clearly related to the kinds of differentiation and support strategies that a teacher may adopt in the classroom. Item 7 and Item 8 belong to Component Two (*Differentiation and inclusion*), indicating the extent to which the teacher takes full account of student differences. However, it seems that compared to their English counterparts, Hong Kong EFL lessons in the sample were less likely to involve different teaching strategies for different groups of students as Item 20 was not retained in the current model. In contrast, varying the due times for handing assignments appears to be a key differentiation strategy for both Hong Kong and English lessons in CVCP and ECP.

In order to achieve differentiation and inclusion, Hong Kong teachers in the sample may have to adjust their instructional skills such as wait time and response strategies (i.e., Item 15). These teachers were more often observed letting students to be independent active learners who can correct their own work (i.e., Item 25). In the lessons observed, Hong Kong teachers might be more keen on motivating passive students in class (i.e., Item 35), while in the English lessons, praise might be more often used to motivate students (i.e., Item 37). This may again reflect some cultural differences in classroom practices. In terms of reliability, the CVCP factor shows stronger internal consistency than the ECP factor (Cronbach's alpha=0.94 vs 0.74).

Factor Clarity and logic of presentation

The fourth factor, *Clarity and logic of presentation*, reflects how the presentation or lesson is structured and related to the teacher's questioning and feedback to the students. Table 4.21 below shows that this factor is most comparable with the first factor of the ECP item-based model.

Table 4.21: **Shared items on Factor *Clarity and logic of presentation* in CVCP and ECP factors of CFA models for ISTOF items**

CVCP Factor N= 76	**Indicator No.**	**Item No.**	**Item Description (T = The teacher; S= student)**	**ECP Factor** (N=79)	**Indicator No.**	**Item No.**	**Item Description (T = The teacher; S= student)**
Clarity and logic of presentation (Cronbach's alpha = 0.91)	IND11	**Item 2**	T provides appropriate feedback to the answers given by Ss.	**Clear and coherent lesson in a supportive learning climate (Cronbach's = 0.84)**			
	IND33	**Item 13**	T presents the lesson with a logical flow that moves from simple to more complex concepts.		IND32	**Item 10**	T communicates in a clear and understandable manner.
	IND33	***Item 14***	*T implements the lesson smoothly moving from one stage to another with*		*IND33*	***Item 14***	*T implements the lesson smoothly moving from one stage to another with*

CVCP Factor N= 76	Indicator No.	Item No.	Item Description (T = The teacher; S= student)	ECP Factor (N=79)	Indicator No.	Item No.	Item Description (T = The teacher; S= student)
			well-managed transition points.				*well-managed transition points.*
	IND42	**Item 18**	The length of the pause following questions varies according to the difficulty level of questions		IND61	**Item 31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.
					IND61	**Item 32**	T shows respect for the Ss in both in his/her behaviour and use of language.

Item 13 and Item 14 are items indicating how well lessons are structured and belong to the Component Three, *Clarity of Instruction,* in the original scale. Smooth transitions and logical flows from basic to complex concepts are key characteristics of a well-structured lesson in rating lessons of the Hong Kong sample. Item 2 and Item 18 come from two other different components (*Assessment and evaluation* and *Instructional skills*, respectively) in the original theoretical model, but they are clearly related to questioning and feedback skills of teachers. It should be noted that the two factors of these two models only share Item 14. The reliability of the current CVCP model is slightly higher than that of the ECP factor (Cronbach's alpha=0.91 vs 0.84). Item 10 in the ECP factor is also in the same component in the original theoretical model as Item 13 and Item 14, but it is assumed to be related to communication skills, rather than logical structuring of presentation. The ECP factor also includes items concerning the extent to which students are valued in a supportive classroom climate (i.e., Item 31 and Item 32). This suggests that in the English lessons the factor *Classroom Climate* was more likely to be linked with the factor *Clarity of Instruction* than the factor *Classroom management* as was found in their Hong Kong counterparts.

Factor Student engagement

Table 4.22 below shows the items of the fifth factor, *Student engagement.* Despite its items originating from the different components in the original, this factor is clearly addressing how the teacher may enhance students' engagement in the learning activities. Its high reliability score suggests that these seven items are strongly internally consistent. This is not surprising as several items contain words like *engage* or *involvement*, as highlighted in Table 4.22.

Table 4.22: **Shared items on Factor *Student engagement* in CVCP and ECP factors of CFA models for ISTOF items**

CVCP Factor N= 76	Indicator No	Item No.	Item Description (T = The teacher; S= student)	ECP Factor N=79	Indicator No	Item No.	Item Description (T = The teacher; S= student)
Student engagement (Cronbach's alpha = 0.94)	IND21	***Item 5***	Ss ***communicate frequently*** with one another ***on task-oriented issues.***	**Engaging students with assignments and activities (Cronbach's alpha = 0.79)**	IND21	***Item 5***	Ss ***communicate frequently*** with one another ***on task-oriented issues***.
	IND21	***Item 6***	*All Ss are actively **engaged** in learning.*		*IND21*	***Item 6***	*All Ss are actively **engaged** in learning.*
	IND41	***Item 16***	*T gives assignments that stimulate all Ss to active **involvement.***		*IND41*	***Item 16***	*T gives assignments that stimulate all Ss to active **involvement.***
	IND54	**Item 29**	T systematically uses material and examples from the Ss' daily life to illustrate the course content.				
	IND54	**Item 30**	Ss are invited to give their own examples.				
	IND62	**Item 33**	T creates purposeful activities that ***engage*** every S in productive work.				
	IND63	**Item 36**	T seeks to ***engage*** all Ss in classroom activities.				

In the original scale, Item 16 is hypothesised to indicate the teacher's instructional skill in engaging students, while Item 5 and Item 6 of *Differentiation and Support* (Component Two) refer to the extent to which the teacher can create an environment in which all students are involved. While Item 29 and Item 30 of *Promoting active learning and developing meta-cognitive skills* (Component Five) specify that the teacher engage students through connecting learning materials to students' real world experiences, Item 30 and Item 36 of *Classroom Climate* (Component Six) respectively specify the teacher's initiation in engaging students and the extent of the student engagement. Table 4.22 shows that this factor consists of all items of the second factor of the ECP model. The extra items that the ECP factor lacks suggest that engagement can be achieved through linking learning with life experiences and promoted in participatory classroom climate. Thus, the inclusion of these items in this factor enriches our understanding of engagement.

Factor Strategies to enhance to learning and lesson focus

As indicated in Table 4.23, the last factor of the current CVCP model consists of three items only. The CVCP factor is also considered more preferable than the ECP factor, because it is generally more desirable to have a factor with at least three items,

Table 4.23: **Shared items on Factor Strategies to enhance to learning and lesson focus in CVCP and ECP factors of CFA models for ISTOF items**

CVCP Factor N=76	Indicator No	Item No.	Item Description (T = The teacher; S= student)	ECP Factor N=79	Indicator No	Item No.	Item Description (T = The teacher; S= student)
Strategies to Enhance Learning and Lesson Focus. (Cronbach's alpha = 0.70)	*IND22*	***Item 4***	*T explains how assignments are aligned to the learning goals of the lesson.*	**Purposive learning (Cronbach's alpha =0.79)**	*IND22*	***Item 4***	*T explains how assignments are aligned to the learning goals of the lesson.*
	IND32	***Item 11***	*T clarifies the lesson objectives at the start of the lesson.*		*IND32*	***Item 11***	*T clarifies the lesson objectives at the start of the lesson.*
	IND64	**Item 37**	T praises children for effort towards realizing their potential.				

Again, all items belong to different components in the original scale. Its reliability is the lowest among all factors, even lower than that of the ECP factor (Cronbach's alpha=0.70 vs 0.74), but it is still acceptable. What is striking is that results of both samples confirmed that Item 4 and Item 11 were empirically linked, despite their origins in different theoretical components in the current ISTOF instrument. Nevertheless, the relationship of Item 37, the extra item that does not appear in the ECP factor, with the other two items is not obvious. This may suggest that praise in the Hong Kong context is goal directed.

Based on the above results, it seems that the six-factor CVCP model is more distinctive than the ECP model discussed earlier in three dimensions. First, as ratings on teacher practices were supposed to be based on items, rather than indicators, a model based on items would reflect dimensions of teacher practices that were more specific. An indicator–based or component-based model may require taking the average weight for all the items of an indicator or a component as suggested in the theoretically driven underlying structure of the instrument. Factors of the present six factor item-based model are coherent in their meanings as reflected in their high reliability scores that indicate high internal consistency. Five factors of the present model have a Cronbach's alpha above 0.9, while none of the factors of the ECP item-based model yielded that high level of internal consistency. Second, the ECP indicator-based model was only more preferable as it took into account of Component Five (i.e., *Promoting active learning and developing metacognitive skills*), while the ECP item-based model was built on multiply-imputed data as most items were deleted due to the presence of a significant number of missing data. Third, except the last factor, all items of the remaining five factors have moderately high loadings, suggesting that the relative weights of these factors in the model may be similar.

4.8 Conclusion

This chapter has described the quantitative classroom observation data collected via the ISTOF instrument and variation across the 76 lessons observed for the four Hong Kong teachers. Results obtained by the confirmatory factor analyses strongly support the view that there are several underlying dimensions in observed teachers' teaching behaviours. Underlying dimensions like clear objective and lesson focus, clear instruction and presentation, effective classroom management, positive lesson climate, differentiation and supportive teaching strategies, meta-cognitive skills teaching and engaging learning activities seem to be important dimensions of effective classroom practices that show the qualities of "basic, generic and replicable" variables (Teddlie et al., 2006). These results provide the ground for further analyses in Chapter 7 for identifying the relative impacts of these dimensions on overall judgment of teaching quality made by the rater and on the individual involvement of students found in the lesson observed by the rater.

While *effective teaching is most likely a multidimensional process*, it is also clear that *different dimensions of teaching behaviours may vary considerably across lessons*. Thus, in Chapter 7, the same identified underlying dimensions of effective classroom practices are used to compare variation among teachers as well as to identify exemplar lessons that were rated highly and lowly along these dimensions. By integrating these new quantitative results with qualitative findings in the field notes and interviews, knowledge about variation among teachers and its impact on departmental and school effectiveness can be enriched.

The current CVCP (ISTOF) Model is well supported by the data. However, results in Section 4.6 indicate that the data also showed considerable support for the seven-factor theoretical model. While obtaining goodness-of-fit for both models is an exciting finding in instrument development, the empirical model seems to better than the theoretical model as it has better convergent validity. However, it still poses a problem of selecting between them because both models were weak in terms of discriminant validity. Thus, results of further analyses that explored their relative strengths will be presented in Chapter 6.

The results of using QoT as observation instrument are presented in the next chapter. The two sets of results obtained from different instruments are intended to be related as both instruments were used on the same occasions in every lesson observed. Results of this chapter and the next chapter illuminate the issues surrounding instrument variation that are to be addressed in depth in Chapter 6. In that chapter, the integrated findings allow us to pursue not only the validity of a single instrument, but also the relative strengths and weaknesses of the instruments if both seem to be ecologically valid and capable of identifying underlying dimensions of effective classroom practices from the observed teachers' behaviours.

Finally, the similarities between factors of the CVCP (ISTOF) Model and those of the ECP studies suggest that the two models can be cross-validated. Cross-validation is usually done to examine the validity of a model across samples, but given the different characteristics of the two samples, it is likely that the two models may not be well supported by the data of another sample. However, it would be of interest to examine the extent to which the two models are supported in the data of different samples because it is also hypothesised that some common dimensions should underlie the teachers' behaviours of both samples.

In the next chapter, results of the descriptive and inferential statistical analyses performed on the quantitative classroom observation data collected using another instrument, QoT, will be reported. Both sets of results will be compared and related to overall indicators of teaching effectiveness in

Chapter 6 and used to characterise the consistency and variation in the observed teaching practices of individual teachers in Chapter 7.

CHAPTER 5 : DIMENSIONS OF AND CONSISTENCY AND VARIATION IN CLASSROOM PRACTICES USING QOT AS THE OBSERVATION INSTRUMENT

5.1 Introduction

The results of systematic classroom observation presented in this chapter are based on a high-inference instrument developed by van de Grift and his colleagues as a tool that can be used in school inspection in different countries (van de Grift et al., 2004, van de Grift, 2007; see Section 3.5.3 for detail). Unlike in the ECP study discussed in Chapter 4, this instrument was used on *the same* occasions as another instrument, ISTOF, the results of which were presented in the last chapter. Using two instruments at the same time allow us to explore *whether similar dimensions of teaching behaviours can be identified as effective classroom practices* despite instrument variation. The employment of two different instruments developed for different purposes certainly would contribute to our knowledge on systematic classroom observation because, as discussed in Chapter 2, instruments have rarely been compared and studied in the past. Issues surrounding instrument variation will be addressed more fully in Chapter 6, but results presented here do highlight some similarities as the discussion moves on. Unlike the ECP study, the unit of analysis is the *lesson*, rather than the *teacher,* because only four teachers were observed. Thus, it should be noted that variation found in the results presented here reflects variation *across lessons* but not always necessarily across teachers. Variation in individual teacher's lessons and variation between these teachers are to be explored in detail in the within-case analyses in Chapter 7 and the cross-case analyses in Chapter 8.

As in the last chapter, five sets of results of systematic observations using QoT as the instrument are presented in the following sections. First, in Section 5.2 the general patterns emerged in descriptive statistics of the QoT indicators are discussed with a focus on those indicators with highest means, lowest means and their frequency distributions. Differences in teaching effectiveness in the two samples may reflect the nature of the sample selected in the two studies. In particular, English lessons in the ECP sample demonstrated more effective teaching behaviours than the Hong Kong lessons (as rated in accordance with the instrument's criteria). In general, most of the effective classroom practices were *less often* observed or found with *less strength or effectiveness* in the lessons of the Hong Kong sample than in those of the English sample.

Second, Sections 5.3 and 5.4 respectively discuss the underlying factors generated in the exploratory and subsequent confirmatory analyses from the ratings obtained using QoT as the observation instrument. These results serve the purpose to illustrate some generic characteristics of effective teaching practices in the Hong Kong sample just like the English sample. It is assumed that exploratory and confirmatory analyses would indicate what underlying factors can characterise the effective classroom practices observed in the lessons of the EFL teachers in Hong Kong. These characteristics are expected to reveal those "basic, generic and replicable" variables (Teddlie et al. 2006) that would contribute to the generic model of teaching or educational effectiveness. A three-factor CFA model seemed to be well supported by the data.

Third, the frequency distributions of the underlying factors of the CFA model found for the Hong Kong sample on QoT data provide some preliminary evidence indicative of differentiated teaching effectiveness. These results presented in Section 5.5 show across individual lessons how teachers' behaviours varied in different dimensions of classroom practices, as they were rated more effective and more often in some but not in the others and thus might reflect where their strengths and weaknesses lied. The magnitudes of variation in these underlying factors across lessons contribute to our understanding of differential teaching effectiveness of the Hong Kong sample. In general, Hong Kong lessons seemed to show strengths in maintaining learning environment and climate orderly and inclusive, structuring lessons with clear instruction, interaction and student participation, and planning the lesson effectively to ensure its objectives are accomplished, but they also tended to be weaker in making the classroom layout effective.

Fourth, Section 5.6 presents and discusses the extent to which the resulted three-factor CFA model could be comparable to a model that shares the hypothetical factor structure with nine theoretical criteria. Two CFA models are introduced, compared and discussed. As one of these two models was based on a recent paper by van de Grift (2007), the major developer of the instrument, it would provide a solution for further exploratory analyses in the next chapter. It is also argued that the three-factor empirical model in Section 5.4 and the rival models presented here seem to reveal that some classroom practices are fundamental, effective practices despite the presence of discrepancies between these models.

Finally, CFA models in the CVCP and the ECP studies in Section 5.7 were compared factor by factor. These comparisons are expected to reveal that there were some generic characteristics of effective teaching practices in these two samples. The factors were strikingly similar except that the CVCP factors seemed to be integrating two or more ECP factors. In other words, more distinctive factors were identified in the ECP model, while CVCP factors tended to cluster together as larger factors than were in the ECP ones. These results suggested factors in the samples were globally alike, but discrepancies might reflect some idiosyncratic properties of the samples.

5.2 Features of strengths and weaknesses in Hong Kong teachers' observed classroom practices and comparisons with the ECP results

Features of strengths and weaknesses in Hong Kong teachers' observed classroom practices can be seen in an excerpt of the descriptive statistics in Table 5.1 (for a full table, see Appendix IX). In the table, 10 of 26 indicators (about 38.5%) are shown with a mean over 3.0, ranging between 3.01 and 3.36. That is, in most lessons, the teacher was rated as strong or effective with respect to the classroom practices these indicators are in concern. These indicators belong mainly to three theoretical criteria[51] of the original scale *Safe and Orderly School Climate* (Criterion 1), *Clear Instruction* (Criterion 4), and *Effective Classroom Organisation* (Criterion 8). In contrast,

51 The theoretical criterion of an indicator is indicated by the first digit of the indicator number. For example, Indicator 11 belongs to Criterion One, while Indicator 31 belongs to Criterion Three.

the number of indicators with a mean below 2.5 (indicating more weaknesses than strengths in the teacher's respective classroom practice) was small. Only two indicators of this kind were found: Indicators 91 and 92 of Criterion 9 (*Effective Classroom Layout*). Among the indicators with a mean above 3, eight had a statistically significant negatively skewed distribution[52], while one had a statistically significant positive kurtosis[53]. *Negative* skewness in these indicators means that the classroom practices as described in these indicators were frequently found *effective*, while a *positive* kurtosis (or leptokurtic distribution) suggests *less* variability across lessons.

Table 5.1: **Features and ranking (descending) of strengths and weaknesses in Hong Kong teachers' observed classroom practices in the CVCP study using the QoT instrument**

The ten Indicators with the highest means in the CVCP Study (N=76)								
Indicat or No.	**Indicator Description (T = The teacher; P = pupil)**	**Mean**	**Std. Deviation**	**Skew-ness**	**z-Skew-ness**	**Kurtosis**	**z-Kurtosis**
IND43	*T gives feedback to Ps*	3.36	0.69	-0.85	*3.09**	0.61	1.12
IND84	T ensures effective classroom management	3.28	0.87	-1.07	*3.88**	0.40	0.73
IND41	*T gives clear instructions and explanations*	3.25	0.66	-0.60	*2.18**	0.70	1.29
IND42	T gives clear explanations of the learning materials and the assignments	3.18	0.74	-0.51	1.86	-0.34	0.62
IND14	*T shows respect for the Ps in behaviour and language use*	3.14	0.80	-.092	*3.34**	0.87	1.60
IND83	T uses learning time efficiently	3.13	0.82	-0.84	*3.06**	0.44	0.81
IND81	T ensures the orderly progression of the lesson	3.09	0.82	-.062	*2.26**	-0.12	0.23
IND11	*T ensures a relaxed atmosphere*	3.08	0.83	-0.58	*2.11**	-0.26	0.47
IND13	*T supports the self-confidence of Ps*	3.01	0.90	-0.59	*2.13**	-0.45	0.82
IND31	T clarifies the lesson objectives at the start of the lesson	3.01	0.55	-0.48	1.73	2.30	*4.23**
Indicators with a mean below 2.5 in the CVCP Study (N=76)							
IND92	*The teaching environment is educational and contemporary*	2.32	0.77	0.29	1.04	-0.13	0.24
IND91	*T ensures that classroom layout supports the P activities*	2.21	0.72	1.00	*3.61**	1.28	*2.35**
IND100	I judge the overall quality of teaching as:	2.93	0.74	-0.72	*2.61**	0.92	1.70

Similar patterns revealed in the characteristics of the descriptive statistics revealed in the ISTOF ratings can also be found for the QoT ratings. First, in general, there were more indicators (22 indicators or 84.6% in total) with means above 3 in the ECP study than the CVCP study and their means were relatively higher. In Table 5.2, which partially shows the descriptive statistics of the indicators in the ECP study, the ten highest means were

52 The value z-skewness was obtained by dividing the skewness value by its standard error. When this z-score of the skewness value (|ske| /s.e.) is above 1.96 it is statistically significant, as indicated by an asterisk in Table 5.1.

53 The value z-kurtosis was obtained by dividing the skewness value by its standard error. When this z-score of the kurtosis value (|β_2| /s.e.) is above 1.96 it is statistically significant, as indicated by an asterisk in Table 5.1.

about 14.3% to 22.6% higher than those in the CVCP study. Unsurprisingly, no indicators were found to have a mean below 2.5 for the purposive sample known for their higher teaching effectiveness in the ECP research. Only four indicators in the ECP study had a mean below 3, but still above 2.5: Indicators 61, 62, 71 and 72, indicating that lessons in the ECP sample were rated as having more strengths than weaknesses in these indicators.

Table 5.2: **Features and ranking (descending) of strengths and weaknesses in English teachers' observed classroom practices in the ECP study using the QoT instrument**

Ten Indicators with the highest means in the ECP study (N=79)								
Indicator No.	**Indicator Description (T = The teacher; P = pupil)**	**Mean**	**Std. Deviation**	**Skew-ness**	**z-Skew-ness**	**Kurtosis**	**z-Kurtosis**
IND13	*supports the self-confidence of Ps*	3.84	0.44	-2.91	*10.16**	8.25	*14.57**
IND14	*shows respect for the Ps in behaviour and language*	3.84	0.40	-2.58	*9.00**	6.41	*11.32**
IND41	*gives clear instructions and explanations*	3.81	0.43	-2.21	*7.71**	4.34	*7.67**
IND43	*gives feedback to Ps*	3.80	0.44	-2.05	*7.16**	3.56	*6.29**
IND12	promotes mutual respect	3.74	0.50	-1.82	*6.35**	2.59	*4.57**
IND82	ensures the orderly progression of the lesson	3.71	0.57	-1.89	*6.59**	2.63	*4.64**
IND91	*ensures the classroom layout supports the P activities*	3.71	0.64	-2.74	*9.54**	8.22	*14.51**
IND92	*the teaching environment is educational and contemporary*	3.71	0.70	-2.85	*9.95**	8.02	*14.16**
IND73	provides interactive instruction and activities	3.71	0.52	-1.58	*5.49**	1.69	*2.97**
IND11	*ensures a relaxed atmosphere*	3.69	0.58	-1.70	*5.93**	1.94	*3.42**
Indicators with a mean below 3 in the ECP study (N=79)							
IND71	ensure that the teaching materials are orientated towards transfer	2.89	0.98	-0.76	*2.33**	-0.30	0.47
IND61	adapts the instruction to the relevant differences between Ps	2.86	1.08	-0.70	*2.44**	-0.74	1.31
IND62	adapts the assignments and processing to the relevant difference between Ps	2.73	1.05	-0.36	1.26	-1.03	1.82
IND72	stimulates the use of control activities	2.72	0.96	-0.47	1.44	-0.62	0.98
IND100	I judge the overall quality of teaching as:	3.71	0.57	-1.89	*6.59**	2.63	*4.64**

Second, Indicators 11, 13, and 14 of *Safe and Orderly School Climate* (Criterion 1) and Indicators 41 and 43 of *Clear Instruction* (Criterion 4) appeared to be the strengths of the lessons in both samples as they are among in the top ten highest means in both Table 5.1 and Table 5.2. These indicators may represent dimensions of teaching behaviours that characterise effective teaching. The Hong Kong lessons in the sample also seemed to show more strength in *Effective Classroom Organisation* (Criterion 8). As Hong Kong teachers usually have a tight teaching schedule and a common curriculum and scheme of work to follow, they have to pay more attention to structure their lessons to ensure the learning time is sufficient and material is well-covered. This cannot be done without effective classroom management, especially in a bigger average class size than is typical for the English sample.

In contrast, indicators of Criterion 9 (*Effective Classroom Layout*) seemed to distinguish the two samples as this was the weakest dimension of the Hong Kong lessons in the CVCP study, but among the strongest dimensions for the English lessons in the ECP study. This distinction probably reflects a cultural difference in classroom practices, rather than a difference in teaching effectiveness between two samples because subject teachers in Hong Kong generally do not have any control over the classroom layout as their English counterparts do. In Hong Kong, schools are designed such that students are not expected to move around to rooms where their teachers station most of the time. Instead, teachers have to move around to different classrooms for different classes. Sometimes both teachers and students have to go to different classrooms for different days in a teaching cycle because there are always not enough classrooms and they have to go to whatever classrooms are available. Due to this movement, teachers are expected not to change any layout of the classroom. In case of they need to change anything, they have to ensure that everything is going to revert back to its original setting at the end of the lesson.

Third, as in the case of ISTOF items, QoT indicators in the ECP study not only had higher means, but also statistically more significant negatively skewed and leptokurtic distributions (all with $p<0.0001$), suggesting little variation between the observed classroom practices in the lessons of effective teachers in the English sample. This pattern also seems to be applicable to those indicators with lower means. Among the four indicators with a mean below 3, Indicators 61 and 71 had a significantly negatively skewed distribution. Since there were more missing data (11 cases for Indicators 61 and 62, but 27 cases for Indicators 71 and 72), caution may be required for any interpretation on these results. However, this overall pattern shows that the English lessons in the ECP study were not only more effective (in terms of high scores) than the Hong Kong lessons in terms of these classroom practices but also tended to show little variation between them with regards to these practices. In contrast, lessons of the Hong Kong sample were not rated as effective as those of the English sample and variability across lessons was larger as kurtosis tended to be normal more often.

Finally, a clear overall greater magnitude of strengths of the English lessons in the sample in comparison with the Hong Kong lessons in the sample can be identified in their frequency distribution of scores. Except for Indicators 43 and 84, the indicators with the highest means in the CVCP study in Table 5.3 indicate that in the majority of the lessons observed, Hong Kong sample showed more strengths than weaknesses or predominant strengths in those classroom practices. Thus, overall in most of the lessons observed, the overall quality of teaching showed more strengths than weaknesses in these areas. However, these teachers tended to show more weaknesses than strengths in indicators with the lowest means (i.e., Indicators 91 and 92).

Table 5.3: **Variability identified in the of distribution of strengths and weaknesses in Hong Kong teachers' ranked observed classroom practices in the CVCP study (N=76) using the QoT instrument (A full table is available in Appendix X)**

Indicator No.	Indicator Description (T = The teacher; P = pupil)	Predominately weak	More weaknesses than strengths	More strengths than weaknesses	Predominantly strong
IND43	*T gives feedback to Ps*	1.32%	7.89%	*44.74%*	***46.05%***
IND84	T ensures effective classroom management	5.26%	11.84%	*32.89%*	***50.00%***
IND41	*T gives clear instructions and explanations*	1.32%	7.89%	***55.26%***	*35.53%*
IND42	T gives clear explanation of the learning materials and the assignments	1.32%	15.79%	***46.05%***	*36.84%*
IND14	*T shows respect for the Ps in behaviour and language use*	5.26%	9.21%	***51.32%***	*34.21%*
IND83	T uses learning time efficiently	5.26%	11.84%	***47.37%***	*35.53%*
IND81	T ensures the orderly progression of the lesson	3.95%	22.37%	***50.00%***	*23.68%*
IND11	*T ensures a relaxed atmosphere*	3.95%	18.42%	***43.42%***	*34.21%*
IND13	*T supports the self-confidence of Ps*	6.58%	19.74%	***39.47%***	*34.21%*
IND31	T clarifies the lesson objectives at the start of the lesson	1.32%	10.53%	***73.68%***	*14.47%*
IND92	*The teaching environment is educational and contemporary*	*11.84%*	***51.32%***	30.26%	6.58%
IND91	*T ensures that classroom layout supports the P activities*	*9.21%*	***68.42%***	14.47%	7.89%
IND100	I judge the overall quality of teaching as:	5.26%	14.47%	***61.84%***	*18.42%*

In the ECP study, for those indicators with the highest means in Table 5.4, more lessons were identified as *predominantly strong* in those areas, rather than just more strengths than weaknesses. The English lessons in the sample seemed to show more effective behaviours than those in the Hong Kong sample as evident in the overwhelming majority of lessons and teachers (74% to 87%) observed that showed the highest quality of teaching. The number of lessons in which the teachers showed more weaknesses than strengths or predominantly weak was only small and negligible (ranging from 1.4% to 7.3%). The difference in the *overall judgment of teaching quality of the lesson* (Indicator 100) in the two tables clearly confirms the strengths of the English sample in teaching effectiveness. In the majority (about 77%) of the lessons observed, the English lessons in the ECP sample showed predominantly strong ratings in overall teaching quality, while in the majority (62%) of the lessons in the CVCP study, the Hong Kong lessons were found with more strengths than weaknesses only.

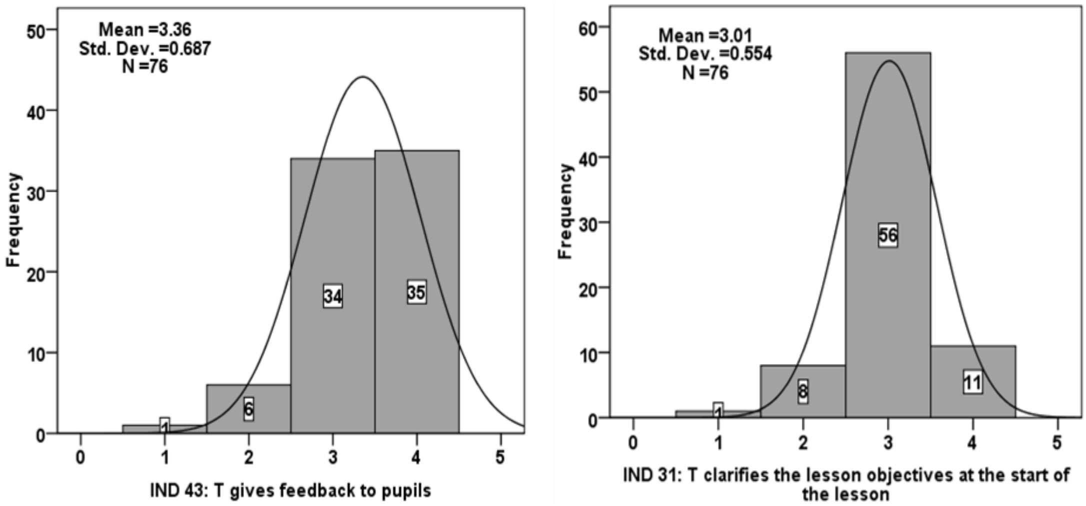
Table 5.4: **Variability identified in the distribution of strengths and weaknesses in English teachers' ranked observed classroom practices in the ECP study (N=79) using the QoT instrument**

Indicator No.	Indicator Description (T = The teacher; P = pupil)	Predominately weak	More weaknesses than strengths	More strengths than weaknesses	Predominantly strong
IND13	*supports the self-confidence of Ps*	0.00%	2.86%	*10.00%*	***87.14%***
IND14	*shows respect for the Ps in behaviour and language*	0.00%	1.43%	*12.86%*	***85.71%***
IND41	*gives clear instructions and explanations*	0.00%	1.43%	*15.71%*	***82.86%***
IND43	*gives feedback to Ps*	0.00%	1.43%	*17.14%*	***81.43%***
IND12	promotes mutual respect	0.00%	2.86%	*20.00%*	***77.14%***
IND73	provides interactive instruction and activities	0.00%	2.90%	*23.19%*	***73.91%***
IND82	ensures the orderly progression of the lesson	0.00%	5.71%	*17.14%*	***77.14%***
IND91	*ensures the classroom layout supports the P activities*	2.86%	1.43%	*17.14%*	***78.57%***
IND92	*the teaching environment is educational and contemporary*	4.29%	1.43%	*12.86%*	***81.43%***
IND11	*ensures a relaxed atmosphere*	5.80%	7.25%	*7.25%*	***79.71%***
IND100	I judge the overall quality of teaching as:	0.00%	5.71%	*17.14%*	***77.14%***

Frequency distributions of individual items often reveal great variation in variability in different observed teachers' teaching behaviour. In both the CVCP and ECP studies, there was no bimodal distribution for the QoT results, probably because there is no neutral rating in the instrument.

Figure 5.1 shows the frequency distributions of two of the indicators with the highest means in Table 5.1. Both indicators had a negative skewness value and positive kurtosis value in Table 5.1. However, while Indicator 43 had a statistically significant negatively skewed (p<0.005) but statistically insignificant leptokurtic distribution, Indicator 31 a statistically significant leptokurtic (p<0.05) but statistically insignificant negatively skewed distribution.

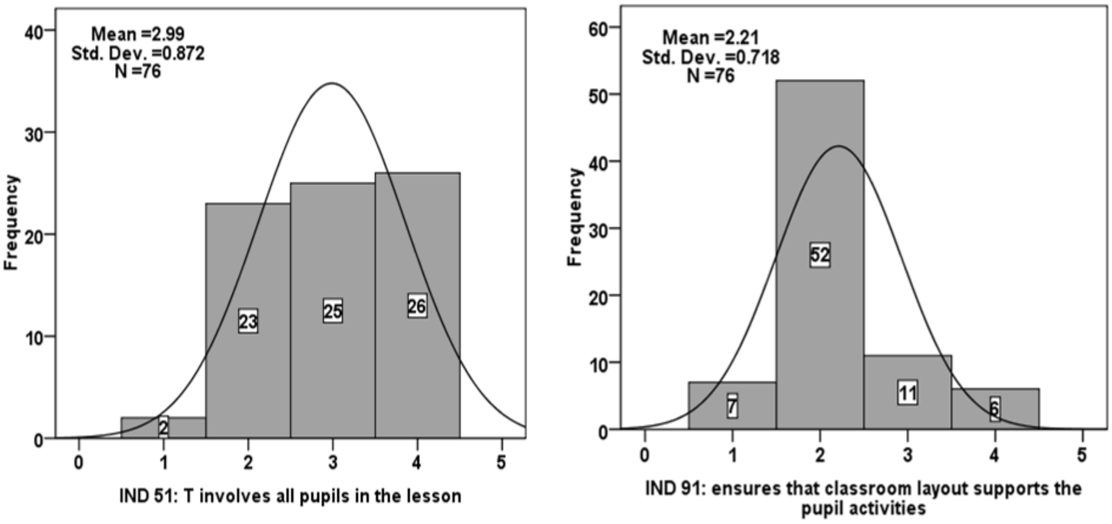
Figure 5.1: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of Indicator 43 & Indicator 31**



Skewness = -0.85; Kurtosis = 0.61 **Skewness = -0.48; Kurtosis = 2.30**

Figure 5.2 below shows two rare cases in the frequency distributions of QoT indicators.

Figure 5.2: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distributions of Indicator 51 & Indicator 91**



Skewness = -0.22; Kurtosis = -1.11 **Skewness = 1.00; Kurtosis = 1.28**

Indicator 51 had a statistically insignificant negatively skewed but statistically significant platykurtic ($p<0.05$) distribution, but Indicator 91 had a statistically significant positively skewed ($p<0.005$) and statistically significant leptokurtic ($p<0.05$) distribution. The distribution of Indicator 51 indicated that in most lessons observed, teachers might be rated as having more strengths than weaknesses and variability across lessons was very likely to be large as there were about equal number of lessons with different ratings. In contrast, the distribution of Indicator 91 revealed a case where teachers were very likely to be found with more weaknesses than strengths and variability across lessons was very likely to be small because only 24 out of 76 lessons (about 32%) had a different rating. This was exactly an opposite case of Indicator 31.

5.3 Preliminary dimensions of effective classroom practices identified in a three-factor model: the results of the exploratory factor analysis

As in the analysis for ISTOF, the EFA analysis performed was also based on the principal component analysis extraction with varimax rotation specified in DATA REDUCTION in SPSS16. As shown in Table 5.5, the three factors generated accounted for almost 80% of the total variances, suggesting that they can be reliably used to build a CFA model. Again, no imputation was required as there were no missing data.

Table 5.5: **Relative importance of EFA factors of QoT indicators as identified in the variances they accounted for in the CVCP study (N=76)**

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**	**Total**	**% of Variance**	**Cumulative %**
1	16.96	65.23	65.23	16.96	65.23	65.23	9.45	36.36	36.36
2	2.76	10.60	75.83	2.76	10.60	75.83	7.09	27.25	63.62
3	1.03	3.94	79.77	1.03	3.94	79.77	4.20	16.15	79.77

In contrast with the six-factor model of ISTOF, the current QoT model shows two dominant factors, *Integrated class management and climate* and *Structured teaching skills*. The first factor, which consists of all the indicators of the two theoretical domains concerning lesson climate, namely, *Safe and orderly school climate* and *Stimulating learning climate*, is comparable to the second factor of ISTOF. It seems that both ratings using ISTOF and QoT showed that classroom management and classroom climate were highly related dimensions of the classroom practices of the four EFL teachers.

The EFA factors of the QoT model in Table 5.6 below are well defined by the indicators in the factors. Both loadings of indicators of factors and their communalities are high. The overall factor loadings of the 26 indicators are mostly high. No indicator has loadings below 0.5. Except Indicators 61 and 41, all indicators have loadings ranging from moderate to high (λ between 0.60 and 0.90). Except Indicators 31 and 41, whose communities are slightly below the high level (h^2 = 0.68 and 0.69, respectively), all indicators have high communalities (with h^2 between 0.70 and 0.92), suggesting that a strong stability for the model that built on these indicators.

Unlike the case in ISTOF, there were far fewer parameters to be analysed with only 26 indicators and 3 factors. This meant that there was no pressure to reduce the number of indicators or factors in the subsequent confirmatory factor analysis. All indicators were retained, even though many indicators showed cross-loadings in the EFA model.

Table 5.6: **Preliminary dimensions of effective classroom practices identified in the EFA factors of QoT indicators and their loadings (above 0.4) in the CVCP study (N=79)**

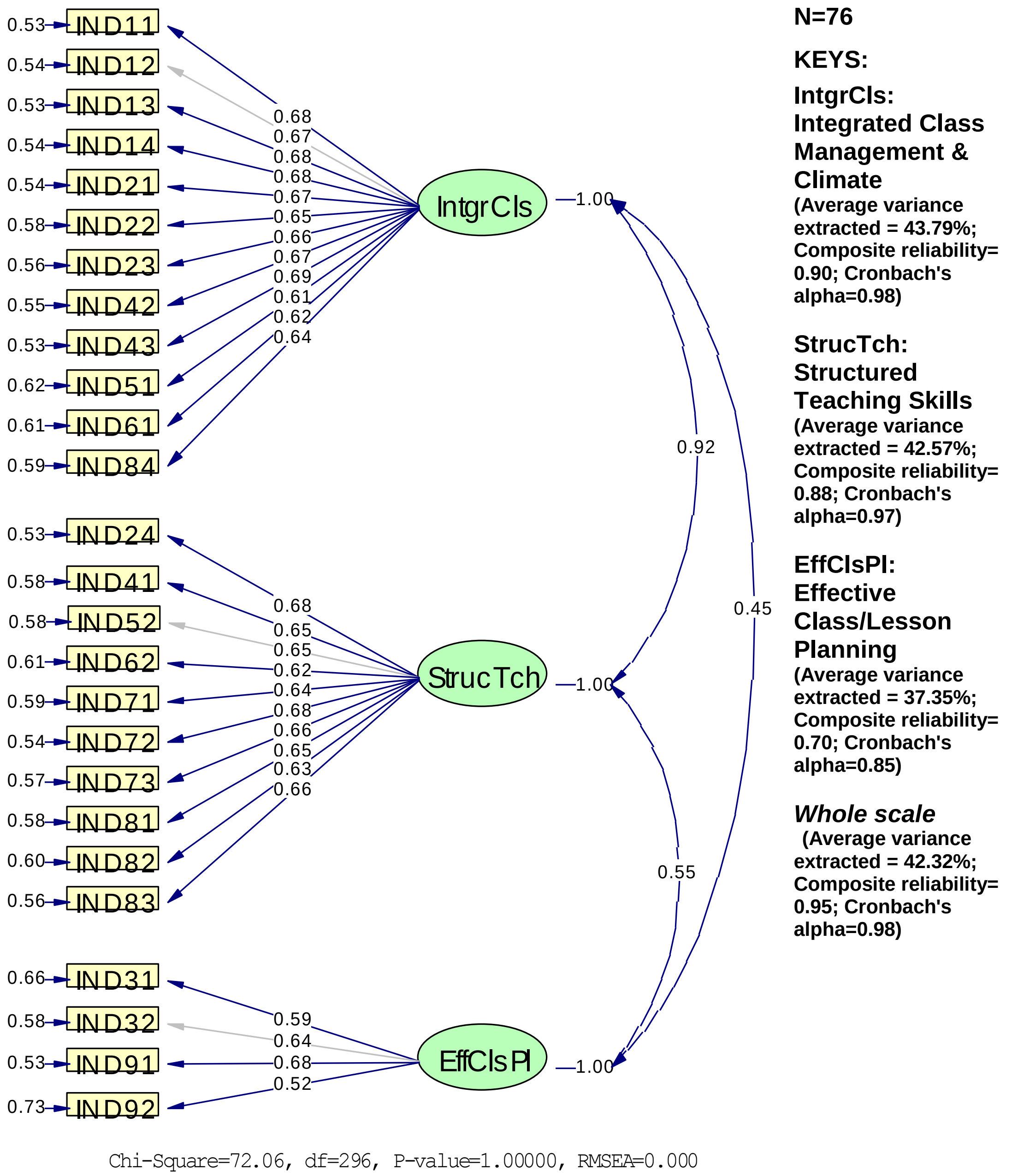
Factor Name	Indicator No.	Indicator description (T = The teacher; P = pupil)	Component 1	Component 2	Component 3
Integrated class management and climate	IND12	The teacher (T) promotes mutual respect	0.90		
	IND11	T ensures a relaxed atmosphere	0.89		
	IND13	T supports the self-confidence of Ps	0.86		
	IND14	T shows respect for the Ps in behaviour and language use	0.86		
	IND21	T ensures cohesion	0.81		
	IND43	T gives feedback to Ps	0.76	0.48	
	IND42	T gives clear explanations of the learning materials and the assignments	0.73	0.50	
	IND23	T promotes cooperation between Ps	0.71	0.40	
	IND51	T involves all Ps in the lesson	0.67	0.50	
	IND22	T stimulates the independence of Ps	0.66	0.43	0.43
	IND84	T ensures effective classroom management	0.64	0.54	
	IND61	T adapts the instruction to the relevant differences between Ps	0.53	0.51	0.50
Structured teaching skills	IND52	T makes use of teaching methods that activate the Ps	0.44	0.80	
	IND73	T provides interactive instruction and activities		0.77	
	IND72	T stimulates the use of control activities	0.53	0.73	
	IND24	There is good individual involvement by the Ps	0.57	0.72	
	IND81	T gives a well structured lesson		0.70	0.45
	IND82	T ensures the orderly progression of the lesson	0.50	0.68	
	IND83	T uses learning time efficiently	0.58	0.67	
	IND62	T adapts the assignments and processing to the relevant differences between Ps		0.62	0.54
	IND71	T ensures that the teaching materials are orientated towards transfer	0.55	0.60	
	IND41	T gives clear instructions and explanations	0.55	0.57	
Effective class/lesson planning	IND32	T evaluates whether the objectives have been achieved at the end of the lesson			0.84
	IND92	The teaching environment is educational and contemporary			0.81
	IND91	T ensures that classroom layout supports the P activities			0.80
	IND31	T clarifies the lesson objectives at the start of the lesson			0.78

5.4 Three dimensions of effective classroom practices identified in the results of the confirmatory factor analysis

The confirmatory factor analysis was conducted on the basis of the EFA model, using the Maximum Likelihood estimation in LISREL 8.72. Figure 5.3

below shows a CFA model of three latent variables (hereafter as CVCP (QoT) Model) for all indicators. Despite the cross loadings, any further reduction of indicators in the current model was not supported by modification indices[54].

Figure 5.3: **The underlying dimensions of effective classroom practices as identified in a three-factor CVCP (QoT) model of 26 QoT indicators with standardised coefficients**



54 The solution reported here seems to be better than a data-driven solution that was built on a full model with CFA results obtained for each individual EFA factor. The results are presented in Appendix XI. As this data-driven solution was not built and based on theoretical grounds, it was considered less preferable. Moreover, its overall goodness-of-fit indices do not support the data-driven solution as superior to the current solution. For example, χ^2 is only improved by 37.81 after a reduction of degree of freedom by 164 with 8 indicators deleted. The loss of information certainly did not justify the negligible gains in the goodness-of-fit indices. Both solutions were initially not positive definite. The matrix became positive definite when LISREL automatically reset the ridge option taken with the ridge constant at 1.0.

Multiple goodness-of-fit indices were employed to evaluate the solution included chi square (χ^2), root mean square error of approximation (RMSEA), normed fit index (NFI), comparative fit index(CFI), incremental fit index (IFI), relative fit index (RFI), standardised root mean square residual (SRMR) and its 90% confidence interval (90% CI) and test of close fit (CFit), and Hoelter's critical N[55]. Each of these goodness-of-fit indices suggested that the three-factor model fit the data well (χ^2 =72.06, df=296, p = 1.0; RMSEA=0.0 with 90%CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.98; CFI=1.00; IFI=1.08; RFI=0.98; SRMR=0.050; Critical N = 425.26; a full list of fit indices available in Appendix XII).

An inspection of standardised residuals and modification indices indicated no localised points of ill fit in the solution (e.g., the largest standardised residual=1.89; the largest modification index=6.23). All freely estimated unstandardised parameters were statistically significant ($p<0.01$). Standardised parameter estimates from this solution are presented in Figure 5.3. As indicated in Figure 5.3, no indicator has a loading below 0.52. Except Indicator 31 and Indicator 92, all indicators of the three hypothesised factors have moderately high loadings somewhere between 0.61 and 0.69 (square multiple correlation or R^2 ranged from 0.27 to 0.46).

The average variance extracted for each factor as well as that for the whole scale was below 50%, indicating *insufficient* convergent validity was found for most factors and the whole model. That is, the CVCP (ISTOF) model presented in Figure 4.5 was relatively better than the QoT model on this regard. However, the composite reliability for each component and that for the whole model were generally high above 0.7. Reliability tests for each factor as well as for the whole scale were also highly positive (Cronbach's alpha ranged from 0.85 to 0.98[56]), suggesting each factor and the scale were good in terms of internal consistency.

55 It should be note that the current results were obtained through setting the ridge option to 0.01 as the matrix was initially not positive definite. The matrix became positive definite when LISREL automatically reset the ridge option taken with the ridge constant at 1.0. The ridge option has been a standard option since LISREL 7 to adjust regression models with near-multicollinearity (see Jöreskog & Sörbom, 1996, pp.24; 167; 169; 322).

56 A split-half test was run with similar results: Cronbach's alpha for First Half is 0.97 and the Second Half is 0.93. Spearman-Brown Equal Length and Unequal Length are both 0.97 and Guttman Split-Half coefficient is 0.93.

Unlike the CVCP (ISTOF) model, this QoT model shows high correlation only between factors *Integrated class management and climate* and *Structured teaching skills* in Figure 5.3. This is not surprising because many items in these two factors showed cross-loadings in Table 5.6. Nevertheless, it was not certain whether these factors might also be weak in discriminant validity. Accordingly, a test was performed to compare the average variances extracted and shared variances and the status of discriminant validity of factors was confirmed as presented in Table 5.7.

Table 5.7: **Average variance extracted and shared variance estimates of the underlying dimensions of the three-factor CVCP (QoT) Model**

Factor Name	Integrated Class Mgt & Climate	Structured Teaching Skills	Effective Class/Lesson Planning
Integrated class management and climate	**0.44**	0.90	0.27
Structured teaching skills	0.95	**0.43**	0.21
Effective class/lesson planning	0.52	0.46	**0.37**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

As expected, all the average variance explained estimates on the diagonal are higher than the shared variance estimates above the diagonal except the one between factors *Integrated Class Management and Climate* and *Structured Teaching Skills*, indicating its discriminant validity is stronger than that of the CVCP (ISTOF) model. Thus, despite weak convergent validity, this three-factor QoT model shows a good fit for the data, high internal consistency and relatively better discriminant validity.

5.5 Variation across lessons of the three dimensions of effective classroom practices in the CVCP (QoT) Model

Based on the operational definition of variation in Chapter One, variability in ratings of indicators of individual factors found across lessons is explored in this section. Variability across lessons is also considered as one of the key indicators of differentiated teaching effectiveness in Chapter Two. Two dimensions of the nature of differentiated teaching effectiveness can be noted in Table 5.8: *teachers vary in their strengths and weaknesses in different dimensions of effective classroom practices and these dimensions vary across lessons.*

Table 5.8: **Characteristics of underlying dimensions of effective classroom practices identified in the CVCP (QoT) Model (N=76)**

Factor Name	Agg. Mean	Median	Mode	Std. Deviation	Skewness	z-skewness	Kurtosis	z-kurtosis
Integrated class management and climate	1.99	1.98	2.48	0.49	-0.63	***2.27****	-0.12	0.23
Structured teaching skills	1.92	1.96	1.96	0.49	-0.51	1.85	-0.17	0.32
Effective class/lesson planning	1.55	1.52	1.52	0.35	0.55	***1.98****	1.19	***2.19****
I judge the overall quality of teaching as	2.93	3.00	3.00	0.74	-0.72	***2.61****	0.93	1.70

An examination of individual factors in the following subsections reveals the magnitude of variation for each factor and indicates the characteristics of the lessons of the four Hong Kong teachers in the sample. Statistically significant negative skewness and positive kurtosis were found, though the means and medians of these factors were very close. Generally, negative skewness suggests that in the majority of the lessons the teacher's behaviours were rated above the mean. This was evident in all factors, but statistically significant in two. Statistically significant positive kurtosis or platykurtic distribution, which indicates little variability in factor scores across lessons occurred in only one factor and negative kurtosis did not seem to strongly present, suggesting variability of ratings in QoT factors was not as great as it was for the ISTOF factors discussed in the previous chapter.

5.5.1 Factor Integrated class management and climate

The first factor of the current model retained all the twelve indicators in the factor in the EFA analysis presented earlier in Section 5.3. The factor loading of each indicator is moderately high, ranging between 0.61 and 0.69. Thus, the square multiple correlations of these indicators are close and within a reasonable range between 0.37 and 0.49, showing their similar relative importance. It seems that the CFA has not resulted in a reduction of indicators but a reduction of factors because the nine theoretical criteria have been reduced into three latent factors.

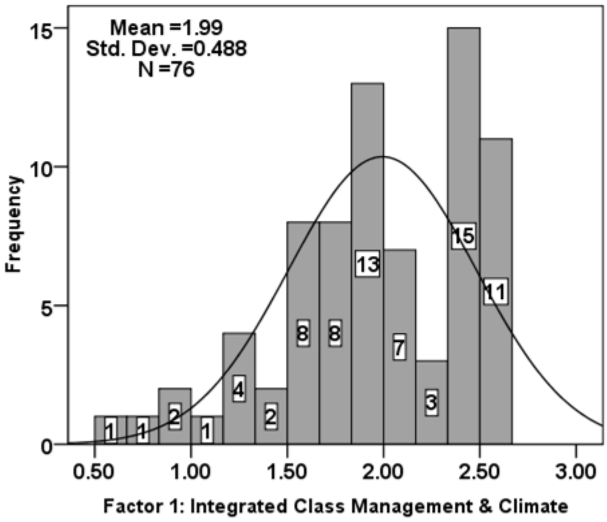
This factor provides empirical support for indicators of three theoretical criteria (shown in italics in Table *5.9*): *Safe and orderly school climate* (Criterion One), *Stimulating learning climate* (Criterion Two), and *Clear Instruction* (Criterion Three). Additional indicators of other criteria also convey strategies to enhance inclusion (Indicator 51) like adapting the instruction to cater for the individual needs (Indicator 61) and suggest the lesson has to be well-managed (Indicator 84) (see the words in bold in Table 5.9). The factor may not have clearly distinguished different theoretical criteria. Instead, it suggests that teaching is a dynamic activity in which the teacher has to do many different things to make learning effective. It may not be appropriate to study individual teaching behaviours and study them individually as they are present together to serve similar purpose. This probably explains why these indicators were shown highly internally consistent in the reliability test (Cronbach's alpha=0.98).

Table 5.9: **Coherence and relative importance of indicators as identified in the factor estimates of Factor *Integrated class management and climate* in the CVCP(QoT) Model (N=76)**

Factor Name	Indicator No.	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Integrated class management and climate (Cronbach's alpha=0.98)	***IND11***	*T ensures a relaxed atmosphere*	0.68	0.83	0.46
	IND12	*T promotes mutual respect*	0.67	Fixed at 1	0.45
	IND13	*T supports the self-confidence of Ps*	0.68	0.95	0.46
	IND14	*T shows respect for the Ps in behaviour and language use*	0.68	1.47	0.46
	IND21	*T ensures cohesion*	0.67	0.72	0.45
	IND22	*T stimulates the independence of Ps*	0.65	0.80	0.42
	IND23	*T promotes cooperation between Ps*	0.66	0.66	0.44
	IND42	*T gives clear explanations of the learning materials and the assignments*	0.67	0.64	0.45
	IND43	*T gives feedback to Ps*	0.69	0.93	0.48
	IND51	T ***involves all Ps*** in the lesson	0.61	0.50	0.37
	IND61	T ***adapts the instruction*** to the relevant differences between Ps	0.62	0.79	0.38
	IND84	T ensures ***effective classroom management***	0.64	1.15	0.41

Although a statistically significant negatively skewed distribution was found for this factor, Figure 5.4 reveals the histogram is bimodal. On the one hand, in about one-third of the lessons, the teachers on average tended to be predominantly strong in the teaching behaviours specified by all or most of the indicators in the factor. This is indicated by the higher peak on the right in Figure 5.4 and the region where the mode (=2.48) lies. On the other hand, in about half the lessons, the teachers on average only showed mixed strengths and weaknesses as indicated by the clustered scores around the second lower peak, where the mean (=1.99) and the median (=1.98) are.

Figure 5.4: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Integrated class management and climate***



Skewness = -0.63; Kurtosis = -0.12

The **lowest** possible value for this factor is 0.66, meaning that the teacher showed **predominantly weak** in the teaching behaviours specified by all the 12 indicators for a particular lesson.

A value of 1.32 may mean that on average the teacher showed **more weaknesses than strengths** in the teaching behaviours specified by most of the 12 indicators for a particular lesson.

A value of 1.98 may mean that on average the teacher showed **more strengths than weaknesses** in the teaching behaviours specified by most of the 12 indicators for a particular lesson.

The **highest** possible value for this factor is 2.64, meaning that the teacher showed **predominantly strong** in the teaching behaviours specified by all the 12 indicators for a particular lesson.

5.5.2 Factor *Structured teaching skills*

Retaining all the ten indicators in the EFA analysis made this factor similar to the first factor as an integration of indicators of diverse nature. A reliability test showed that its high internal consistency also (Cronbach's alpha=0.97) rivals that of the last factor. Factor loadings and square multiple correlations of indicators show narrower ranges, 0.62 to 0.68 and 0.38 to 0.46, respectively than the first factor.

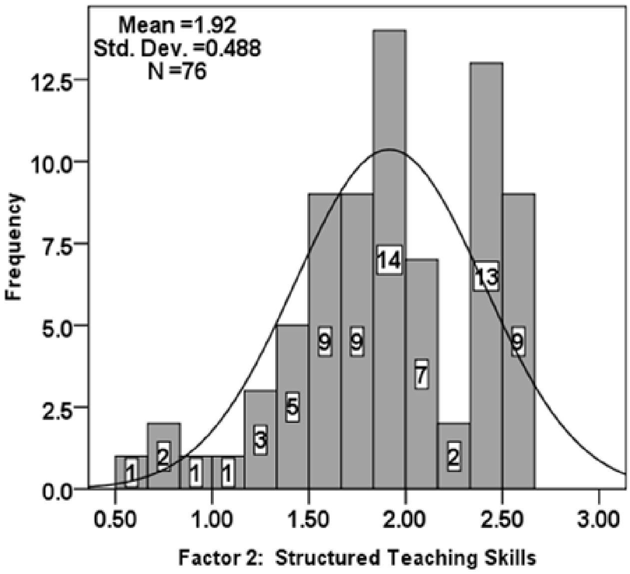
The balance found between the factor estimates of the indicators however should not be mistaken to suggest the factor is equally representing different theoretical criteria. A dominance of indicators of *Teaching learning strategies* (Criterion Seven) and *Effective classroom organisation* (Criterion Eight) was evident (shown in italics in Table 5.10). Again, like the first factor, additional indicators originated from other criteria are also related to inclusion (Indicator 62), instruction (Indicator 41), and student engagement strategies (Indicators 24 and 52). Taken all these indicators together, the factor seems to display how the teacher structured the lesson to ensure that learning can happen when the teacher can instruct the students clearly (Indicator 41) and adapt assignment and processing for diverse ability groups (Indicator 62) such that students are activated (Indicator 52) and involved individually (Indicator 24).

Table 5.10: **Coherence and relative importance of indicators as identified in the factor estimates of Factor *Structured teaching skills* in the CVCP (QoT) Model (N=76)**

Factor Name	Indicator No.	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Structured teaching skills (Cronbach's alpha=0.97)	**IND24**	There is **good individual involvement by the Ps**	0.68	1.13	0.46
	IND41	T gives **clear instructions and explanations**	0.65	1.02	0.42
	IND52	T makes use of teaching methods that **activate the Ps**	0.65	Fixed at 1	0.42
	IND62	T **adapts the assignments and processing** to the relevant **differences between Ps**	0.62	0.84	0.38
	IND71	*T ensures that the teaching materials are orientated towards transfer*	0.64	0.92	0.41
	IND72	*T stimulates the use of control activities*	0.68	1.30	0.46
	IND73	*T provides interactive instruction and activities*	0.66	1.45	0.44
	IND81	*T gives a well structured lesson*	0.65	0.82	0.42
	IND82	*T ensures the orderly progression of the lesson*	0.63	0.94	0.40
	IND83	*T uses learning time efficiently*	0.66	1.40	0.44

The histogram of this factor in Figure 5.5 shows a bimodal distribution very similar to the first factor, except there were fewer lessons in which the teacher received a higher rating and thus the higher peak, where the mode (=1.96) lies, is on the left.

Figure 5.5: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Structured teaching skills***



Skewness = -0.51; Kurtosis = -0.17

The **lowest** possible value for this factor is 0.65, meaning that the teacher showed **predominantly weak** in the teaching behaviours specified by all the 10 indicators for a particular lesson.

A value of 1.30 may mean that on average the teacher showed **more weaknesses than strengths** in the teaching behaviours specified by most of the 10 indicators for a particular lesson.

A value of 1.96 may mean that on average the teacher showed **more strengths than weaknesses** in the teaching behaviours specified by most of the 10 indicators for a particular lesson.

The **highest** possible value for this factor is 2.61, meaning that the teacher showed **predominantly strong** in the teaching behaviours specified by all the 10 indicators for a particular lesson.

5.5.3 Factor *Effective class/lesson planning*

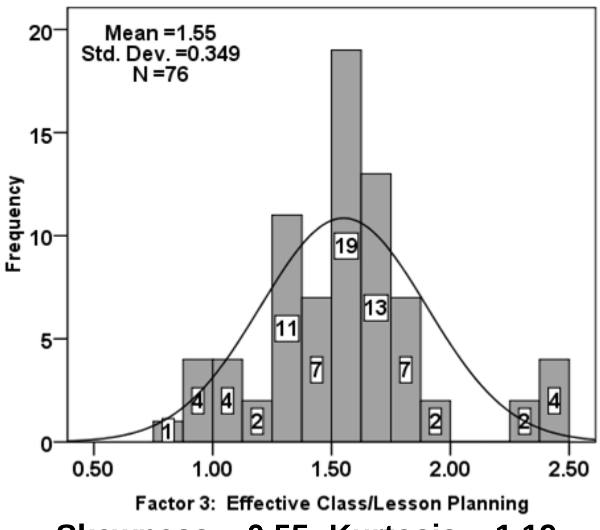
Consisting of only four indicators, the last factor of the current CFA solution in Table 5.11 seems to be much less eclectic than the first and second factors and clearly linking only two theoretical criteria: *Clear* *objectives* (Criterion Three) and *Effective classroom layout* (Criterion 9). Despite fewer indicators, its internal consistency (Cronbach's alpha= 0.85) is within high range but slightly lower than the last two factors. The factor loadings and square multiple correlations of the indicators of this factor are on average lower than those of the last two factors too, ranging between 0.52 and 0.68 and between 0.27 and 0.46, respectively.

Table 5.11: **Coherence and relative importance of indicators as identified in the factor estimates of Factor *Effective class/ lesson planning* in the CVCP (QoT) Model (N=76)**

Factor Name	Indicator No.	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Effective class/ lesson planning (Cronbach's alpha=0.85)	**IND31**	T clarifies the lesson objectives at the start of the lesson	0.59	1.24	0.35
	IND32	T evaluates whether the objectives have been achieved at the end of the lesson	0.64	Fixed at 1	0.41
	IND91	T ensures that classroom layout supports the P activities	0.68	0.72	0.46
	IND92	The teaching environment is educational and contemporary	0.52	0.75	0.27

All the three central tendency indicators, the mean (=1.55), (the median (=1.52) and the mode (=1.52), of this factor are lower than those of the other two factors. Accordingly, a statistically significant positively skewed distribution ($p<0.05$) was found, though the histogram in Figure 5.6 looks more like a normal distribution. This is a statistically significant platykurtic distribution ($p<0.05$), indicating low variability in ratings was found across lessons as ratings are clustered around the mode or the peak in Figure 5.6***.***

Figure 5.6: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of Factor *Effective class/lesson planning***



Skewness = 0.55; Kurtosis = 1.19

The **lowest** possible value for this factor is 0.61, meaning that the teacher showed **predominantly weak** in the teaching behaviours specified by all the 4 indicators for a particular lesson.

A value of 1.22 may mean that on average the teacher showed **more weaknesses than strengths** in the teaching behaviours specified by most of the 4 indicators for a particular lesson.

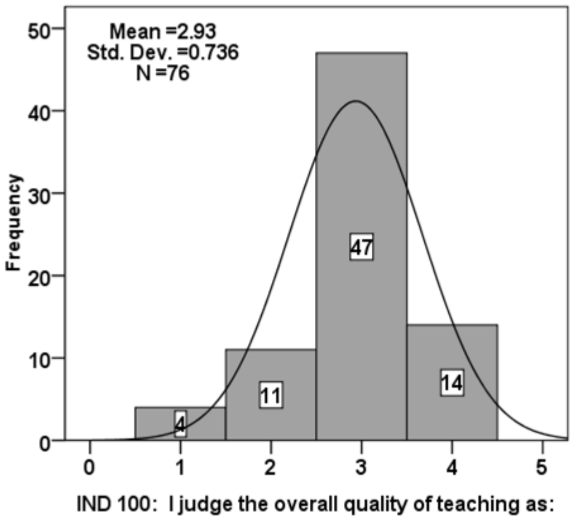
A value of 1.82 may mean that on average the teacher showed **more strengths than weaknesses** in the teaching behaviours specified by most of the 4 indicators for a particular lesson.

The **highest** possible value for this factor is 2.43, meaning that the teacher showed **predominantly strong** in the teaching behaviours specified by all the 4 indicators for a particular lesson.

5.5.4 Indicator 100: Overall judgment of lesson quality

As QoT is a high-inference instrument (see Section 2.2.4) designed on an inspection model of professional judgments of inspectors (see Section 3.5.3), it includes an indicator that serves to indicate the overall teaching quality (i.e., Indicator 100)[57]. This indicator is expected to represent a summative or final global judgment of the rater. Its negative skewed distribution in Figure 5.7 was statistically significant ($p<0.05$), suggesting that in more lessons, the teacher was seen as having more strengths than weaknesses.

Figure 5.7: **Relative variability of observed teacher's strengths and weaknesses identified in the frequency distribution of *Overall judgment of lesson quality***



Skewness = -0.72; Kurtosis = 0.93

The **lowest** possible value for this factor is 1, meaning that the teacher showed **predominantly weak** in the overall teaching quality for a particular lesson.

A value of 2 may mean that on average the teacher showed **more weaknesses than strengths** in the overall teaching quality for a particular lesson.

A value of 3 may mean that on average the teacher showed **more strengths than weaknesses** in the overall teaching quality for a particular lesson.

The **highest** possible value for this factor is 4, meaning that the teacher showed **predominantly strong** in the overall teaching quality for a particular lesson.

5.6 The extent of similarities and discrepancies between CVCP factors and theoretical factors

5.6.1 Patterns of differences in the shared indicators of CVCP factors and theoretical factors

Despite the excellent goodness-of-fit indices of the solution presented in the previous section, this model was very different from the nine criteria theoretical model because only three latent factors were identified. Table 5.12 shows the relationship between the three latent factors discussed above

57 In a recent paper, van de Grift (2007) did not use Indicator 100 as an indicator of overall judgment of teaching quality. Instead he standardized all scale scores by "dividing the sum score by the product of the number of items and the number of response categories" (van de Grift, 2007, p.140). Clearly, van de Grift was not using CFA to explore the validity of QoT. This indicator is important in the present study as it is the overall impression of a teacher's effectiveness in a particular lesson. Further analyses based on this indicator and factor scores are discussed in detail in the next chapter.

and the respective theoretical criteria that were hypothesised to be associated with the indicators in those factors. The fifth column shows all the indicators originally in the theoretical criteria, with the one(s) that were also found in the latent factors underlined. Two characteristics can be identified in those relationships.

Table 5.12: **Shared indicators in the CVCP (QoT) factors and the theoretical criteria of QoT (N=76)**

Factor Name	Indicator No.	Indicator Description (T = The teacher; P =pupil)	Original Theoretical Criterion	Indicator No. in concern
Integrated class management and climate (Cronbach's alpha=0.98)	***IND11*** ***IND12*** ***IND13*** ***IND14***	*T ensures a relaxed atmosphere* *T promotes mutual respect* *T supports the self-confidence of Ps* *T shows respect for the Ps in behaviour and language use*	**Safe and orderly school climate**	**11, 12, 13. 14**
	IND21 ***IND22*** ***IND23***	*T ensures cohesion* *T stimulates the independence of Ps* *T promotes cooperation between Ps*	**Stimulating learning climate**	**21, 22, 23,** 24
	IND42 ***IND43***	*T gives clear explanations of the learning materials and the assignments* *T gives feedback to Ps*	**Clear instruction**	**41, 42, 43**
	IND51	T involves all Ps in the lesson	**Activating Pupils**	**51,** 52
	IND61	T adapts the instruction to the relevant differences between Ps	**Adaptation of teaching**	**61,** 62
	IND84	T ensures effective classroom management	**Effective classroom organization**	81, 82, 83, **84**
Structured teaching skills (Cronbach's alpha=0.97)	***IND24***	There is good individual involvement by the Ps	**Stimulating learning climate**	21, 22, 23, **24**
	IND41	T gives clear instructions and explanations	**Clear instruction**	**41, 42, 43**
	IND52	T makes use of teaching methods that activate the Ps	**Activating pupils**	51, **52**
	IND62	T adapts the assignments and processing to the relevant differences between Ps	**Adaptation of teaching**	61, **62**
	IND71 ***IND72*** ***IND73***	*T ensures that the teaching materials are orientated towards transfer* *T stimulates the use of control activities* *T provides interactive instruction and activities*	**Teaching learning strategies**	**71,72, 73**
	IND81 ***IND82*** ***IND83***	*T gives a well structured lesson* *T ensures the orderly progression of the lesson* *T uses learning time efficiently*	**Effective classroom organization**	**81, 82, 83,** 84
Effective class/ lesson planning (Cronbach's alpha=0.85)	***IND31*** ***IND32***	*T clarifies the lesson objectives at the start of the lesson* *T evaluates whether the objectives have been achieved at the end of the lesson*	**Clear Objectives**	**31, 32**
	IND91 ***IND92***	*T ensures that classroom layout supports the P activities* *The teaching environment is educational and contemporary*	**Effective classroom layout**	**91, 92**

First, the coherence of some theoretical criteria was confirmed. *Safe and orderly school climate*, *Teaching learning strategies, Clear objectives* and *Effective classroom layout* are criteria with all indicators (highlighted in *italics*) present in a single latent factor. In contrast, *Stimulating learning climate*, *Clear instruction* and *Effective classroom organisation* are less coherent because its indicators (highlighted in ***red***) appear in two latent factors.

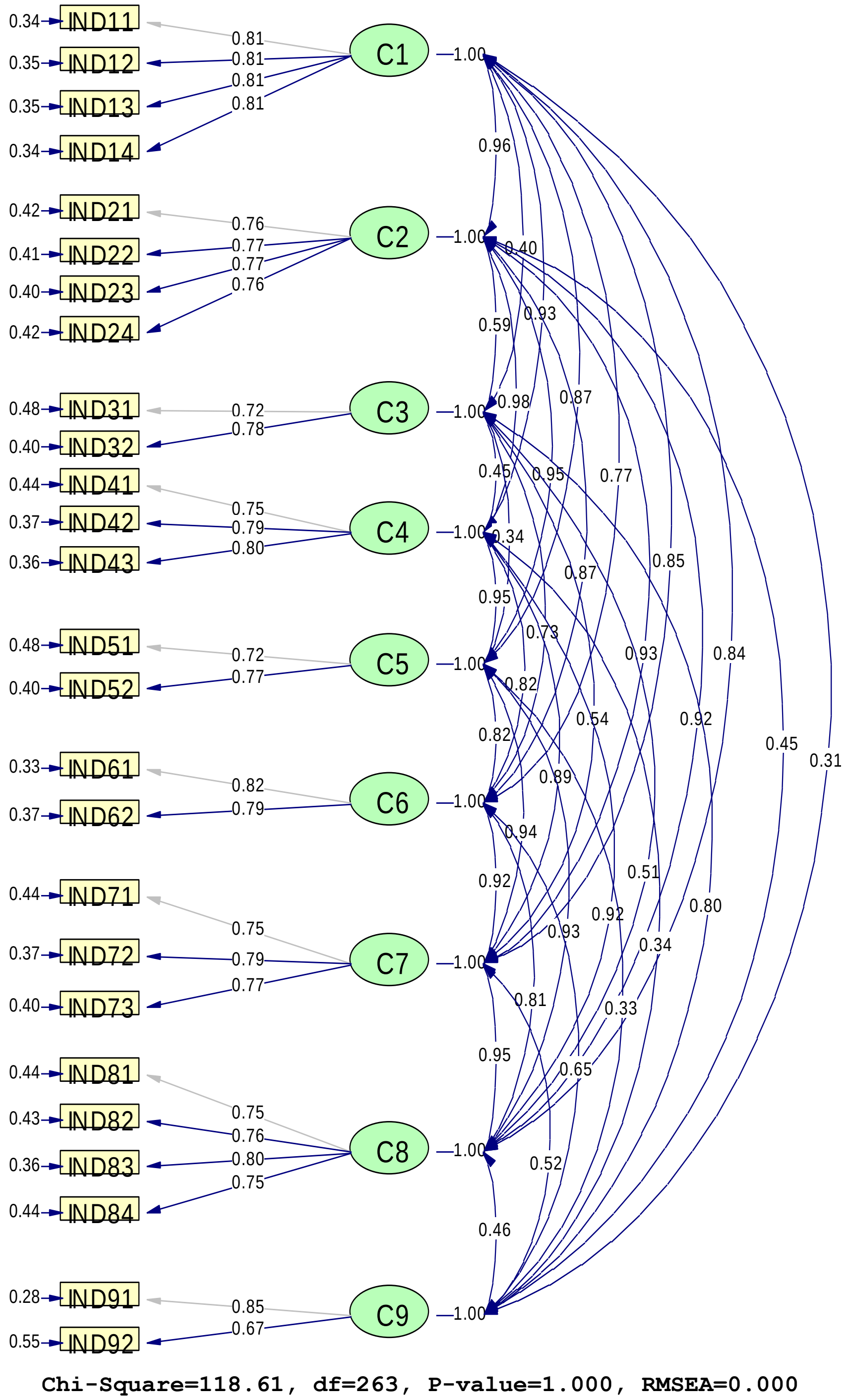
Second, criteria like *Activating pupils* and *Adaptation of teaching* seem to be less coherent as the indicators appear in two latent factors. However, since the criteria *Activating pupils* and *Adaptation of teaching* only have two indicators, it is inappropriate to conclude their incoherence without reference to their internal consistency under reliability tests (see these results in the next section). However, results of item-total statistics in SPSS16 showed that when these indicators were deleted, reliability would only diminish very little with a negligible difference (Cronbach's alpha decreased from 0.975 to 0.974), but this also applied to other indicators because the reliability would decline very little if any single indicator was to be deleted (no more than a loss of 0.04 in Cronbach's alpha).

5.6.2 Exploring the convergent and discriminant validity of theoretical models of QoT

To explore the convergent and discriminant validity of the theoretical nine-criterion model, several confirmatory analyses were conducted to compare a model with nine underlying factors (Model A) and two other rival models (Model B and Model C) with a simpler underlying structure. Figure 5.8 displays the theoretical model (or Model A) with all twenty six indicators in nine theoretical criteria. Although the number of indicators of this model is the same as the current solution discussed above, the number of parameters of this model has increased dramatically from 55 to 88. Accordingly, LISREL8.72 automatically flatted a warning message to suggest that the solution might not be reliable with the number of parameters estimated exceeding the number of sample size (N=76)[58]. This meant that the solution was more complex and became less reliable with the amount of data used to generate estimation.

58 Two other warning messages were flagged in LISREL. First, a ridge option of 0.5 was taken as the matrix was not positive definite. Second, a warning message of that PHI was not positive, but an inspection of it showed that all the t-values for PHI were statistically significant, indicating no apparent problem was there.

Figure 5.8: **A CFA model based on NINE theoretical criteria and TWENTY SIX Indicators with standardised coefficients (N=76) [Model A]**



KEYS:

C 1 =
Safe and orderly school climate
(Cronbach's alpha=0.98)

C 2 =
Stimulating learning climate
(Cronbach's alpha=0.93)

C 3 =
Clear Objectives
(Cronbach's alpha=0.76)

C 4 =
Clear instruction
(Cronbach's alpha=0.92)

C 5 =
Activating Pupils
(Cronbach's alpha=0.85)

C 6 =
Adaptation of teaching
(Cronbach's alpha=0.93)

C 7 =
Teaching learning strategies
(Cronbach's alpha=0.93)

C 8 =
Effective classroom organization
(Cronbach's alpha=0.93)

C 9 =
Effective classroom layout
(Cronbach's alpha=0.82)

Despite such a warning message, a CFA analysis using the maximum likelihood estimation method in LISREL 8.72 went smoothly and arrived at a solution after only 17 iterations. Moreover, each of the overall goodness-of-fit indices of this solution seemed to suggest that a CFA model based on all nine theoretical criteria might also fit the data well (χ^2 =118.61, df=263, p = 1.0; RMSEA=0.0 with 90%CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.98;

CFI=1.00; IFI=1.03; RFI=0.98; SRMR=0.057; Critical N = 218.36).,These indices were slightly better than those obtained for the three-factor empirical model discussed in the last three sections. All standardised residuals and modification indices (e.g., largest standardised residual = 2.47; the largest modification index = 10.14) looked acceptable, indicating the absence of localised weaker points in the solution. Standardised parameter estimates from this solution are presented in Figure 5.8 and Table 5.13 below. All freely estimated unstandardised parameters were statistically significant ($p < 0.05$).

Table 5.13: **Coherence and relative importance of indicators as identified in the factor estimates and reliability of a CFA model based on NINE theoretical criteria and TWENTY SIX indicators (N=76) [Model A]**

Factor Name	Indicator No.	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Safe and orderly school climate	**IND11**	T ensures a relaxed atmosphere	0.81	Fixed at 1	0.66
	IND12	T promotes mutual respect	0.81	1.21	0.65
	IND13	T supports the self-confidence of Ps	0.81	1.13	0.65
	IND14	T shows respect for the Ps in behaviour and language use	0.81	1.77	0.66
Stimulating learning climate	**IND21**	T ensures cohesion	0.76	Fixed at 1	0.58
	IND22	T stimulates the independence of Ps	0.77	1.16	0.59
	IND23	T promotes cooperation between Ps	0.77	0.95	0.60
	IND24	There is good individual involvement by the Ps	0.76	1.38	0.58
Clear Objectives	**IND31**	T clarifies the lesson objectives at the start of the lesson	0.72	Fixed at 1	0.52
	IND32	T evaluates whether the objectives have been achieved at the end of the lesson	0.78	0.79	0.60
Clear instruction	**IND41**	T gives clear instructions and explanations	0.75	Fixed at 1	0.56
	IND42	T gives clear explanations of the learning materials and the assignments	0.79	0.74	0.63
	IND43	T gives feedback to Ps	0.80	1.07	0.64
Activating Pupils	**IND51**	T involves all in the lesson	0.72	Fixed at 1	0.52
	IND52	T makes use of teaching methods that activate the Ps	0.77	1.76	0.60
Adaptation of teaching	**IND61**	T adapts the instruction to the relevant differences between Ps	0.82	Fixed at 1	0.67
	IND62	T adapts the assignments and processing to the relevant differences between Ps	0.79	0.88	0.63
Teaching learning strategies	**IND71**	T ensures that the teaching materials are orientated towards transfer	0.75	Fixed at 1	0.56
	IND72	T stimulates the use of control activities	0.79	1.42	0.63
	IND73	T provides interactive instruction and activities	0.77	1.60	0.60
Effective classroom organization	**IND81**	T gives a well structured lesson	0.75	Fixed at 1	0.56
	IND82	T ensures the orderly progression of the lesson	0.76	1.19	0.57
	IND83	T uses learning time efficiently	0.80	1.79	0.64
	IND84	T ensures effective classroom management	0.75	1.68	0.56
Effective classroom layout	**IND91**	T ensures that classroom layout supports the P activities	0.85	Fixed at 1	0.72
	IND92	The teaching environment is educational and contemporary	0.67	1.09	0.45

The factor loadings and square multiple correlations of the indicators of this model are even higher than those of the solution of a three- factor model. Factor loading estimates between 0.67 and 0.81 suggested that all indicators were strong in their relations with their purported latent factors (square multiple correlation or R^2 ranged from 0.45 to 0.66).

Moreover, the average variance extracted for each factor as well as that for the whole scale were at least 55%, indicating *stronger* convergent validity was found for all factors and for this model than were for the empirical QoT factors and for the CVCP (QoT) Model. As shown in Table 5.14, both types of reliability tests for each component and for the whole model were generally high, ranging from 0.71 to 0.97 for the composite reliability and from 0.76 to 0.98 for Cronbach's alpha[59]. These results suggested the scale were good in terms of internal consistency. The reliability for the whole scale would be the same as the three-factor model as the two models have the same indicators (Cronbach's alpha = 0.98). However, reliability tests run for each theoretical criterion showed that each had high internal consistency, even for those criteria with only two indicators. This probably explains why the internal consistency for the whole scale was so high.

Table 5.14: **Average variance extracted, composite reliability and Cronbach's alpha reliability of the CFA model based on NINE theoretical criteria [Model A]**

Factor Name	**Average Variance Extracted**	**Composite reliability**	**Cronbach's Alpha**
Safe and orderly school climate	0.66	**0.88**	0.98
Stimulating learning climate	0.59	**0.85**	0.93
Clear Objectives	0.55	**0.71**	0.76
Clear instruction	0.61	**0.82**	0.92
Activating Pupils	0.56	**0.72**	0.85
Adaptation of teaching	0.65	**0.79**	0.93
Teaching learning strategies	0.59	**0.82**	0.93
Effective classroom organization	0.59	**0.85**	0.93
Effective classroom layout	0.59	**0.74**	0.82
Whole Scale	0.60	0.97	0.98

A mixture of high and low intercorrelations between the theoretical criteria in Figure 5.8 suggested that some of these criteria might be weak in discriminant validity. A test performed to compare the average variances extracted and shared variances confirmed some insufficient discriminant validity was found for more than half of the criteria. As indicated in Table 5.15, criteria *Clear objectives* and *Effective classroom layout* are most distinctive

59 A split-half test was run for the whole scale with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.96. Spearman-Brown Equal Length and Unequal Length are both 0.94 and Guttman Split-Half coefficient is 0.93.

for having all shared variances lower than their average variances extracted, followed by *Safe and orderly school climate* and *Adaptation of teaching,* each of which has three shared variances below their average variances explained. Again, these results suggest that this theoretical model is better than the empirical model in terms of convergent and discriminant validity

Table 5.15: **Average variance extracted and shared variance estimates of a CFA model based on NINE theoretical criteria [Model A]**

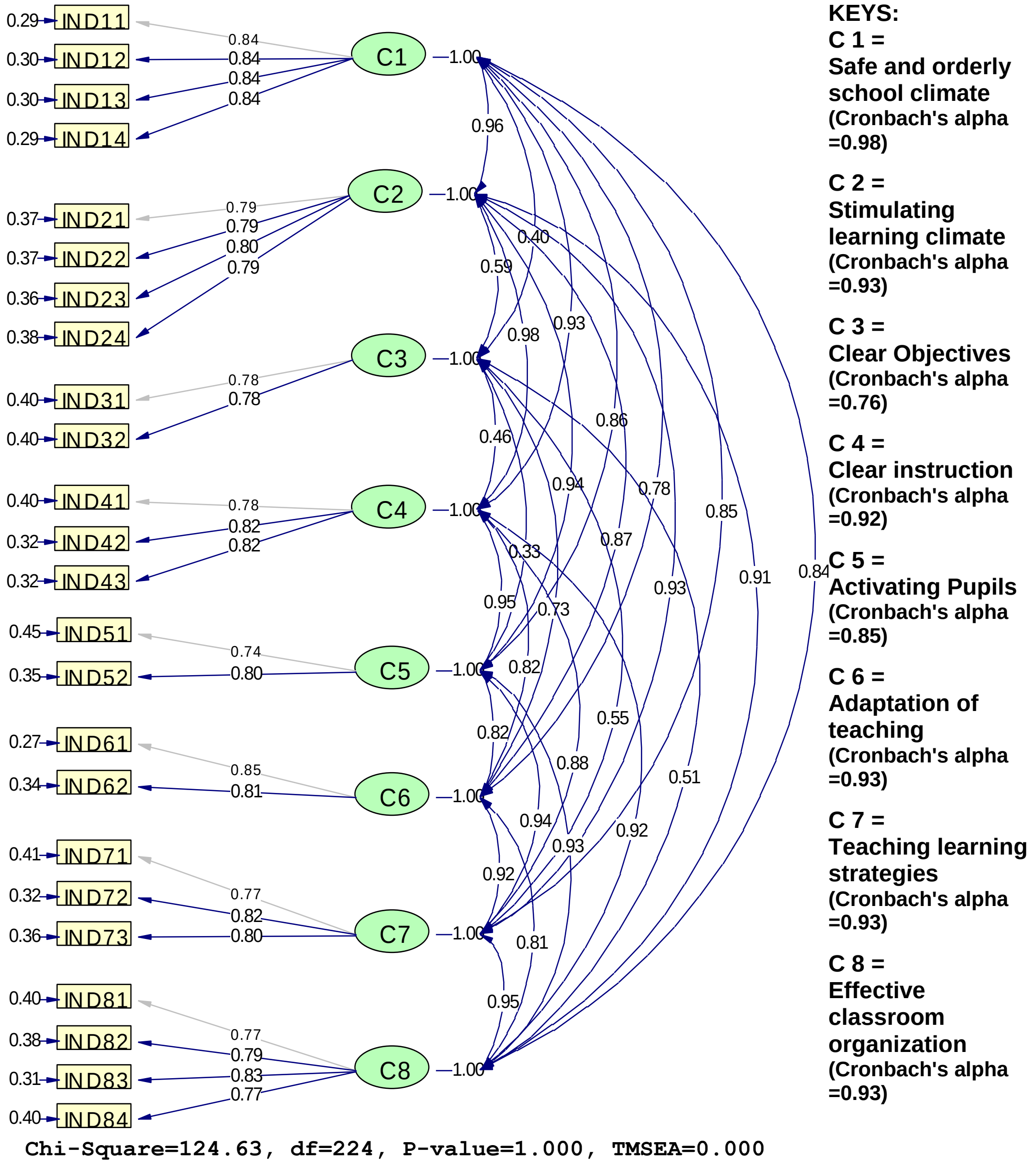
Factor Name	Safe and orderly school climate	Stimulating learning climate	Clear Objectives	Clear instruction	Activating Pupils	Adaptation of teaching	Teaching learning strategies	Effective classroom organ-ization	Effective classroom layout
Safe & orderly school climate	**0.66**	0.92	*0.16*	0.86	0.76	*0.59*	0.72	0.71	*0.10*
Stimulating learning climate	0.96	**0.59**	*0.35*	0.96	0.90	0.76	0.86	0.85	*0.20*
Clear Objectives	0.40	0.59	**0.55**	*0.20*	*0.12*	*0.53*	*0.29*	*0.26*	0.64
Clear instruction	0.93	0.98	0.45	**0.61**	0.90	0.67	0.79	0.85	*0.12*
Activating pupils	0.87	0.95	0.34	0.95	**0.56**	0.67	0.88	0.86	*0.11*
Adaptation of teaching	0.77	0.87	0.73	0.82	0.82	**0.65**	0.85	0.66	*0.42*
Teaching learning strategies	0.85	0.93	0.54	0.89	0.94	0.92	**0.60**	0.90	*0.27*
Effective classroom organization	0.84	0.92	0.51	0.92	0.93	0.81	0.95	**0.58**	*0.21*
Effective classroom layout	0.31	0.45	0.80	0.34	0.33	0.65	0.52	0.46	**0.59**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

Given such good results, one may have to reconsider this solution as a tentative alterative to the three-factor one because of its theoretical base. The problem and uncertainty surrounded by a solution with a sample size smaller than the number of parameters seemed to reflect a rule of thumb and there is no research in the literature that has compared solutions with a small number of actors (like three) and a large number of factors (like nine) while the number of indicators unchanged. However, Marsh and Hau (1999) found that the acceptability of a solution increased with more factors when the number of indicators and sample size unchanged. That is, there is a trade off between complexity of a solution and the number of sample size required for a reliable solution. A nine-factor model is certainly more complex than a three-factor one. From a theoretical perspective, a nine-factor model may be attractive as it is what the instrument is intended to measure, but a three-factor model is more preferable as it is more parsimonious, given that both models are well supported by the data.

One attempt to reduce the number of factors without deviating too much from the original theoretical structure underlying the instrument is to delete Criterion Nine, *Effective classroom layout*. Figure 5.9 below shows Model B, which is the same as the last model except the last factor is deleted. Deleting Criterion Nine was also motivated because, as discussed above, this criterion did not seem to be applicable in the Hong Kong contexts as it had been in the English contexts.

Figure 5.9: **A CFA model based on EIGHT theoretical criteria and TWENTY FOUR indicators with standardised coefficients (N=76) [Model B]**



Like the previous models, Model B was generated by employing the Maximum Likelihood estimation in LISREL 8.72, based on the first eight theoretical criteria. The solution generated after 15 iterations seemed to fit the data equally well as previous models[60] because while some goodness-of-fit indices of this solution were slightly poorer (χ^2 =124.63, df=224, *p*=1.0; IFI=1.02; RFI=0.98; Critical N = 183.58), while some were remained unchanged or better (e.g., RMSEA=0.0 with 90% CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.98; CFI=1.00; SRMR=0.041). Standardised parameter estimates from this solution are presented in Figure 5.9 above and Table 5.16 below.

Table 5.16: **Coherence and relative importance of indicators as identified in the factor estimates and reliability of a CFA model based on EIGHT theoretical criteria and TWENTY FOUR indicators (N=76) [Model B]**

Factor Name	Indic-ator No.	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Safe and orderly school climate	IND11	T ensures a relaxed atmosphere	0.84	Fixed at 1	0.71
	IND12	T promotes mutual respect	0.84	1.21	0.70
	IND13	T supports the self-confidence of Ps	0.84	1.13	0.70
	IND14	T shows respect for the Ps in behaviour and language use	0.84	1.77	0.71
Stimulating learning climate	IND21	T ensures cohesion	0.79	Fixed at 1	0.63
	IND22	T stimulates the independence of Ps	0.79	1.16	0.63
	IND23	T promotes cooperation between Ps	0.80	0.95	0.64
	IND24	There is good individual involvement by the Ps	0.79	1.33	0.62
Clear Objectives	IND31	T clarifies the lesson objectives at the start of the lesson	0.78	Fixed at 1	0.60
	IND32	T evaluates whether the objectives have been achieved at the end of the lesson	0.78	0.73	0.60
Clear instruction	IND41	T gives clear instructions and explanations	0.78	Fixed at 1	0.60
	IND42	T gives clear explanations of the learning materials and the assignments	0.82	0.74	0.68
	IND43	T gives feedback to Ps	0.82	1.06	0.68
Activating Pupils	IND51	T involves all Ps in the lesson	0.74	Fixed at 1	0.55
	IND52	T makes use of teaching methods that activate Ps	0.80	1.77	0.65
Adaptation of teaching	IND61	T adapts the instruction to the relevant differences between Ps	0.85	Fixed at 1	0.73
	IND62	T adapts the assignments and processing to the relevant differences between Ps	0.81	0.87	0.66
Teaching learning strategies	IND71	T ensures that the teaching materials are orientated towards transfer	0.77	Fixed at 1	0.59
	IND72	T stimulates the use of control activities	0.82	1.42	0.68
	IND73	T provides interactive instruction and activities	0.80	1.60	0.64
Effective classroom organization	IND81	T gives a well structured lesson	0.77	Fixed at 1	0.60
	IND82	T ensures the orderly progression of the lesson	0.79	1.20	0.62
	IND83	T uses learning time efficiently	0.83	1.80	0.69
	IND84	T ensures effective classroom management	0.77	1.68	0.60
Reliability of 24 indicators without Indicators 91 and 92: Cronbach's alpha =0.98[61]					

60 Like previous ones, the ridge option had to be reset because the matrix was not positive definite. This solution was obtained by taking a ridge option at 0.5. Some other ridge options might yield better goodness-of-fit indices but they were rejected because PHI was not positive definite with some insignificant estimates. This solution produced the best goodness-of-fit indices without any insignificant estimates.

61 A split-half test was run with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.97. Spearman-Brown Equal Length and Unequal Length are both 0.96 and Guttman Split-Half coefficient is 0.96. The split-half reliability test for a model without Indicators 91 and 92 was slightly better than a model with al indicators but the gain is very marginal and probably negligible.

In general, standardised residuals looked acceptable (e.g., largest standardised residual = 2.30), but there were two modification indices over 10, one between Factor 3 and Indicator 81 (=10.95) and another between Factor 5 and Indicator 24 (= 10.14), indicating that some better fit solutions could be obtained. However, these tentative solutions[62] are not of interest here because the current approach is intended to drive at a model or solution that is grounded on theoretical grounds rather than purely driven by data mining. All freely estimated unstandardised parameters were statistically significant (p<0.05). The factor loadings and square multiple correlations of the indicators of this model are even higher than those of the solution of either the three-factor model or the nine-factor model. Factor loading estimates between 0.77 and 0.85 suggested that all indicators were strong in their relations with their purported latent factors (square multiple correlation or R^2 ranged from 0.59 to 0.73).

Compared with the Model A, the average variance extracted of each factor as well as that of the whole scale was higher, at 61% or above, indicating *stronger* convergent validity was found for all factors and this model than the CVCP (QoT) Model and Model A. As shown in Table 5.17, both types of reliability tests for each component and the whole model were generally high, ranging from 0.75 to 0.98 for the composite reliability and from 0.76 to 0.98 for Cronbach's alpha[63].

Table 5.17: **Average variance extracted, composite reliability and Cronbach's alpha reliability of the CFA model based on EIGHT theoretical criteria [Model B]**

Factor Name	Average Variance Extracted	Composite reliability	Cronbach's Alpha
Safe and orderly school climate	0.71	**0.91**	0.98
Stimulating learning climate	0.63	**0.87**	0.93
Clear Objectives	0.61	**0.75**	0.76
Clear instruction	0.65	**0.85**	0.92
Activating Pupils	0.59	**0.75**	0.85
Adaptation of teaching	0.69	**0.82**	0.93
Teaching learning strategies	0.64	**0.84**	0.93
Effective classroom organization	0.62	**0.87**	0.93
Whole Scale	0.64	0.98	0.98

62 For example, another solution obtained by deleting Indicator 81: *The teacher gives a well structured lesson* and linking Indicator 24: *There is good individual involvement by the pupils* with *Activating Pupils* (Factor 5) did produce much better results (see Appendix XIII for the model diagram and fit indices and Appendix XIV for the factor estimates and reliability).

63 A split-half test was run for the whole scale with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.96. Spearman-Brown Equal Length and Unequal Length are both 0.94 and Guttman Split-Half coefficient is 0.93.

The reliability for each theoretical criterion was the same as that for each factor in the nine-criterion model because each factor had the same indicators in the two models. That is, the high internal consistency found for each factor remained unchanged. The reliability for the whole scale was identical with the previous one after deleting Criterion Nine. Thus, these results suggested the scale were good in terms of internal consistency, but Model B can be a slightly parsimonious alternative with an underlying structure similar to that of the original scale and similar goodness-of-fit.

Figure 5.9 also shows a mixture high and low intercorrelations between the theoretical criteria, suggesting that some of these criteria might be weak in discriminant validity. Table 5.18 shows the results confirming the insufficient discriminant validity found for more than half of the criteria. In particular, with its average variance explained higher than all of its shared variances, *Clear objectives* is the only distinctive criterion. Again, these results suggest that this modified theoretical QoT model is better than the empirical model in terms of convergent and discriminant validity.

Table 5.18: **Average variance extracted and shared variance estimates of a CFA model based on EIGHT theoretical criteria [Model B]**

Factor Name	Safe and orderly school climate	Stimulating learning climate	Clear Objectives	Clear instruction	Activating Pupils	Adaptation of teaching	Teaching learning strategies	Effective classroom organ-ization
Safe and orderly school climate	**0.71**	0.94	*0.16*	0.86	0.77	*0.59*	0.72	0.76
Stimulating learning climate	0.97	**0.63**	*0.41*	0.96	0.88	0.76	0.81	0.77
Clear Objectives	0.40	0.64	**0.61**	*0.21*	*0.14*	*0.53*	*0.30*	*0.18*
Clear instruction	0.93	0.98	0.46	**0.65**	0.92	0.69	0.79	0.83
Activating pupils	0.88	0.94	0.38	0.96	**0.59**	0.69	0.94	0.86
Adaptation of teaching	0.77	0.87	0.73	0.83	0.83	**0.69**	0.85	*0.59*
Teaching learning strategies	0.85	0.90	0.55	0.89	0.97	0.92	**0.64**	0.88
Effective classroom organization	0.87	0.88	0.43	0.91	0.93	0.77	0.94	**0.62**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

Another alternative way to simplify the nine-factor model is to reduce the number of factors on theoretical grounds. The current version of QoT was adopted from an earlier version in which Criterion Nine was included. In a more recent version, this criterion was deleted and some criteria were combined (van de Grift, 2007). As shown in Table 5.19, indicators of Criteria

Safe and orderly school climate and *Stimulating learning climate* are combined to form a new factor (in *italics*)[64] of seven indicators labelled *Safe and stimulating learning climate*. Similarly, indicators of *Clear objectives, Clear instruction* and *Activating pupils* are combined as one new factor of seven indicators that has retained *Clear Instruction* as its name. Three theoretical criteria, *Adaptation of teaching, Teaching learning strategies,* and *Efficient classroom management,* remain regarded as distinctive factors in the 2007 version of QoT.

Table 5.19: **Comparisons between two versions of QoT by indicators**

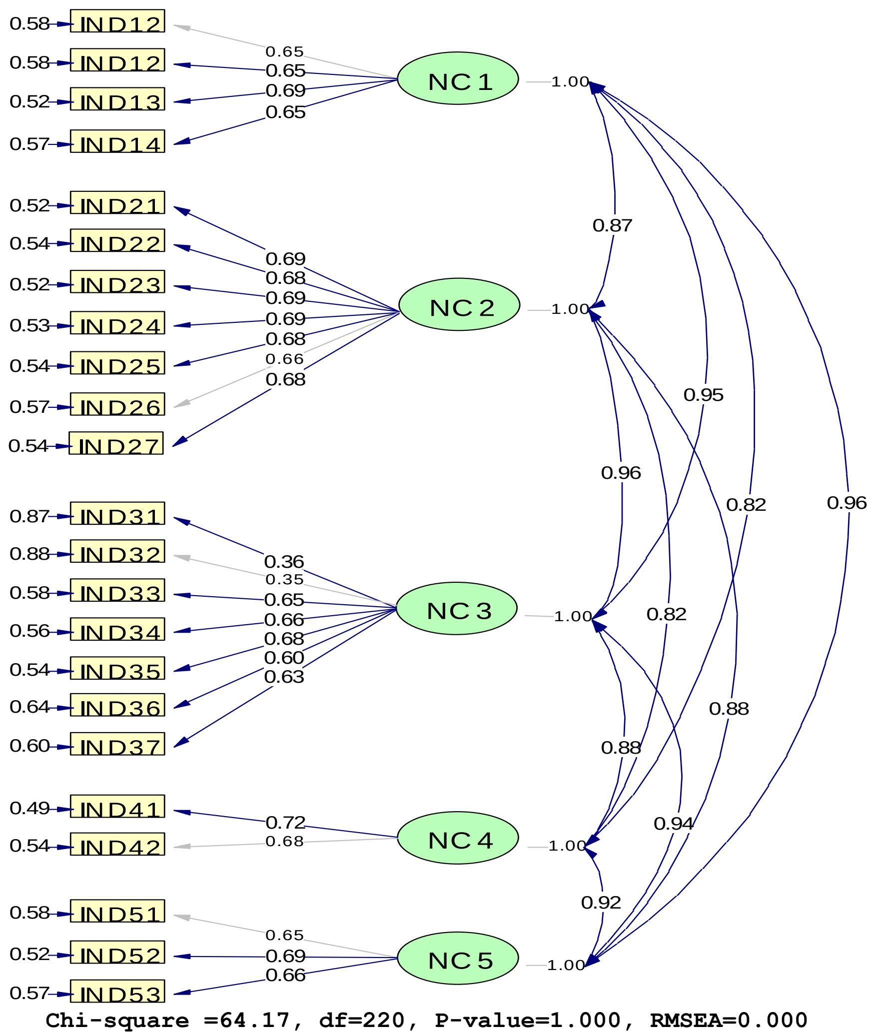
2007 Scale Name	2007 Indicator No.	2004 Indicator No.	Indicator Description (T = The teacher; P = pupil)
Safe and stimulating learning climate	**IND21**	**IND11**	The teacher ensures a relaxed atmosphere
	IND22	**IND12**	The teacher promotes mutual respect
	IND23	**IND13**	The teacher supports the self-confidence of Ps
	IND24	**IND14**	The teacher shows respect for the Ps in behaviour and language
	IND25	**IND21**	The teacher ensures cohesion
	IND26	**IND22**	The teacher stimulates the independence of Ps
	IND27	**IND23**	The teacher promotes cooperation between Ps
	IND61	**IND24**	**There is good individual involvement by the Ps**
Clear instruction	**IND31**	**IND31**	The teacher clarifies the lesson objectives at the start of the lesson
	IND32	**IND32**	The teacher evaluates whether the objectives have been achieved at the end of the lesson
	IND33	**IND41**	The teacher gives clear instructions and explanations
	IND34	**IND42**	The teacher gives clear explanations of the learning materials and the assignments
	IND35	**IND43**	The teacher gives feedback to Ps
	IND36	**IND51**	The teacher involves all Ps in the lesson
	IND37	**IND52**	The teacher makes use of teaching methods that activate the Ps
Adaptation of teaching	**IND41**	**IND61**	The teacher adapts the instruction to the relevant differences between Ps
	IND42	**IND62**	The teacher adapts the assignments and processing to the relevant differences between Ps
Teaching learning strategies	**IND51**	**IND71**	The teacher ensures that the teaching materials are orientated towards transfer
	IND52	**IND72**	The teacher stimulates the use of control activities
	IND53	**IND73**	The teacher provides interactive instruction and activities
Efficient classroom management	**IND12**	**IND81**	The teacher ensures the orderly progression of the lesson
	IND11	**IND82**	The teacher gives a well structured lesson
	IND13	**IND83**	The teacher uses learning time efficiently
	IND14	**IND84**	The teacher ensures effective classroom management
	Deleted	**IND91**	The teacher ensures classroom layout supports the P activities
	Deleted	**IND92**	The teaching environment is educational and contemporary
	IND71	**IND100**	**I judge the overall quality of teaching as.**

Five factors, two new theoretical factors (i.e., *Safe and stimulating learning climate* and *Clear instruction*) and three original factors (i.e., *Adaptation of teaching, Teaching learning strategies*, and *Efficient classroom*

64 Van de Grift (2007) used the term scale instead of factor used here. It is rather unusual to have a scale that consists of only two items as Factor Four (Adaptation of Teaching). To be consistent, the term scale is used to refer to the whole instrument, while its dimensions, either theoretical or empirical, has been named as factor, component or criterion in different contexts.

management), proposed by van de Grift (2007) simplified the old theoretical underlying structure of nine criteria. A model based on this new underlying structure of the instrument was tested by Maximum Likelihood estimation in LISREL 8.72. Generated in 13 iterations, the solution for this new theoretical model in Figure 5.10 (or Model C, with indicator number showed as in the 2007 version of QoT in Table 5.19) seemed to fit the data better than the previous models[65]. Most of the goodness-of-fit indices of this solution improved (χ^2 =64.17, df=220, *p*=1.0; RMSEA=0.0 with 90%CI=0.0, 0.0 and p-value for CFit= 1.0; NFI= 0.98; CFI=1.00; SRMR=0.040, IFI=1.06; RFI=0.98; Critical N = 404.75).

Figure 5.10: **A CFA model based on FIVE theoretical factors and TWENTY THREE indicators with standardised coefficients (N=76) [Model C]**



KEYS:

NC 1 = Efficient classroom management (Cronbach's alpha=0.93)

NC 2 = Safe and stimulating learning climate (Cronbach's alpha=0.97)

NC 3 = Clear Instruction (Cronbach's alpha=0.88)

NC 4 = Adaptation of teaching (Cronbach's alpha=0.93)

NC 5 = Teaching learning strategies (Cronbach's alpha=0.93)

Whole Scale (Cronbach's alpha=0.98)

65 The ridge option had to be reset because the matrix was not positive definite. This solution was obtained by taking a ridge option at 0.5. Some other ridge options might yield better goodness-of-fit indices but they were rejected because PHI was not positive definite with some insignificant estimates. This solution produced the best goodness-of-fit indices without any insignificant estimates.

Generally, standardised residuals and modification indices looked reasonably good except between Indicator 31 and Indicator 32 where the largest standardised residual (= 2.97) and modification index (=8.83) were found. Standardised parameter estimates from this solution are presented in Figure 5.10 above and Table 5.20 below. All freely estimated unstandardised parameters were statistically significant (p<0.05). Both the factor loadings and square multiple correlations of the indicators of this model were lower than those of Model A and Model B, but similar to those of the three-latent factor model. Except for Indicators 31 and 32, factor loading estimates were between 0.65 and 0.72, suggesting that most indicators were moderately strong in their relations with their purported latent factors (square multiple correlations or R^2 ranged from 0.42 to 0.51).

Table 5.20: **Factor loadings of a CFA model based on FIVE theoretical factors and TWENTY THREE indicators (N=76) [Model C]**

2007 Scale Name	Indicator No. 2007(2004)	Indicator Description (T = The teacher; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Safe and stimulating learning climate	IND21(11)	T ensures a relaxed atmosphere	0.69	Fixed at 1	0.48
	IND22(12)	T promotes mutual respect	0.68	1.20	0.46
	IND23(13)	T supports the self-confidence of Ps	0.69	1.13	0.48
	IND24(14)	T shows respect for the Ps in behaviour and language use	0.69	1.76	0.47
	IND25(21)	T ensures cohesion	0.68	0.86	0.46
	IND26(22)	T stimulates the independence of Ps	0.66	0.96	0.43
	IND27(23)	T promotes cooperation between Ps	0.68	0.80	0.46
Clear instruction	IND31(31)	T clarifies the lesson objectives at the start of the lesson	0.36	Fixed at 1	0.13
	IND32(32)	T evaluates whether the objectives have been achieved at the end of the lesson	0.35	0.70	0.12
	IND33(41)	T gives clear instructions and explanations	0.65	2.07	0.42
	IND34(42)	T gives clear explanations of the learning materials and the assignments	0.66	1.50	0.44
	IND35(43)	T gives feedback to Ps	0.68	2.17	0.46
	IND36(51)	T involves all Ps in the lesson	0.60	1.15	0.36
	IND37(52)	T makes use of teaching methods that activate the Ps	0.63	1.97	0.40
Adaptation of teaching	IND41(61)	T adapts the instruction to the relevant differences between Ps	0.72	Fixed at 1	0.51
	IND42(62)	T adapts the assignments and processing to the relevant differences between Ps	0.68	0.86	0.46
Teaching learning strategies	IND51(71)	T ensures that the teaching materials are orientated towards transfer	0.65	Fixed at 1	0.42
	IND52(72)	T stimulates the use of control activities	0.69	1.41	0.48
	IND53(73)	T provides interactive instruction and activities	0.66	1.57	0.43
Efficient classroom management	IND11(82)	T gives a well structured lesson	0.65	1.19	0.42
	IND12(81)	T ensures the orderly progression of the lesson	0.65	Fixed at 1	0.42
	IND13(83)	T uses learning time efficiently	0.69	1.79	0.48
	IND14(84)	T ensures effective classroom management	0.65	1.69	0.43

Compared with the previous QoT models, the average variance extracted of each factor as well as that of the whole scale was much lower, varying from 0.33 to 0.49. This indicated convergent validity found for all factors and this model was the lowest among the QoT models. As shown in Table 5.21, reliability scores based on composite reliability of the factors are lower than those based on Cronbach's alpha.

***Table 5.21:* Average variance extracted, composite reliability and Cronbach's alpha reliability of the CFA model based on FIVE theoretical factors [Model C]**

Factor Name	Average Variance Extracted	Composite reliability	Cronbach's Alpha
Safe and stimulating learning climate	0.44	**0.76**	0.93
Clear instruction	0.46	**0.86**	0.97
Adaptation of teaching	0.33	**0.77**	0.88
Teaching learning strategies	0.49	**0.66**	0.93
Efficient classroom management	0.44	**0.71**	0.93
Whole Scale	0.42	0.94	0.98

There was little change in the reliability for the whole scale with only Indicator 24 excluded in this model after Model B (Cronbach's alpha = 0.98)[66]. As shown in Figure 5.10, the reliability for the three original factors, *Adaptation of teaching, Teaching learning strategies*, and *Efficient classroom management*, remained the same as in Model A and Model B. The reliability for the new factor *Safe and stimulating learning climate* (Cronbach's alpha=0.97) was very high and seemed to be slightly above the average of the reliability of the two theoretical criteria that form this factor (i.e., an average of Cronbach's alpha of 0.98 and 0.93). Similarly, the reliability for the new factor *Clear instruction* (Cronbach's alpha=0.88) seemed to be 5% above the average of the reliability of the three theoretical criteria that form this factor (i.e., an average of Cronbach's alpha of 0.76, 0.92, and 0.85). The item-total statistics of this factor indicates that its reliability would increase if either Indicator 31 or Indicator 32 was to be deleted. This suggests that the coherence or internal consistency of this factor would be enhanced if these two indicators are excluded.

Figure 5.10 also shows high intercorrelations between the new theoretical factors, suggesting that they might be very weak in discriminant validity. Insufficient discriminant validity was confirmed in Table 5.22, which

66 A split-half test was run with similar results: Cronbach's alpha for First Half is 0.96 and the Second Half is 0.97. Spearman-Brown Equal Length and Unequal Length are both 0.95 and Guttman Split-Half coefficient is 0.96.

shows that for each factor, the average variance explained was much lower than higher than its shared variances above the diagonal. These results suggest that this new modified theoretical QoT model is the weakest among all QoT models in terms of, convergent and discriminant validity as well as composite reliability.

Table 5.22: **Average variance extracted and shared variance estimates of a CFA model based on FIVE theoretical criteria in 2007 [Model C]**

Factor Name	Safe and stimulating learning climate	Clear instruction	Adaptation of teaching	Teaching learning strategies	Efficient classroom management
Safe and stimulating learning climate	**0.44**	0.76	0.90	0.67	0.92
Clear instruction	0.87	**0.46**	0.92	0.67	0.77
Adaptation of teaching	0.95	0.96	**0.33**	0.77	0.88
Teaching learning strategies	0.82	0.82	0.88	**0.49**	0.85
Efficient classroom management	0.96	0.88	0.94	0.92	**0.44**

Note: Correlations are below the diagonal, squared correlations are above the diagonal and average variance explained estimates are presented on the diagonal.

Taken the results of three models and the three-factor model together, it is rather surprising to find that the current data lent support for all these models. There are minor discrepancies in terms of reliability of the individual underlying dimensions of each model as well as reliability of estimation of individual models surrounding with the problem of generating goodness-of-fit solutions from a small sample size. Since a parsimonious model is always more desirable, Model C and the three-factor CVCP (QoT) Model are thus considered preferable. However, although Model C may have provided a solution that has theoretical grounds as well as considerable empirical support[67], it lacks adequate convergent and discriminant validity. This means that the parsimony principle sometimes may be misleading in scale development. Again, these results clearly indicate that comparisons of empirical and theoretical models should not be based only on CFA results and justify further comparisons based on other different analyses as those presented in Chapter 6. Meanwhile, these models also suggest three deeper problems.

67 More importantly, van de Grift (2007) also presented the new version of QoT with results of four different countries. These results would be useful for further comparisons to be discussed later in the next chapter.

First, it seems to be a fact that effective teaching is multi-dimensional. For example, effective teaching concerns clear instruction and presentation, effective classroom management, stimulating lesson climate and others. Thus, this leads to the second problem concerning the distinctiveness of these dimensions. It seems that some dimensions are less distinctive than others because they are more easily integrated together than others, especially when they are highly correlated. It may be of theoretical interest to keep different dimensions as distinctive as possible, but in reality these dimensions have to be integrated to make teaching and learning possible. This means that different dimensions are less likely to measure as distinctive dimensions as they would be in theories. Moreover, the distinctiveness of dimensions may vary across contexts and across samples, or both. Finally, if distinctiveness of dimensions is such a difficult thing to measure, a better instrument would be one that can capture it better. Thus, it is expected that comparisons of the results of the two instruments would illuminate their relative reliability and validity in measuring distinctiveness of dimensions of effective teaching in the next chapter.

5.7 The extent of similarities and discrepancies between CVCP and ECP factors

As the major research aim of the present study is to identify some common dimensions of effective classroom practices, it is assumed that two samples in CVCP and ECP studies are comparable on these dimensions. That is, it is expected that the two samples would show some generic characteristics of effective teaching behaviours that can characterise across samples and contexts while some idiosyncratic characteristics may be considered as unique properties that define the samples or their contexts. As in the last chapter, the underlying dimensions of the two empirical CFA models in the two studies were compared factor by factor.

Factor Integrated classroom management and climate

Table 5.23 below shows the combination of indicators of the first factor of the CVCP model, *Integrated classroom management and climate.* Sharing six indicators with the first factor of the QoT model in the ECP study, *Supportive and stimulating lesson climate* in the factors makes them highly comparable. This factor also has indicators that were in three different factors in the ECP model (i.e., *Well organized lessons with clear objectives*, *Proactive lesson management* and *Adaptation of teaching*). These additional indicators of different criteria in the original scale make this CVCP factor more global than was the ECP factor. Despite the diversity of its indicators, this factor has very high internal consistency (Cronbach's alpha=0.98).

Table 5.23: **Shared indicators on Factor Integrated class management and climate in CVCP and ECP factors of CFA models for QoT indicators**

CVCP Factor (N=76)	Indicator No.	Indicator Description	ECP Factor (N=79)	Indicator No.	Indicator Description
Integrated Class Management & Climate (Cronbach's alpha = 0.98)	***IND11***	*T ensures a relaxed atmosphere*	Supportive and stimulating lesson climate (Cronbach's alpha= 0. 87)	***IND11***	*T ensures a relaxed atmosphere*
	IND12	*T promotes mutual respect*		***IND12***	*T promotes mutual respect*
	IND13	*T supports the self-confidence of Ps*		***IND13***	*T supports the self-confidence of Ps*
	IND14	T shows respect for the Ps in behaviour and language use			
	IND21	*T ensures cohesion*		***IND21***	*T ensures cohesion*
	IND22	*T stimulates the independence of Ps*		***IND22***	*T stimulates the independence of Ps*
	IND23	*T promotes cooperation between Ps*		***IND23***	*T promotes cooperation between Ps*
				IND24	There is good individual involvement by the Ps
	IND42	*T gives clear explanations of the learning materials and the assignments*	Well organized lessons with clear objectives (Cronbach's alpha= 0. 73)	***IND42***	*T gives clear explanations of the learning materials and the assignments*
	IND43	*T gives feedback to Ps*		***IND43***	*T gives feedback to Ps*
	IND51	*T involves all Ps in the lesson*	Proactive lesson management (Cronbach's alpha= 0. 88)	***IND51***	*T involves all Ps in the lesson*
	IND61	*T adapts the instruction to the relevant differences between Ps*	Adaptation of teaching (Cronbach's alpha= 0. 90)	***IND61***	*T adapts the instruction to the relevant differences between Ps*
	IND84	T ensures effective classroom management			

Factor Structured teaching skills

Similarly in Table 5.24, the second factor of the CVCP model, *Structured teaching skills*, is dominant by indicators of two criteria of the original scale (i.e., *Teaching learning strategies* and *Effective classroom organization*) and thus comparable to the two factors of the ECP model (i.e., *Teaching learning strategies* and *Proactive lesson management*) that were associated with these theoretical criteria. Again, this factor consists of indicators of several other factors in the ECP model and shows similar very high internal consistency (Cronbach's alpha=0.97).

Table 5.24: **Shared indicators on Factor Structured teaching skills in CVCP and ECP factors of CFA models**

CVCP Factor N=76	Indicator No.	Indicator Description	ECP Factor N=79	Indicator No.	Indicator Description
Structured Teaching Skills (Cronbach's alpha = 0.97)	***IND24***	*There is good individual involvement by the Ps*	**Supportive and stimulating lesson climate (Cronbach's alpha= 0. 87)**	***IND24***	*There is good individual involvement by the Ps*
	IND41	*T gives clear instructions and explanations*	**Well organized lessons with clear objectives (Cronbach's alpha= 0. 73)**	***IND41***	*T gives clear instructions and explanations*
	IND52	T makes use of teaching methods that activate the Ps			
	IND62	*T adapts the assignments and processing to the relevant differences between Ps*	**Adaptation of teaching (Cronbach's alpha= 0. 90)**	***IND62***	*T adapts the assignments and processing to the relevant differences between Ps*
	IND71	*T ensures that the teaching materials are orientated towards transfer*	**Teaching learning strategies (Cronbach's alpha= 0. 69)**	***IND71***	*T ensures that the teaching materials are orientated towards transfer*
	IND72	*T stimulates the use of control activities*		***IND72***	*T stimulates the use of control activities*
	IND73	*T provides interactive instruction and activities*		***IND73***	*T provides interactive instruction and activities*
	IND81	*T gives a well structured lesson*	**Proactive lesson management (Cronbach's alpha= 0. 88)**	***IND81***	*T gives a well structured lesson*
	IND82	*T ensures the orderly progression of the lesson*		***IND82***	*T ensures the orderly progression of the lesson*
	IND83	*T uses learning time efficiently*		***IND83***	*T uses learning time efficiently*

Factor Effective class/Lesson planning

Table 5.25 reveals that the third factor of the present QoT model, *Effective class/Lesson planning* is combination of two factors of the ECP model (i.e., *Well organized lessons with clear objectives* and *Effective classroom layout*).

Table 5.25: **Shared items on Factor Effective class/lesson planning in CVCP and ECP factors of CFA models for QoT**

CVCP Factor N=76	Indicator No.	Indicator Description	ECP Factor N=79	Indicator No.	Indicator Description
Effective Class/Lesson Planning (Cronbach's alpha = 0.85)	***IND31***	*T clarifies the lesson objectives at the start of the lesson*	**Well organized lessons with clear objectives (Cronbach's alpha= 0. 73)**	***IND31***	*T clarifies the lesson objectives at the start of the lesson*
	IND32	*T evaluates whether the objectives have been achieved at the end of the lesson*		***IND32***	*T evaluates whether the objectives have been achieved at the end of the lesson*
	IND91	*T ensures that classroom layout supports the P activities*	**Effective classroom layout (Cronbach's alpha= 0. 79)**	***IND91***	*T ensures that classroom layout supports the P activities*
	IND92	*The teaching environment is educational and contemporary*		***IND92***	*The teaching environment is educational and contemporary*

5.8 Conclusion

This chapter has described the quantitative classroom observation data collected via the QoT instrument and variation across the 76 lessons observed for the four Hong Kong teachers. The key finding of the confirmatory factor analyses in this chapter suggests that the underlying dimensions of observed teachers' teaching behaviours for the Hong Kong case study sample seems to be very similar to those in the ECP study, but look more global in term of structure. Whether in a global or specific form, dimensions like clear instruction and presentation, effective classroom management, and stimulating lesson climate appear to be important dimensions of effective classroom practices that show the qualities of basic, generic and replicable" variables. In addition to concluding that *effective teaching is most likely a multidimensional process*, it is also clear that *different dimensions of teaching behaviours may vary considerably across lessons*. Thus, the identification of these dimensions leads into further analyses to identify their relative impacts on overall judgment of teaching quality made by the rater and the individual involvement of students found in the lesson observed by the rater. The results of these analyses are presented in the next chapter. The identified underlying dimensions of effective classroom practices are also used to compare variation between teachers and to identify exemplar lessons that were rated highly and lowly for integrating with qualitative findings to enrich the case descriptions in Chapter 7.

The current CVCP (QoT) Model is well supported by the data. However, the various models examined in Section 5.6 indicate that the data also showed considerable support for these models. Obtaining goodness-of-fit for theoretical models is an exciting finding for instrument development, but it also poses a problem of selecting the best model. If the parsimonious principle is to be applied, the CVCP (QoT) Model is most attractive, but its rather global factor structure also means that the underlying dimensions may not be as distinctive as it has been hypothesised. The results regarding the convergent and discriminant validity of the various models indicate the relative strengths of the empirical and theoretical models and justify comparisons of these models should not be limited to only CFA results but extended further to other different analyses. Thus, like the ISTOF instrument, results of further analyses to explore their relative strengths are presented in the next chapter.

Together with the results of using ISTOF as observation instrument, we are able to pursue not only the construct validity of a single instrument, but also the relative strengths and weaknesses of the instruments as both seem to be ecologically valid and capable of identifying underlying dimensions of effective classroom practices from the observed teachers' behaviours. Issues surrounding instrument variation are thus addressed in a greater depth in the next chapter.

Finally, the CFA models for the instruments in the CVCP and the ECP studies will be cross-validated in both samples in the next chapter. Given the different characteristics of the two samples, it is likely that the two models will not be well supported by the data of another sample. However, as it is also hypothesised that some common dimensions should underlie the teacher's behaviours of both samples, it would be of interest to examine the extent to which the two models are supported in the data of different samples.

In the next chapter, results of this chapter and the last chapter will be compared and then related to two overall indicators of teaching effectiveness. The various CFA models reviewed in these chapters are compared in terms of their predictability of these indicators.

CHAPTER 6 : MULTIPLE DIMENSIONS OF TEACHING QUALITY AND THEIR RELATIVE IMPACTS, INSTRUMENT VARIATION AND SAMPLE VARIATION

6.1 Introduction

In the last two chapters, several distinctive underlying dimensions[68] of effective classroom practices have been identified through confirmatory factor analyses. The results seem to confirm two views: first, teaching is a multidimensional construct and second, teachers vary in strengths in these different dimensions and variation in these teacher practices occurs across lessons. The existence of distinctive dimensions of teaching that is consistently found across two different samples and contexts is important as this support the view that some sets of teaching behaviours are likely to be "basic, generic and replicable" (Teddlie et al., 2006). That is, a generic concept of teaching effectiveness can be supported by empirical evidence as demonstrated in Chapters 5 and 6. However, this concept does not necessarily contradict the alternative differentiated concept of teaching effectiveness, according to which a teacher may not be equally strong in every dimension of teaching and his/her strengths in these different dimensions may vary across lessons when contexts and students change as also illustrated in the two preceding chapters.

The main foci of this chapter are 1) to compare the underlying dimensions and the various good-fitted models identified in the last two chapters; and 2) to see how far such dimensions can predict two global indicators of overall judgement of teaching quality and the overall individual involvement, which are likely to influence both learning and attainment. Comparisons were made using Pearson correlation and multiple regression in SPSS 16 and cross-validation of CFA models in LISREL 8.72.

68 It should be noted that different names were used for different levels of units in previous studies:

Instruments	Basic Unit	Higher Unit	Highest Unit
ISTOF	Item	Indicator	Component
QoT Old	Indicator	Criterion	
QoT New	Indicator	Factor	
CVCP	Item/indicator	Factor /Dimension	

Section 6.2 examines how the various underlying dimensions in each model are related to each other in the model and how these dimensions are associated with the two global indicators of overall teaching effectiveness. Thus, bivariate correlations between dimensions and between dimensions and indicators of teaching effectiveness highlight the relative strength of their relationships.

Based on the results of Section 6.2, it is natural to ask how the various dimensions of effective classroom practices identified may have contributed to the predictions of the two global indicators of overall teaching effectiveness, namely, the judgment of overall teaching quality and the judgment of good individual involvement of pupils. Here the two indicators of teaching effectiveness are expected to be related and affected by different classroom practices. The overall judgment of teaching quality conforms to an inspection model of professionals who evaluate the teachers' behaviours based on their knowledge and experience, while the good individual involvement of pupils refers to the engagement of observed pupils' behaviours as well as the less observable cognitive activities. This indicator is more associated with a theory of learning. Section 6.3 reports the results of multiple regression analyses using these two indicators of teaching effectiveness as dependent variables and different dimensions of effective classroom practices as predictors in each CFA model to explore the relative impacts of different dimensions as predictors and eventually the relative predictability of different models.

Following the logic of inquiry in Section 6.3, Section 6.4 addresses the relative predictive power of the two instruments in relating to the two indicators of teaching effectiveness in multiple regression. The focus of comparisons is between the two instruments, rather than between the models of the same instruments. These comparisons do not necessarily imply the superiority of one of the classroom observation instruments, but their relative explanatory power of the whole instruments or their individual dimensions in their relations to the two indicators of teaching effectiveness.

Finally, Section 6.5 presents the results of analyses using LISREL 8.72 to cross-validate the CFA models in the CVCP study with the ECP sample.

These results are important as they show the extent to which the CFA models can be generalised to other samples, despite the inherent characteristics, especially differences, in the characteristics of the samples.

6.2 Relationships between dimensions of observed teachers' behaviours and indicators of teaching effectiveness

In order to explore the relationship between underlying dimensions of various models and indicators of teaching effectiveness, Pearson correlation analyses were used[69]. In general, all were positively correlated and nearly all were statistically significant at $p<0.001$ in the one-tail test. These results show that effective classroom practices tend to be strongly positively correlated with each other. However, very strong correlations (i.e., correlation coefficient reaches 0.8 or above) found between some underlying dimensions may also suggest that there may be a problem of multicollinearity (Tabachnick & Fidell, 2006). Yet, despite their strong correlations, teachers' classroom behaviours shown in the last two chapters varied differently along these dimensions across lessons. The results of the relationships between underlying dimensions of each model are presented in the following sections.

6.2.1 Relationships between dimensions of effective classroom practices of CVCP (ISTOF) Model and indicators of teaching effectiveness

Table 6.1 below shows that all the six dimensions of observed teachers' behaviours identified in CVCP (ISTOF) Model are fairly strongly associated (most with Pearson correlation above 0.5) and their relationships are statistically highly significant. Among the six dimensions, the dimension *Student Engagement* is strongly associated with more dimensions, with correlations ranging from 0.63 to 0.82. In contrast, the dimensions *Strategies to enhance learning and lesson focus* and *Meta-cognitive skills teaching* appear to be relatively weakly associated with other dimensions, as their correlations with them ranged similarly between 0.49 and 0.67 (i.e., 22.4% to 28.6% lower than those found for the above two dimensions).

69 It should be noted that the correlation used in this chapter would be different from those presented in the last two chapters, where the correlation between factors are based on CFA models that also take into account of measurement errors. Comparisons of the two sets of correlation values seem to show that Pearson correlations tend to be lower.

Table 6.1: **Relationships between dimensions of observed teachers' behaviours in CVCP (ISTOF) Model and indicators of teaching effectiveness as indicated in the Pearson correlation matrix (N=76)**

Dimension/indicator Name	Meta-cognitive skills teachin	Classroom manage-ment & climate	Different-iation & support	Clarity & logic of present-ation	Student engage-ment	Strategies to enhance learning & lesson focus	Overall quality of teaching	Good individual involvement by pupils
Meta-cognitive skills teaching	1.00							
Classroom Manage-ment & Climate	0.67**	1.00						
Differentiation & support	0.62**	***0.76*** **	1.00					
Clarity & Logic of Presentation	0.67**	0.67**	***0.74****	1.00				
Student engagement	0.67**	***0.71****	***0.82****	***0.78****	1.00			
Strategies to enhance learning and lesson focus	*0.49***	*0.47***	0.67**	0.63**	0.65**	1.00		
Overall quality of teaching	0.63**	***0.74****	***0.80****	***0.73****	***0.82****	0.67**	1.00	
Good individual involvement by pupils	0.69**	***0.77****	***0.81****	***0.74****	***0.84****	0.53**	***0.77****	1.00

Note: ** p<.001

The correlation between the two indicators of teaching effectiveness is strong (*r=0.77, p<0.001*). Both indicators show even stronger associations with two dimensions, *Student engagement* and *Differentiation and support*, suggesting the relative importance of these factors. The indicator *Overall quality of teaching* is less strongly associated with the dimension *Meta-cognitive skills teaching* (*r=0.63, p<0.001*), while the indicator *Good individual involvement by the pupils* is less strongly associated with the dimension *Strategies to enhance learning and lesson focus* (*r=0.53, p<0.001*).

6.2.2 Relationships between dimensions of effective classroom practices of the seven-component theoretical ISTOF Model and indicators of teaching effectiveness

In Table 6.2 below, the relationships between the seven components of teachers' behaviours in the theoretical model of ISTOF generally seem to be stronger than those between the dimensions of the CVCP (ISTOF) Model as discussed in the last section because no correlation is below 0.5 and their relationships are statistically highly significant. Among the seven components, *Differentiation and inclusion*, *Instruction skills,* and *Classroom climate* seem to be strongly associated with more other components with correlation coefficient ranging from 0.68 to 0.87. Interestingly, the component *Classroom management* is less strongly associated with most components, as its correlation coefficients with other components ranged only between 0.59 and

0.73 (i.e., 15.3% to 19.2% lower than those of the three components mentioned).

Table 6.2: **Relationships between dimensions of observed teachers' behaviours in theoretical ISTOF Model and indicators of teaching effectiveness as indicated in the Pearson correlation matrix (N=76)**

Dimension/indicator Name	Assessment and Evaluation	Different-iation and Inclusion	Clarity of Instruction	Instruct-ional Skills	Promoting active learning & developin g meta-cognitive skills	Classroom Climate	Classroom Manage-ment	Overall quality of teaching	Good individual involve-ment by pupils
Assessment and evaluation	1.00								
Differentiation and inclusion	***0.72****	1.00							
Clarity of instruction	***0.77****	***0.74****	1.00						
Instructional skills	***0.83****	***0.85****	***0.81****	1.00					
Promoting active learning and developing metacognitive skills	***0.76****	0.69**	***0.70****	***0.79****	1.00				
Classroom climate	***0.78****	***0.86****	***0.74****	***0.87****	***0.74****	1.00			
Classroom management	0.59**	0.68**	0.60**	0.68**	***0.74****	***0.73****	1.00		
Overall quality of teaching	0.69**	***0.83****	***0.76****	***0.79****	0.69**	***0.83****	***0.71****	1.00	
Good individual involvement by pupils	***0.71****	***0.83****	***0.73****	***0.86****	***0.75****	***0.83****	***0.74****	***0.77****	1.00

Note: ** p<.001

Both the global indicators of teaching effectiveness show strong associations with three components, *Differentiation and inclusion*, *Instruction skills, and Classroom climate*, suggesting their relative importance. Two components, *Instruction skills* and *Promoting active learning and developing metacognitive skills,* seem to be less strongly associated (about 8.8% weaker) with the indicator *Overall quality of teaching* (*r=0.79* and *0.69*, respectively) than with the indicator *Good individual involvement by the pupils* (*r=0.86* and 0.75, respectively).

6.2.3 Relationships between dimensions of effective classroom practices of CVCP (QoT) Model and indicators of teaching effectiveness

The relationships between the three dimensions of observed teachers' behaviours in Table 6.3 below seem to be either very strong (*r=0.89*) or only moderate (*r=0.39* to 0.48), though all relationships are statistically highly significant. This is a rather different pattern to those described above for the ISTOF dimensions.

Table 6.3: **Relationships between dimensions of observed teachers' behaviours in CVCP (QoT) Model and indicators of teaching effectiveness as indicated in the Pearson correlation matrix (N=76)**

Dimension/indicator Name	Integrated class management & climate	Structured teaching skills	Effective class/ lesson planning	Overall quality of teaching	Good individual involvement by the pupils
Integrated class management & climate	1.00				
Structured teaching skills	***0.89*****	1.00			
Effective class/lesson planning	*0.39***	*0.48***	1.00		
Overall quality of teaching	***0.80*****	***0.86*****	0.60**	1.00	
Good individual involvement by pupils	***0.86*****	***0.92*****	*0.34***	***0.77*****	1.00

Note: ** p<.001

Interestingly, the dimension *Effective class/lesson planning* is associated with the other two dimensions much less strongly, as their correlation coefficients were only 0.39 and 0.48, that is, 85.4% to 128.2% weaker than the relationship between dimensions *Integrated class management and climate* and *Structured teaching skills*. Both indicators of teaching effectiveness show strong associations with only two dimensions, namely, *Integrated class management and climate* and *Structured teaching skills*, suggesting their particular importance. Both dimensions seem to be slightly more strongly associated (about 8.7% to 8.9% weaker) with the global indicator *Good individual involvement by the pupils* (*r=0.86* and *0.92*, respectively) than with the global indicator *Overall quality of teaching* (*r=0.80* and *0.86*, respectively). Between the two dimensions, *Structured teaching skills* shows slightly stronger associations with both indicators than does *Integrated class management and climate.*

6.2.4 Relationships between dimensions of effective classroom practices of the eight-criterion theoretical QoT Model (Model B) and indicators of teaching effectiveness

Table 6.4 below shows that except the criterion *Clear objectives,* the relationships between other criteria of teachers' behaviours in the theoretical model of QoT generally are very strong because no correlation coefficient is below 0.7 and their relationships are statistically highly significant.

Table 6.4: **Relationships between dimensions of observed teachers' behaviours in eight-criterion theoretical QoT Model (Model B) and indicators of teaching effectiveness as indicated in the Pearson correlation matrix (N=76)**

Dimension/ indicator Name	Safe & orderly school climate	Stimulating learning climate	Clear Objectives	Clear instruction	Activating Pupils	Adaptation of teaching	Teaching learning strategies	Effective classroom organisa-tion	Overall quality of teaching	Good individual involve-ment by pupils
Safe & orderly school climate	1.00									
Stimulating learning climate	0.90**	1.00								
Clear objectives	0.34**	0.80**	1.00							
Clear instruction	0.85**	0.89**	0.37**	1.00						
Activating pupils	0.78**	0.85**	0.26*	0.83**	1.00					
Adaptation of teaching	0.70**	0.80**	0.61**	0.73**	0.71**	1.00				
Teaching learning strategies	0.79**	0.87**	0.45**	0.79**	0.82**	0.85**	1.00			
Effective classroom organisation	0.80**	0.85**	0.43**	0.84**	0.81**	0.73**	0.86**	1.00		
Overall quality of teaching	0.75**	0.82**	0.60**	0.74**	0.70**	0.76**	0.82**	0.85**	1.00	
Good individual involvement by pupils	0.79**	0.92**	0.35**	0.84**	0.85**	0.75**	0.89**	0.83**	0.77**	1.00

Note: ** p<.001

Apparently, *Clear objectives* is the only criterion that is associated with five criteria in much less strength, with their correlation coefficients ranging from *0.26 (p<0.05)* to *0.45 (p<0.001)*, but it shows stronger associations with the criteria *Adaptation of teaching* (*r=0.61, p<0001*) and with *Stimulating learning climate* (*r=0.80, p<0001*). In contrast, criterion *Stimulating learning climate* is associated with all criteria most strongly, as the coefficients of their correlations were all above 0.80. In comparison with the criterion *Stimulating learning climate*, the criterion *Effective classroom organisation* shows slightly lower strength in its associations with other criteria, particularly with criterion *Clear objectives* (where *r*=0.43, about 86% weaker than that between the criteria *Stimulating learning climate* and *Clear objectives*).

The fact that the strongest relationship (*r*=0.90, p<0.001) is found between the criteria *Safely and orderly school climate* and *Stimulating learning climate* seems to justify the theoretical consideration to combine the two criteria in the new version of QoT (see Section 5.6). By contrast, the present correlation results in Table 6.4 do not seem to support the idea to combine the criterion *Clear objectives* with the criteria *Clear instruction* and *Activating pupils* on empirical grounds because its connections with them are not strong (*r=0.37* and *0.26*, respectively). Both indicators of teaching effectiveness show strong associations with most criteria except *Clear objectives*. However, between the two indicators, while the indicator *Overall quality of teaching* may have shown stronger relationship with criterion *Clear objectives* (*r=0.60, p<0.001*), the indicator *Good individual involvement by the pupils* shows stronger relationships with the criteria *Stimulating learning climate, Clear instruction*, *Activating pupils* and *Teaching learning strategies* (about 8.5% to 21.4% stronger than those of the indicator *Overall quality of teaching* and these criteria).

6.2.5 Relationships between dimensions of effective classroom practices of the five-factor theoretical QoT Model (Model C) and indicators of teaching effectiveness

Overall, as seen in Table 6.5, the relationships between the five factors of teachers' behaviours in the theoretical model of QoT (Model C) are found to be stronger than those between the dimensions of the CVCP (QoT) Model and between the criteria of Model B discussed in the last two sections because no correlation has a coefficient below 0.76 and all relationships are statistically highly significant.

Table 6.5: **Relationships between dimensions of observed teachers' behaviours in five-factor theoretical QoT Model (Model C) and indicators of teaching effectiveness as indicated in the Pearson correlation matrix (N=76)**

Dimension/ indicator Name	Efficient classroom management	Safe and stimulating learning climate	Clear instruction	Adaptation of teaching	Teaching learning strategies	Overall quality of teaching	Good individual involvement by pupils
Efficient classroom management	1.00						
Safe and stimulating learning climate	***0.83****	1.00					
Clear instruction	***0.88****	***0.90****	1.00				
Adaptation of teaching	***0.73****	***0.76****	***0.80****	1.00			
Teaching learning strategies	***0.86****	***0.82****	***0.86****	***0.85****	1.00		
Overall quality of teaching	***0.85****	***0.79****	***0.80****	***0.76****	***0.82****	1.00	
Good individual involvement by pupils	***0.83****	***0.83****	***0.88****	***0.75****	***0.89****	***0.77****	1.00

Note: ** p<.001

Among the five factors, *Clear instruction,* shows the strongest associations with other factors with correlation coefficients at 0.86 or above. Like the factor *Clear instruction*, the factor *Teaching learning strategies* is another key factor in this model that is more strongly associated with other factors as well as the two indicators of teaching effectiveness. The "weakest" yet strong relationship is found between the factors *Efficient classroom management* and *Adaptation of teaching* (*r=0.76, p<0.001*). The factor *Adaptation of teaching* shows relatively weaker correlations with other factors as well with the two indicators of teaching effectiveness. Indicator *Overall quality of teaching'* shows stronger associations with the factors *Efficient classroom management* and *Teaching learning strategies,* while the indicator *Good individual involvement by the pupils* is more strongly associated with the factors *Clear instruction* and *Teaching learning strategies*. However, comparing the two indicators, the indicator *Overall quality of teaching* tends to have relatively weaker relationships (by about 5% to 10%) than does the indicator *Good individual involvement by the pupils*.

From the above results, it seems that the underlying factors of the theoretical model of ISTOF and those of the new theoretical model of QoT (Model C) show stronger correlations to one another as well as with the two global indicators of teaching effectiveness. The findings for Model C were very similar to what were obtained by van de Grift (2007) earlier. However, it is not clear whether a model with more closely correlated underlying factors is necessary better than a model that shows relatively weaker correlations between the factors. However, correlations only measure associations rather than any testing of possible causal links. Nonetheless, two things are clear: first, some underlying factors are related to other factors as well as with the two global indicators more strongly than others. Second, it seems that the underlying factors affected the two indicators in different strengths.

6.3 Relative impacts of the dimensions of effective teaching on indicators of teaching effectiveness

Multiple regression analyses[70] were employed to explore the relative impacts of underlying dimensions of a model on the two indicators of teaching effectiveness, namely, the *Overall quality of teaching'* and the *Good individual involvement of the pupils* as rated by the field researcher. Unlike the Pearson correlations, where high and statistically significant correlations

70 After inspecting the distributions of the various factors of the models and their correlations, it was decided that linear multiple regression, rather than ordinal regression, would be robust and appropriate for the current exploratory analyses presented in this and the subsequent sections.

are often desirable, high intercorrelations between independent variables, like the relationships of the dimensions of various models examined in the last section, often pose a challenge of multicollinearity that may bias the interpretation of solutions found statistically significant[71]. It was recognised that the stepwise methods may require good theoretical reasons to specify the order of entering predictions (Hair et al., 2006) and that stepwise methods may yield biased estimations (Muijs, 2004), it was decided to adopt the Enter method. Accordingly, for each model, standard multiple regression analysis was performed initially using the default Enter method of LINEAR REGRESSION function in SPSS 16 to identify predicators that might contribute significantly to the prediction such that their unique contributions could be confirmed again in the subsequent analyses using the Backward and then the Forward stepwise methods. It was intended that these methods would sufficiently exhaust all possible models that can be generated, while taking into consideration of potential contributions of individual predictors and potential multicollinearity. For all solutions, scatter plots not presented, here were checked to explore whether any outlier might have seriously biased the estimations.

6.3.1 Relative impacts of dimensions of effective teaching of CVCP (ISTOF) Model on indicators of teaching effectiveness

Overall quality of teaching

Standard multiple regression was conducted with the overall quality of teaching (i.e., IND100) as the dependent variable and the six underlying dimensions of observed teachers' behaviours as independent variables. Multiple *R* for regression was shown statistically significant: $F(6, 69)= 37.34, p<.001, R^2$ *adj.=.74*. Only three of the six independent variables contributed significantly to the prediction of the overall judgement of teaching quality. Regression results summarized in Table 6.6 below presents the results from the default Enter analysis. This indicates that *Student Engagement* was the strongest predictor with the highest standardised beta value and t-value ($\beta=.37;\ t(69)=3.07,\ p<.001$), followed by *Classroom Management and*

71 The problem of multicollinearity has been addressed in length in various texts (see Hair et al., 2006; Myers, Gamst & Guariono, 2006; Muijs, 2004; Stevens, 2002; Tabachnick & Fidell, 2006).

Climate (*β=.25; t(69)=2.47, p<.005*) and *Strategies to enhance learning and lesson focus (β=.19; t(69)=2.23, p<.005)*.

Table 6.6: **Relative impacts of dimensions of observed teachers' behaviours in CVCP (ISTOF) Model on the overall quality of teaching as indicated in the regression analysis using the default Enter method**

Dependent: Overall quality of teaching	***B***	***SE B***	***β***
(Constant)	0.30	0.23	
Meta-cognitive skills teaching	0.01	0.10	.01
Classroom management and climate	***0.22***	0.09	***.25******
Differentiation and support	0.15	0.13	.15
Clarity & Logic of Presentation	0.05	0.13	.04
Student engagement	***0.39***	0.13	***.37*******
Strategies to enhance learning and lesson focus	***0.31***	0.14	***.19******

Note: R^2=.77 (N=76, ps<.001); *** p<.005, **** p<.001

The respective unique variances of these variables were 3.2%, 2.7% and 2.1%. For other variables, their standardised coefficient beta values are too small to show any significant impact on the prediction. Despite the strong correlations identified between some of the predictors (see Section 6.2.1), both collinearity statistics and collinearity diagnostics indicated that there seemed to be no serious problem of multicollinearity. Casewise diagnostics indicated that no lesson seemed to be an outlier as all standardised residuals were within the +/-2.5 limits. Subsequent stepwise results also identified the same dimensions as the only statistically significant predictors.

Good individual involvement of the pupils

Six underlying dimensions of observed teachers' behaviours were again used as independent variables in the standard multiple regression analysis to predict individual involvement of the pupils (i.e., IND24). The multiple regression was shown to be statistically significant: *F(6, 69)= 44.33, p<.001,* R^2 *adj.=.78*. Only two of the six independent variables contributed significantly to the prediction of the individual involvement of pupils using the Enter method. Regression results summarized in Table 6.7 below indicate that with the highest standardised beta value and t-value, *Student engagement* was again the strongest predictor (*β=.40; t(69)=3.59, p<.001*), followed by *Differentiation and support* (*β=.29; t(69)=2.5, p<.05*).

Table 6.7: **Relative impacts of dimensions of observed teachers' behaviours in CVCP (ISTOF) Model on the overall quality of teaching as indicated in the regression analysis using the default Enter method**

Dependent: Good individual involvement of the	*B*	*SE B*	*β*
(Constant)	0.22	0.27	
Meta-cognitive skills teaching	0.18	0.11	.13
Classroom management and climate	0.18	0.10	.17
Differentiation and support	***0.36***	0.14	***.29****
Clarity and Logic of Presentation	0.15	0.15	.10
Student engagement	***0.53***	0.15	***.40*******
Strategies to enhance learning and lesson focus	-0.27	0.16	-.13

Note: R^2=.79 (N=76, ps<.001); * p<.05, **** p=.001

The respective unique variances of these variables were small, only at 3.8% and 1.9%. For other variables, their standardised coefficient beta values are too small to show any significant impact on the prediction. Both collinearity statistics and collinearity diagnostics indicated that there seemed to be no serious problem of multicollinearity, although some strong correlations were identified between some of the predictors (see Section 6.2.1). Casewise diagnostics indicated that one lesson seemed to be an outlier as its standardised residual was slightly below the 2.5 limit[72].

Both the Backward and Forward stepwise method produced different results. The Backward stepwise solution included *Meta-cognitive skills teaching* as a predictor in the model, but its contribution was found insignificant (p=.085). In contrast, in the Forward stepwise solution, Classroom *management and climate* and *Differentiation and support* were the second and third statistically most significant predictors. In other words, the role of *Classroom management* was stronger while that of *Differentiation and support* became weaker. Overall, the total unique variances accounted by the three variables in the Forward stepwise solution was relatively larger (i.e., about 10.6 %) as other variables were left out in the model.

In general, the dimensions of CVCP (ISTOF) seemed to predict the ratings of overall teaching quality and good individual involvement of the pupils well. Among the dimensions, *Student engagement* is most important in terms of its contribution to the predictions. The dimension *Classroom management and climate* may be also important in predicting the results of both indicators (i.e., when the Forward stepwise solution is adopted), but its

72 Its standardised residual, predicted value and actual value were -.2.62; 3.12, and 2.00, respectively.

significance is relatively smaller comparing to the dimension *Student engagement.* This is not surprising because this suggests that, to be effective, the teacher has to go beyond basic skills in maintaining appropriate classroom management and climate. S/he has to be able to engage students in learning. The differences in the predictable power of two dimensions, *Strategies to enhance learning and lesson focus* and *Differentiation and support,* for different indicators of teaching effectiveness seems to suggest that these indicators are affected by similar but slightly different variables.

6.3.2 Relative impacts of dimensions of effective classroom practices of the seven-component theoretical ISTOF Model on indicators of teaching effectiveness

Overall quality of teaching

The seven theoretical components of teachers' behaviours of the ISTOF theoretical model were used as independent variables and the overall judgement of teaching quality (i.e., IND100) as the dependent variable for conducting the standard multiple regression analysis. Multiple regression using the Enter method was found statistically significant: $F(7, 68)= 33.69, p<.001, R^2 adj.=.75$. Only two of the seven independent variables, only the components *Differentiation and inclusion* and *Clarity of instruction* were found to contribute significantly to the prediction of the overall judgement of teaching quality.

In Table 6.8, where regression results were summarized, the component *Differentiation and inclusion* is the strongest predictor with the highest standardised beta value and t-value ($\beta=.38; t(68)=3.00, p<.005$), while the component *Clarity of instruction* is the second strongest predictor ($\beta=.24; t(68)=2.21, p<.05$). These variables accounted for small unique variances, only at 3.0% and 1.6%, respectively. Standardised coefficient and beta values of other variables (including the negative ones) are too small to show any significant impact on the prediction in the standard solution. None of the collinearity statistics and collinearity diagnostics suggested that multicollinearity might be a concern. As indicated in casewise diagnostics, no lesson seemed to be an outlier as all standardised residuals were well within the +/-2.5 limits.

Table 6.8: **Relative impacts of dimensions of observed teachers' behaviours in the seven-component theoretical ISTOF Model on the overall quality of teaching as indicated in the multiple regression analysis using the default Enter method**

Dependent: Overall quality of teaching	*B*	*SE B*	*β*
(Constant)	0.35	0.28	
Assignment and evaluation	-0.04	0.11	-.04
Differentiation and inclusion	***0.28***	0.09	***.38****
Clarity of instruction	***0.25***	0.11	***.24****
Instructional skills	-0.19	0.14	-.23
Promoting active learning and developing metacognitive skills	0.13	0.09	.17
Classroom climate	0.22	0.12	.29
Classroom management	0.10	0.07	.15

Note: R^2=.78 (N=76, ps<.001); *p<.05, ***p<.005

The Backward stepwise solution yielded similar results, with the components *Differentiation and inclusion* and *Clarity of instruction* as the strongest and the second strongest predictor. Two other variables, *Classroom management* and *Classroom climate,* were also included in the final step of regression, but both failed to make significant contributions to the prediction. However, the component *Classroom climate* also appeared in the Forward stepwise solution as the strongest predictor in the model, but its contribution was as significant (*p=.005*) as the other two variables. The amounts of unique variance attributable to these predictors were almost the same, 2.8% for the component *Classroom climate* and 2.9% for the component *Differentiation and inclusion*. The third predictor, the component *Clarity of instruction* only explained a unique variance of about 1.6%. Thus, the results between the standard and the Forward stepwise solutions are larger. These results indicated that components, like *Classroom climate* and *Differentiation and inclusion,* which showed high correlations with most other components (see Section 6.2.2) tended to be strong predictors, but it was the component *Clarity of instruction*, rather than the component *Instructional skills*, which appeared to be another important predictor.

Good individual involvement of the pupils

To predict the good individual involvement of the pupils (i.e., IND24), the seven theoretical components of teachers' behaviours of the theoretical ISTOF model were again used as independent variables in the standard multiple regression analysis. Again statistically significant multiple regression was evident: *F(7, 68)= 37.64, p<.001,* R^2 *adj.=.77*. Among the seven components, only *Differentiation and inclusion* and *Clarity of Instruction skills* were found contributing significantly to the prediction. Regression results obtained y the default Enter method are summarized in Table 6.9.

Table 6.9: **Relative impacts of dimensions of observed teachers' behaviours in the seven-component theoretical ISTOF Model on the good individual involvement of the pupils as indicated in the multiple regression analysis using the default Enter method**

Dependent: Good individual involvement of the pupils	***B***	***SE B***	***β***
(Constant)	0.18	0.33	
Assignment and evaluation	-0.12	0.13	-.10
Differentiation and inclusion	***0.24***	0.11	***.27****
Clarity of instruction	0.01	0.13	.01
Instructional skills	***0.35***	0.16	***.34****
Promoting active learning and developing metacognitive skills	0.15	0.11	.16
Classroom climate	0.11	0.14	.12
Classroom management	0.13	0.08	.16

Note: R^2=.80 (N=76, ps<.001); *p<.05

With the slightly higher standardized beta value and t-value (*β= .34; t(68)= 2.18, p<.05*), the component *Instructional skills* seemed to be a stronger predictor than the component *Differentiation and inclusion* (*β=.27; t(68)= 2.19, p<.05*). Both components uniquely accounted for about 1.4% of variance respectively. Both collinearity statistics and collinearity diagnostics suggested no serious problem of multicollinearity, though some strong correlations were noted between the components *Instructional skills* and *Differentiation and inclusion* with other components (see Section 6.2.2). Casewise diagnostics indicated that no lesson seemed to be an outlier as its standardised residual was well within the +/-2.5 limits.

Both the Backward and Forward stepwise solutions differed from the standard solution in their inclusion of the component *Classroom management* as the third statistically significant predictor, while the components *Differentiation and inclusion* and *Clarity of instruction* remained as the strongest and the second strongest predictors. The amounts of unique variance attributable to these predictors in the stepwise solutions were also more distinctive than they were in the standard solution.

As in the results of the CVCP (ISTOF) Model, the teacher's differentiated support and inclusion strategies (i.e., *Differentiation and inclusion*) is found to be an important theoretical compone*nt that predicts* *enhanced students' participation. Referring back to* Table 4.15 in Section 4.6.1, it would be clear that three items (namely, Items 5, 6, and 16) of the underlying dimension *Student engagement* of the CVCP (ISTOF) Model were originated in the components *Instruction skills* and *Differentiation and inclusion,* the same theoretical components shown here as the most important predictors of student involvement.

6.3.3 Relative impacts of dimensions of effective classroom practices of the CVCP (QoT) Model on indicators of teaching effectiveness

Overall quality of teaching

The three underlying dimensions of observed teachers' behaviours of the CVCP (QoT) Model were employed as independent variables to predict the overall quality of teaching (i.e., IND100) in standard multiple regression analysis. Regression results using the default method are summarized in Table 6.10. All the three independent variables contributed significantly to the prediction of the overall judgement of teaching quality (*$F(3,72)$ =90.19, p<.001, R^2 adj. =.78*). However, among these dimensions, *Structured teaching skills* showed its prominent importance in prediction with a much higher standardised beta value and t-value, (*β=.53; $t(72)$=4.22, p<.001*) with a unique variance about 5.2%.

Table 6.10: **Relative impacts of dimensions of observed teachers' behaviours in CVCP(QoT) Model on the overall quality of teaching as indicated in the multiple regression analysis using the default Enter method**

Dependent: Overall quality of teaching	***B***	***SE B***	***β***
(Constant)	-0.12	0.21	
Integrated class management and climate	***0.36***	0.18	***.24****
Structured teaching skills	***0.80***	0.19	***.53*******
Effective class/lesson planning	***0.52***	0.13	***.25*******

Note: R^2=.79 (N=76, ps<.001). *p<.05 and ****p<.001

The dimensions *Integrated class management and climate* and *Effective class/lesson planning* showed similar standardised beta values (*β=.24* and *.25* respectively) but different t-values (*$t(72)$=1.99, p=.05* and *$t(72)$=4.01, p<.001*, respectively) probably due to their different amounts of unique variance explained (1.2% and 4.7% respectively). These results showed that the dimension *Effective class/lesson planning* is a better predictor than the dimension *Integrated class management and climate*. Both collinearity statistics and collinearity diagnostics indicated that there might be no serious multicollinearity in the solution. However, casewise diagnostics showed that the standardised residuals of two out of 76 lessons were above 2.5, suggesting that they might be outliers[73].

The Forward stepwise method produced a different result in which the dimension *Integrated class management and climate* was excluded as its significance (*p=.00505*) was slightly above the probability-of-F-to-enter criterion (p<=0.05). Given such a small difference in significance, it was considered better to include the dimension in the model. The results of the Backward stepwise method produced the same results as the default method.

The regression results showed that the CVCP (QoT) Model seemed to be consistent with the bivariate correlation results discussed in Section 6.2.3, which indicated the importance of the dimension *Structured teaching skills,* but correlation results could not reveal the lower unique contribution of the dimension *Integrated class management* in the prediction.

Good individual involvement of the pupils

All the three underlying dimensions of observed teachers' behaviours of the CVCP (QoT) Model were employed as independent variables to predict the good individual involvement of the pupils (i.e., IND24) in standard multiple regression analysis. However, since the dimension Structured teaching skills also incorporated the dependent variable (see Table 5.10), it would be inappropriate to use the original average factor score in the prediction. It was also recognised that leaving this dimension from the prediction would be inappropriate because this dimension was especially important for a model with only three factors. Accordingly, the average factor score of this dimension was recomputed such that it would not include the factor loading of IND24.

As indicated in Table 6.11, all three dimensions of the model were found contributing significantly to the prediction in the multiple regression performed ($F(3,72) = 122.72, p<.001, R^2$ *adj.* $=.83$): *Structured teaching skills*

73 The standardised residuals of the two lessons were 2.71 and 2.55. The actual scores were 3, while the predicted scores were 2.07 and 2.12.

(Recomputed) (*t(72)=6.62 p<.001*), *Integrated class management and climate* (*t(72)=2.67 p<.01*), and *Effective class/lesson planning* (*t(72)=-2.20 p<.05)*. Both collinearity statistics and collinearity diagnostics indicated that there might be no multicollinearity for this solution. However, casewise diagnostics showed that the standardised residuals of two out of 76 lessons were beyond the +/- 2.5 limits[74], suggesting that these lessons were outliers. As further multiple regression analyses with these lessons excluded did not produce more plausible models[75], it seemed that the dimension *Structured teaching skills* is the single best predictor of the good involvement of the pupils.

Table 6.11: **Relative impacts of dimensions of observed teachers' behaviours in CVCP (QoT) Model on the good individual involvement of the pupils as indicated in the multiple regression analysis using the default Enter method**

Dependent: Good individual involvement of the pupils	*B*	*SE B*	*β*
(Constant)	-0.18	0.23	
Integrated class management and climate	***0.51***	***0.19***	***.27****
Structured teaching skills (Recomputed)	***1.36***	0.21	***.72*******
Effective class/lesson planning	***-0.32***	0.14	***-.12****

Note: R^2=.84 (N=76, ps<.001); * p<.05, ****p<.001

There are two things unusual in these results. First, the unique variance attributable to the dimension *Structured teaching skills* was 10.0% was more than six times of that attributable to the dimension *Integrated class management and climate* (1.6%) and about ten times of that attributable to the dimension *Effective class/lesson planning* (1.1%). In other words, the significance of the dimension *Structured teaching skills* in the prediction of the good individual involvement of the pupils was prominent. Its prominence was clear even after the Indicator 24, the dependent variable, was removed from the dimension here. As it has been discussed in Section 5.5.2, this dimension has a composite structure which shows a dominance of indicators of *Teaching learning strategies* (Criterion Seven) and *Effective classroom organisation* (Criterion Eight) as well as those indicators describing teacher behaviours facilitating pupil involvement like inclusion (Indicator 62),

74 The standardised residuals of the two lessons were -3.98 and 3.08; the actual scores were 2, while the predicted scores were 3.34 and 0.96.

75 Multiple regression analyses were conducted again with these two cases deleted. The new model with the dimension *Structured teaching skills* as the single predictor showed higher adjusted R^2=.89 (ps.<.001), but the new models with other predictors did not improve the plausibility as the standardised beta value of the dimension *Effective class/lesson planning* remained negative in these models with more new outliers identified. These results suggested deleting the outliers of the current model would not help identifying a better model.

instruction (Indicator 41), and student engagement strategies (Indicators 24 and 52).

Second, the negative standardized beta value of the dimension *Effective class/lesson planning* suggested a negative influence of the dimension on the prediction, which was rather unexpected and difficult to interpret, though its unique variance was relatively small. These results were replicated in the solutions using the Forward and Backward stepwise methods. However, the Forward stepwise method also showed a solution ($F(2,73) = 172.57$, $p<.005$, R^2 *adj.* $=.82$) in the 2nd step with only two statistically significant dimensions *Structure teaching skills* ($t(73)=6.09$ $p<.001$) and *Integrated class management and climate* ($t(73)=2.88$, $p<.005$). As shown in Table 6.12, without the negative effect of the dimension *Effective class/lesson planning*, the standardised coefficient beta of *Structured teaching skills (Recomputed)* was slightly lower, but that of *Integrated class management and climate* increased substantially with stronger significance. These dimensions respectively accounted for about 8.9% and 2.0%. These results suggest that this solution might be less biased than other solutions, regardless of the methods of estimation, which included the dimension *Effective class/lesson planning.*

Table 6.12: **Relative impacts of dimensions of observed teachers' behaviours in CVCP (QoT) Model on the good individual involvement of the pupils as indicated in the multiple regression analysis using the Forward stepwise method**

Dependent: Good individual involvement of the pupils	*B*	*SE B*	*β*
Step 2			
(Constant)	-0.47	0.19	
Structured teaching skills (Recomputed)	***1.20***	0.20	***.63*******
Integrated class management and climate	***0.56***	0.19	***.30******

Note: $R^2=.83$ (N=76, ps<.01); *** p<.005, ****p<.001

6.3.4 Relative impacts of dimensions of effective classroom practices of the eight-criterion theoretical QoT Model (Model B) on indicators of teaching effectiveness

Overall quality of teaching

Eight theoretical criteria of teachers' behaviours of the original QoT Model (Model B) were employed as independent variables to predict the overall quality of teaching (i.e., IND100) in standard multiple regression analysis. These criteria are identical to the first eight theoretical criteria of the original nine-criterion theoretical model (van de Grift et al., 2004). Regression results using the default method are summarized in Table 6.13.

Table 6.13: **Relative impacts of dimensions of observed teachers' behaviours in eight-criterion theoretical QoT Model (Model B) on the overall quality of teaching as indicated in the multiple regression analysis using the default Enter method**

Dependent: Overall quality of teaching	***B***	***SE B***	***β***
(Constant)	-0.14	0.30	
Safe and orderly school climate	0.18	0.14	.17
Stimulating learning climate	0.13	0.22	.11
Clear objectives	***0.42***	0.13	***.25****
Clear instruction	-0.18	0.19	-.12
Activating Pupils	-0.03	0.14	-.03
Adaptation of teaching	0.00	0.13	.00
Teaching learning strategies	0.19	0.16	.17
Effective classroom organisation	***0.61***	0.16	***.49*****

Note: R^2=.81(N=76, ps<.001); ***p<.005, ****p<.001

Statistically significant multiple regression was evident in the prediction based on the eight theoretical criteria: $F(8,67)=36.61, p<.001, R^2$ *adj. =.79*. However, only two criteria showed significant contributions to the prediction. With higher standardised beta value and t-value (*β=.49; t(67)=3.86, p<.001*), the criterion *Effective classroom organization* had a unique variance of about 4.1%. As for the criterion *Clear objectives,* its lower standardised beta value (*β=.25*), t-value (*t(67)=3.21, p<.005*), and unique variance (2.9%) all consistently reflected it as the second strongest predictor. The solution using the Backward stepwise method produced similar results as those of the default method except including two other criteria, *Safe and orderly school climate* and *Teaching learning strategies*, which showed statistically insignificant contributions. However, in the Forward stepwise solution, the criterion *Safe and orderly school climate* showed statistically significant contribution (*β=.21; t(67)=3.47, p<.05*).

The importance of the criterion *Effective classroom organization* as a predictor was suggested in the bivariate correlation results discussed in Section 6.2.4, which the same correlation results did not show equal support for the criteria *Clear objectives* and *Safe and orderly school climate* as key predictors. No serious bias of multicollinearity was found in both collinearity statistics and collinearity diagnostics. However, casewise diagnostics revealed two potential outliers with standardised residuals of slightly below the -2.5 boundary[76].

Good individual involvement of the pupils

The same standard multiple regression procedure was administered with the eight theoretical criteria of teachers' behaviours of the QoT Model (Model B) as independent variables and the good individual involvement of pupils (i.e., IND24) as the dependent variable. Like the dimension *Structured teaching skills* of the CVCP (QoT) Model, the average factor score of the criterion *Stimulating learning climate,* which included the factor loading of IND24 (see Figure 5.8 and Table 5.13), was recomputed before performing the regression analyses. Multiple *R* for regression obtained by the default Enter method was found statistically significant: $F(8,67) = 53.23$, $p<.001$, R^2 *adj.=.85*. As shown in Table 6.14, only two theoretical criteria contributed to the prediction significantly: *Teaching learning strategies* ($\beta=.42$; $t(67)=4.09$, $p<.001$), *Stimulating learning climate (Recomputed)* ($\beta=.57$ $t(67)=2.31$ $p<.05$). Their respective unique variances explained of these criteria are 4.5% and 1.1%. However, the presence of negative beta values for three variables in this prediction posed some difficulties of interpretation, because it was assumed that all betas of these variables should be positive when there was no negative correlation identified between the predictors in Table 6.4.

Table 6.14: **Relative impacts of dimensions of observed teachers' behaviours in eight-criterion theoretical QoT Model (Model B) on the good individual involvement of the pupils in the multiple regression analysis using the default Enter method**

Dependent: Good individual involvement of the pupils	***B***	***SE B***	***β***
(Constant)	-.18	.31	
Safe and orderly school climate	-.23	.15	-.17
***Stimulating learning climate* (Recomputed)**	***.50***	.22	***.33****
Clear objectives	-.18	.14	-.09
Clear instruction	.38	.19	.22
Activating pupils	.22	.15	.15
Adaptation of teaching	-.11	.13	-.09
Teaching learning strategies	***.77***	.16	***.57*****
Effective classroom organisation	.00	.17	.00

Note: R^2=.91(N=76, ps<.001); **p<.0.1, ****p<.001

However, these conflicting results might be related to multicollinearity. The tolerance and variance inflation factor (VIF) statistics for the criteria

76 The standardised residuals of the two lessons were -2.50 and -2.62. The actual scores were 2, while the predicted scores were 2.83 and 2.87.

Stimulating learning climate (Recomputed) are 0.1 and 10.25 and for *Teaching learning strategies* are .14 and 7.31 respectively, indicating the presence of multicollinearity[77]. Further inspection of collinearity diagnostics showed that the condition index was 48.78 and two variance proportions were above 0.50, suggesting the criteria of multicollinearity were met (see Meyers, Gamst & Guarino, 2006; Tabachnick & Fidell, 2006)[78]. These results also revealed the limitations of the default Enter method in its "lack of consideration of such factors as multicollinearity, the identification of outliners and influentials, and the interpretability of the results" (Hair et al., 2006, p.213).

The final solution obtained in the fourth step of the Backward stepwise method had four statistically significant predictors: *Teaching learning strategies*, *Stimulating learning climate (Recomputed), Clear instruction and Clear objectives.* However, this solution was considered improper because two criteria, *Safe and orderly school climate* and *Clear objectives*, in the prediction still showed negative betas. Moreover, when some of the statistically insignificant variables were gradually excluded from the prediction, the VIF for *Stimulating learning climate (Recomputed)* decreased slightly to 9.19, but was still indicative of multicollinearity. This also suggests that some more variables might have to exclude to obtain a proper solution. Accordingly, the Forward stepwise method was attempted.

The three solutions of the Forward stepwise presented in Table 6.15 are different from those of the Enter and Backward stepwise methods, but looked more plausible because there is no negative beta. The multiple regression of the last model produced similar statistics as those obtained by using the Enter method: $F(3,72) = 137.24$, $p < .001$, R^2 *adj.*$=.84$, but interestingly, the three statistically significant criteria did not include the criterion *Stimulating learning climate (Recomputed)* as it was in the solutions obtained by using other methods: *Teaching learning strategies* ($\beta=.51$; $t(72)=6.02$, $p<.001$) *Clear instruction* ($\beta=.24$; $t(72)=2.84$, $p<.01$), and

77 Stevens (2002) suggested a tolerance value of .1or less or VIF at 10 or over as a heuristic for checking multicollinearity, but Hair et al. (2006) were more conservative and indicated that tolerance values below .19 (or above a VIF of 5.3) might be suspicious of multicollinearity.

78 It was suggested that the condition index over 30 and at least two variance proportions above 0.50 be indicative of multicollinearity.

Activating pupils (β=.23; t(72)=2.50, p<.05). These predictors were expected to account for about 7.5%, 1.5% and 1.3% of unique variance. Both collinearity statistics and collinearity diagnostics suggested no immediate concern for multicollinearity in this solution, while casewise diagnostics revealed that three potential outliers with exceptionally low standardised residuals beyond the +/- 2.5 limits[79]. However, an inspection of the scatter plots showed that they did not seem to have biased the estimation.

Table 6.15: **Relative impacts of dimensions of observed teachers' behaviours in eight-criterion theoretical QoT Model (Model B) on the good individual involvement of the pupils in the multiple regression analysis using the Forward stepwise method**

Dependent: Good individual involvement of the pupils	*B*	*SE B*	*β*
Step 1			
(Constant)	.22	.17	
Teaching learning strategies	***1.20***	.07	***.89*******
Step 2			
(Constant)	-.56	.22	
Teaching learning strategies	***.82***	.10	***.61*******
Clear instruction	***.62***	.13	***.36*******
Step 3			
(Constant)	-.50	.22	
Teaching learning strategies	***.69***	.11	***.51*******
Clear instruction	***.43***	.15	***.24******
Activating Pupils	***.33***	.13	***.23****

Note: R^2=.79 (ps<.001) for Step1; ΔR^2= .05 (ps<.001) for Step 2; ; ΔR^2= .01 (ps<.05) for Step 3

Despite the discrepancies found in solutions of these methods, the prominence of the criterion *Teaching learning strategies* was confirmed because none of the other variables contributed more than 2% of unique variance. The solutions of the Forward stepwise method also suggest that the criterion *Stimulating learning climate* may not be associated strongly associated with pupils' individual participation.

6.3.5 Relative impacts of dimensions of effective classroom practices of the five-factor theoretical QoT Model (Model C) on indicators of teaching effectiveness

Overall quality of teaching

Standard multiple regression was conducted with the five factors of teachers' classroom behaviours of the new theoretical QoT Model (Model C)

79 The results of the case diagnostics were as follow:

Lesson No.	Standardised residual	Actual Score	Predicted Value	Residual
7	-2.52	3	3.90	-.90
9	-3.44	2	3.23	-1.23
71	2.84	2	0.99	1.01

as independent variables and the overall quality of teaching (i.e., IND100) as the dependent variable. Table 6.16 shows a statistically significant multiple regression solution obtained by the default Enter method that showed the factor *Efficient classroom management* as the single statistically significant predictor (*β=.52; t(70)=3.80, p<.001*): *F(5,70) =46.60, p<.001,* R^2 *adj.=.75*.

Table 6.16: **Relative impacts of dimensions of observed teachers' behaviours in five-factor theoretical QoT Model (Model C) on the overall quality of teaching as indicated in the multiple regression analysis using the default Enter method**

Dependent: Overall quality of teaching	*B*	*SE B*	*β*
Constant	0.41	0.25	
Efficient classroom management	***0.77***	0.20	***.52*******
Safe and stimulating learning climate	0.28	0.19	.21
Clear instruction	-0.22	0.37	-.10
Adaptation of teaching	0.27	0.14	.21
Teaching learning strategies	0.13	0.19	.10

Note: R^2=.77(N=76, ps<.001); ****p<.001

However, the tolerance values for four factors were below 0.19 (or a VIF above 5.3), suggesting there might be multicollinearity. An inspection of the collinearity diagnostics (condition index=42.84 plus more than two variance proportions over 0.5.) further confirmed multicollinearity in the solution obtained using the default method. Accordingly, the Backward and Forward stepwise methods were employed to explore other plausible models. The final solution in four steps of the Backward stepwise methods indicated that factors *Efficient classroom management* and *Adaptation of teaching* were the only two predictors that yielded statistically significant contributions. As shown in Table 6.17, the Forward stepwise solution produced similar results in two steps:

Table 6.17: **Relative impacts of dimensions of observed teachers' behaviours in five-factor theoretical QoT Model (Model C) on the overall quality of teaching as indicated in the multiple regression analysis using the Forward stepwise method**

Dependent: Overall quality of teaching	*B*	*SE B*	*β*
Step 1			
Constant	0.35	0.19	
Efficient classroom management	***1.26***	0.09	***.85*******
Step 2			
Constant	0.32	0.18	
Efficient classroom management	***0.94***	0.13	***.63*******
Adaptation of teaching	***0.37***	0.10	***.29*******

Note: R^2=.71 (N=76, ps<.001) for Step1; ΔR^2= .04 (ps<.001) for Step 2. ****p<.001

Two statistically significant models were found. With the factor *Efficient classroom management* as the only predictor, the model of Step 1 was statistically significant: $F(1,74) = 187.77$, $p < .001$, R^2 *adj.* $= .71$. The addition of the factor *Adaptation of teaching* as a predictor in the Step 2 model was also statistically significant: $F(2,73) = 114.27$, $p < .001$, R^2 *adj.* $= .75$. Between these two predictors, the dimension *Efficient classroom management* contributed more to the prediction with its much higher standardised beta value and t-value ($\beta = .63$; $t(73) = 7.51$, $p < .001$), accounting for a unique variance of about 18.7%. By contrast, the factor *Teaching learning strategies* only accounted for 4.1% of unique variance with its lower ($\beta = .29$; $t(73) = 3.5$, $p < .001$). The relative importance of the factors *Efficient classroom management* and *Adaptation of teaching* as predictors was *not* suggested in the bivariate correlation results discussed in Table 6.5 of Section 6.2.5, which shows that these two factors are highly correlated with the overall judgement of teaching quality ($r = 0.85$ and 0.76, respectively), but their correlations are not the highest. Both collinearity statistics and diagnostics indicated multicollinearity was not present as a serious bias in this solution. No lesson was identified as potential outliers with standardised residuals of exceeding the +/ -2.5 boundaries in the casewise diagnostics.

Good individual involvement of the pupils

Multiple regression was conducted with the five factors of observed teachers' classroom behaviours of the new theoretical QoT Model (Model C) as independent variables and the good individual involvement of pupils (i.e., IND24) as the dependent variable[80]. Regression results using the default Enter method summarized in Table 6.18 displays a statistically significant model ($F(5,70) = 78.26$, $p < .001$, R^2 *adj.* $= .84$) with only two predictors *Teaching learning strategies* ($\beta = .61$; $t(70) = 5.07$, $p < .001$) *and Clear instruction* ($\beta = .45$; $t(70) = 3.29$, $p < .005$). However, both collinearity statistics and collinearity diagnostics of this solution were indicative of multicollinearity. The tolerance values of four factors were below 0.19 (or a VIF above 5.3) and the condition index was 42.84 with more than two variance proportions over 0.5. There were also 3 out of 76 lessons as outliers with standardised residuals of exceeding the +/ -2.5 boundaries in the casewise diagnostics.

80 It should be noted that none of the factors of QoT Model C included IND24 of the original scale in 2004 version. Accordingly, no recomputation was required prior to trhe multiple regression analysis.

Table 6.18: **Relative impacts of dimensions of observed teachers' behaviours in five-factor theoretical QoT Model (Model C) on the good individual involvement of the pupils as indicated in the multiple regression analysis using the default Enter method**

Dependent: Good individual involvement of the pupils	B	SE B	β
Constant	-0.70	0.25	
Efficient classroom management	-0.06	0.20	-.03
Safe and stimulating learning climate	0.12	0.18	.07
Clear instruction	***1.21***	0.37	***.45****
Adaptation of teaching	-0.24	0.14	-.16
Teaching learning strategies	***0.98***	0.19	***.61*****

Note: R^2=.77(N=76, ps<.001); ***p<.005, ****p<.001

Again, the Backward and Forward stepwise methods were performed to explore other plausible models. The final solution in the third step of the Backward stepwise method included the factor *Adaptation of teaching* as a statistically *insignificant* predictor in addition to the same two factors in solution of the Enter method as statistically significant predictors. Table 6.19 shows the Forward stepwise solution that produced results similar to the Enter method but more clearly in two steps:

Table 6.19: **Relative impacts of dimensions of observed teachers' behaviours in five- factor theoretical QoT Model (Model C) on the good individual involvement of the pupils as indicated in the multiple regression analysis using the Forward stepwise method**

Dependent: Good individual involvement of the pupils	*B*	*SE B*	*β*
Step 1			
Constant	0.22	0.17	
Teaching learning strategies	***1.44***	0.09	***.89*****
Step 2			
Constant	-0.69	0.24	
Teaching learning strategies	***0.83***	0.15	***.52*****
Clear instruction	***1.18***	0.24	***.44*****

Note: R2=.79 (N=76, ps<.001) for Step1; Δ R2= .05 (ps<.01) for Step 2. ****p<.001

In Step 1, the dimension *Teaching learning strategies* was the only predictor: *F(1,74) =278.74, p<.001,* R^2 *adj. =.79*. In Step 2, the additional predictor that contributed significantly to the prediction was the dimension *Clear instruction*: *F(2,73) =193.46, p<.001,* R^2 *adj. =.84*. Both the standardised beta value and t-value of these two predictors were close: *β=.52; t(73)=5.73, p<.001* for the dimension *Teaching learning strategies* and *β=.44; t(73)=4.85, p<.001* for the dimension *Clear instruction.* Accordingly, the amounts of unique variance attributable to them were also close: 7.1% and 5.1%, respectively. Bivariate correlation results in Table 6.5 of Section 6.2.5 were compatible with the current multiple regression results. Both predictors were most highly correlated with the dependable variable, the indicator *Good individual involvement of the pupils* (*r=0.89* and *0.88*, respectively). Despite an overall high correlation results, both collinearity statistics and collinearity diagnostics indicated multicollinearity did not seem to have biased this solution. However, two potential outliers with standardised residuals of exceeding the+/ -2.5 boundaries were identified in the casewise diagnostics[81].

6.3.6 Relative strengths of different models in terms of their explanatory power and predictability on indicators of teaching effectiveness

Overall quality of teaching

Derived from previous tables (i.e., Tables 6.6, 6.8, 6.10, 6.13, and 6.17), Table 6.20 compares the standardised beta values, unique variance explained of each predictor (i.e., dimension, component, criterion or factor in different models) and the adjusted *R* square of each model. These results indicate the relative strengths of the predictors and the models in terms of their explanatory power and predictability on the overall quality of teaching.

Table 6.20: **Relative strengths of different models in terms of their explanatory power and predictability on the overall quality of teaching (N=76)**

Dependent: Overall quality of teaching	***β***	***Unique variance explained***	***Adjusted R^2***
CVCP (ISTOF) Model (Enter)			.74
Student engagement	***.37*******	***3.2%***	
Classroom management and climate	.25***	2.7%	
Strategies to enhance learning and lesson focus	.19***	1.7%	
Seven-component theoretical ISTOF Model (Enter)			.75
Differentiation and inclusion	***.38******	***3.0%***	
Clarity of instruction	.24*	1.6%	
CVCP (QoT) Model (Enter)			.78
Structured teaching skills	***.53*******	***5.2%***	
Effective class/lesson planning	***.25*******	***4.7%***	
Integrated class management and climate	.24*	1.7%	
Eight-criterion theoretical QoT Model (Model B) (Enter)			.79
Effective classroom organisation	***.55*******	***4.1%***	
Clear objectives	.29***	2.9%	
Five-factor theoretical QoT Model (Model C) (Forward)			.75
Efficient classroom management	***.63*******	***18.7%***	
Adaptation of teaching	.29****	4.1%	

Note: *p<.05, **p<.01, ***p<.005, ****p<.001

81 The standardised residuals of the two lessons were -3.26 and 2.80. The actual scores were 2, while the predicted scores were 3.19 and .98, respectively.

Regarding the two models of ISTOF in Table 6.20, the theoretical ISTOF model has a probably negligible advantage in adjusted R square over the CVCP (ISTOF) Model, but the amount of unique variance explained by its predictors, either individually or in total, is slightly smaller than that of the predictors of the CVCP (ISTOF) Model. Thus, it seems that the CVCP (ISTOF) model is slightly stronger than the theoretical model as its predictors have a little better explanatory power and predictability in predicting the overall judgment of teaching quality. It should be noted that the dimension *Student engagement* has a composite structure which also includes two items of the theoretical component *Differentiation and inclusion* (see Table 4.15). However, given the structural differences between the predictors of these two models, further comparison may be required (see Section 6.4.1 below) to establish their relative strengths.

Evaluating the three models of QoT in Table 6.20 is less straightforward. In terms of the predictability of the whole model, the original eight-criterion theoretical QoT model is the best, but in terms of the amount of unique variance explained by the predictors, the new five-factor theoretical model on average is higher. Consisting of the same indicators, the dimensions *Effective classroom organisation* and *Efficient classroom management* are essentially the same theoretical construct under different names or labels (see Section 5.6.2 and Table 5.19 for details). The explanatory power of this construct also varied only slightly in different models. It should be noted that its 18.7% unique variance shown in Table 6.20 might have been inflated by the Forward method. Although the solution of the Enter method might have been biased by multicollinearity, the unique variance of this factor (i.e., 4.8%) appeared to be similar to those in Model A and Model B. The criterion *Effective classroom organisation* is the most important predictor of the original theoretical model, but becomes a little more important under the new label as *Efficient classroom management,* when it was compared to the new construct *Safe and stimulating learning climate* in Model C, which combined the dimensions *Safe and orderly school climate* and *Stimulating learning climate.* This suggests that the new construct of Model C may not have a better construct validity, though it may be motivated and consistent with the view that classroom management is crucial to ensure basic teaching quality but positive learning climate may distinguish typical and effective teaching better. In short, it is reasonable to consider the original theoretical model is the least parsimonious, but being parsimonious does not guarantee that the new theoretical model *has* improved the construct validity of the underlying structure of the scale. This tends to contradict the findings of van de Grift (2007) and shows that correlation results he employed are not always reliable to address the predictive validity of the models. Combining the results here and those results on the convergent and discriminant validity of the models in Section 5.6.2 seems to suggest that the Model C is weaker in terms of various types of validity, while Model A and Model B showed relative strengths.

Good individual involvement of the pupils

Like the last table, Table 6.21 compares the standardised beta values, unique variance explained of each predictor (i.e., dimension, component, criterion or factor in different models) as well as the adjusted *R* square derived from previous tables regarding the prediction of individual involvement of the pupils (i.e., Tables 6.7, 6.9, 6.12, 6.15, and 6.19).

Table 6.21: **Relative strengths of different models in terms of their explanatory power and predictability on the good individual involvement of the pupils (N=76)**

Dependent: Good individual involvement of the pupils	***β***	***Unique variance explained***	***Adjusted R^2***
CVCP (ISTOF) Model (Enter)			.78
Student engagement	***.40*****	***3.8%***	
Differentiation and support	.29*	1.9%	
Seven-component theoretical ISTOF Model (Enter)			.77
Instructional skills	.34*	1.4%	
Differentiation and inclusion	. 27*	1.4%	
CVCP (QoT) Model (Enter)			.83
Structured teaching skills (Recomputed)	***.63*****	***8.9%***	
Integrated class management and climate	.30***	2.0%	
Eight-criterion theoretical QoT Model (Model B) (Forward)			.84
Teaching learning strategies	***.51*****	***7.5%***	
Clear instruction (2004 version, 3 indicators)	.24***	1.5%	
Activating Pupils	.23*	1.3%	
Five-factor theoretical QoT Model (Model C) (Forward)			.84
Teaching learning strategies	***.52*****	***7.1%***	
Clear instruction (2007 version, 7 indicators)	.44****	5.1%	

Note: *p<.05, **p<.01, ***p<.005, ****p<.001

The results in Table 6.21 indicate the relative strengths of the predictors as well as the models in terms of their explanatory power and predictability on the good individual involvement of the pupils. Although the CVCP (ISTOF) Model shows a negligible advantage over the theoretical ISTOF model in the adjusted *R* square, the amount of unique variance explained by each of its predictors is larger than that of the predictors of the theoretical ISTOF model.

The inclusion of the dimension *Differentiation and support* and the component *Differentiation and inclusion* respectively in the empirical and theoretical models also confirms the importance of teachers' differentiated supports to cater for diverse needs of students. The relatively stronger explanatory power of the dimension *Student engagement* suggests its relatively strong construct validity and confirms its contribution to make CVCP (ISTOF) model a better model in predicting the overall judgement of teaching quality. The strength of this dimension may be related to the fact that it is the largest dimension of the CVCP (ISTOF) model with seven items (see Table 4.15 and Table 4.22).

Again, evaluating the three models of QoT is less straightforward. As for the predictability of the whole model in terms of adjusted R square, the differences among these models are actually negligible. In terms of the amount of unique variance explained by the predictors, the relative predictability of the dimension *Structured teaching skills* of the CVCP (QoT) model is the largest. The unique variances of the predictors of the new five-factor theoretical model (or Model C) are marginally higher than those of the original theoretical model (or Model B) in total, but contingent on the individual differences among the predictors.

The criterion *Teaching learning strategies* appears to be slightly more important in Model B and its importance still persists in the Model C. However, there is no apparent theoretical reason why the factor *Teaching learning strategies* might have become more important in the new theoretical model. The significance of the new factor *Clear instruction* formed by combining seven indicators of three criteria of the original model (i.e., *Clear objectives, Clear instruction* and *Activating pupils*) seems to have enhanced as reflected in a unique variance larger than that in Model B when it only included three indicators. It is also surprised that that the new construct *Safe and stimulating learning climate* after combining the indicators of the two criteria in the original model, namely, *Safe and orderly school climate* and *Stimulating learning climate,* would have a statistically *insignificant* contribution to the prediction. Thus, it seems that it is better to conclude that all QoT models are tentatively capturing some essential dimensions that may be viewed as necessary to enhance student participation but further results would be needed to examine reveal their relative significance in more depth using larger samples.

6.4 Relative strengths between ISTOF and QoT as indicated in instrument variation in explanatory power and predictability

6.4.1 Relative strengths of the two instruments in predicting the overall judgement of teaching quality

Since it is discussed in the last section that the CVCP (ISTOF) Model seems to be stronger than its theoretical model in the association with the overall judgement of teaching quality, all comparisons between ISTOF and QoT were performed only with the CVCP (ISTOF) Model. Regarding the QoT models, it was decided that all three models would be employed to compare with the CVCP (ISTOF) Model as this would help illuminate the variation among these instruments. In each comparison between the two instruments where standard multiple regression was conducted, the dependent variable was the overall quality of teaching (i.e., IND100) would be, while the independent variables would be the underlying dimensions (or components, criteria or factors in different theoretical models) of teachers' classroom behaviours of the CVCP (ISTOF) Model and one of the three QoT models.

One comparison was attempted between the CVCP (ISTOF) Model and all the three QoT models together. This comparison was intended to explore the relative strengths of these QoT models when they were all present in the multiple regression analysis. However, since the results showed that multiple regression models became unstable, when all QoT models were included for prediction, an additional comparison was run between the CVCP (ISTOF) Model and two theoretical QoT models (i.e., Model B & Model C). Thus, totally five comparisons were made between the two instruments. For these comparisons, the intercorrelations of all the underlying dimensions of various models are shown in Appendix XV.

For each comparison, standard multiple regression analysis was performed initially using the default Enter method of LINEAR REGRESSION function in SPSS 16 to identify predicators that might contribute significantly to the prediction such that their unique contributions could be explored or confirmed again in the subsequent analyses using the Backward and then the Forward stepwise methods. A summary of the mentioned five comparisons of models using different methods of estimations is presented in Table 6.22.[82] This table lists the all the predictors that were found statistically significant and ranks these predictors in terms of their standardised coefficient betas. The fourth column also indicates some problems identified regarding the presence or absence of negative betas, multicollinearity, and outliers.

Table 6.22: **A summary of comparisons of models regarding their explanatory power and predictability on the overall quality of teaching using different methods of estimations**

Comparison	**Method of estimations**	**Statistically significant predictors identified**	**Presence of multicollinearity and outlier(s)**
CVCP (ISTOF) vs CVCP (QoT)	*Enter*	9 variables, 2 sig. predictors: • *Structured teaching skills (QoT);* • *Effective class/lesson planning (QoT)*	• *2 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *No outlier was found*
	Backward	6th of 6 steps, 4 predictors 3 sig. : • **Structured teaching skills (QoT);** • **Effective class/lesson planning (QoT);** • Integrated class management and climate (ISTOF)	• No negative beta was found for this final step solution • Both collinearity statistics diagnostics of this final step solution was NOT indicative of multicollinearity • One lesson as outlier
	Forward	**2nd of 2 steps, 2 sig. predictors:** • **Structured teaching skills (QoT);** • **Effective class/lesson planning (QoT)**	• No negative beta was found • No apparent multicollinearity problems • One lesson as outlier

82 Whenever a method yielded solution(s) indicative of multicollinearity, the corresponding statistically significant predictors are highlighted in *italics*. All significant predictors that will be further discussed in the next table are highlighted in **bold** . Significant predictors found in both stepwise solutions are also highlighted in **bold** when the solutions were not found indicative of multicollinearity. Predictors are ranked by the sizes of their standardised betas here.

Comparison	Method of estimations	Statistically significant predictors identified	Presence of multicollinearity and outlier(s)
CVCP (ISTOF) vs QoT Original Theoretical Model (Model B)	*Enter*	14 variables, 2 sig. predictors: • *Effective classroom organisation(B);* • *Strategies to enhance learning and lesson focus (ISTOF);*	• *6 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *One lesson as outlier*
	Backward	10th of 10 steps, 5 predictors 4 sig.: • **Effective classroom organisation (B)**; • Student engagement (B); • **Strategies to enhance learning and lesson focus (ISTOF);** • **Clear objectives (B)**	• One predictor with a negative beta for this final step solution • Both collinearity statistics and diagnostics were NOT indicative of multicollinearity • One lesson as outlier
	Forward	**4th of 4 steps, 4 sig. predictors:** • **Effective classroom organisation (B);** • **Strategies to enhance learning and lesson focus (ISTOF);** • **Clear objectives (B);** • **Safe and orderly school climate (B)**	• No negative beta was found • No apparent multicollinearity problems • One lesson as outlier
CVCP (ISTOF) vs QoT New Theoretical Model (Model C)	*Enter*	11 variables, 3 sig. predictors: • *Efficient classroom management (C);* • *Strategies to enhance learning and lesson focus (ISTOF);* • *Safe and stimulating learning climate (C)*	• *5 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *One lesson as outlier*
	Backward	6th of 6 steps, 6 predictors, 3 sig.: • *Efficient classroom management (C);* • *Strategies to enhance learning and lesson focus (ISTOF);* • *Safe and stimulating learning climate (C)*	• *2 predictors with a negative beta* • *Collinearity diagnostics of the solution of the last step was indicative of multicollinearity* • *One lesson as outlier*
	Forward	**3rd of 3 steps, 3 sig. predictors:** • **Efficient classroom management (C);** • **Strategies to enhance learning and lesson focus (ISTOF);** • **Safe and stimulating learning climate (C)**	• No negative beta was found • No apparent multicollinearity problems • Two lessons as outliers
CVCP (ISTOF) vs Models B & C	*Enter*	14 of 19 variables, 2 sig. predictors: • *Efficient classroom management (C);* • *Strategies to enhance learning and lesson focus (ISTOF)*	• *7 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of* ***extreme*** *multicollinearity* • *One lesson as outlier*
	Backward	11th of 11 steps, 5 predictors, 4 sig.: • **Efficient classroom management (C);** • Student engagement (B); • **Strategies to enhance learning and lesson focus (ISTOF);** • **Clear objectives (B)**	• One predictor with a negative beta for this final step solution • Both collinearity statistics and diagnostics were NOT indicative of multicollinearity • One lesson as outlier
	Forward	**4th of 4 steps, 4 sig. predictors:** • **Efficient classroom management (B);** • **Strategies to enhance learning and lesson focus (ISTOF);** • **Clear objectives (B);** • **Safe and stimulating learning climate (C)**	• No negative beta was found • No apparent multicollinearity problems • One lesson as outlier

Comparison	Method of estimations	Statistically significant predictors identified	Presence of multicollinearity and outlier(s)
CVCP (ISTOF) vs All QoT Models	*Enter*	17 of 22 variables, 3 sig. predictors: • *Structured teaching skills (QoT)* • *Strategies to enhance learning and lesson focus (ISTOF);* • *Clarity and logic of presentation (B)*	• *10 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of **extreme** multicollinearity* • *One lesson as outlier*
	Backward	8th of 8 steps, 10 predictors, 8 sig.: • *Structured teaching skills(QoT)* • *Teaching learning strategies (C)* • *Safe and stimulating learning climate (C)*	• *4 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of **extreme** multicollinearity* • *One lesson as outlier*
	Forward	***2nd of 4 steps, 2 sig. predictors:*** • ***Structured* teaching skills (QoT);** • **Clear objectives (B)**	• No negative beta was found for this 2nd step solution • Both collinearity statistics and diagnostics were indicative of multicollinearity for solutions in **Steps 3 and 4** • One lesson as outlier

Multiple regression results indicated all models generated using the default Enter method were subject to multicollinearity, suggesting the estimation might be biased. Multicollinearity generally became more serious with extreme collinearity statistics and diagnostics (e.g., VIF and condition index over 100) when, as expected, there were more variables in the prediction. The solutions of the Backward stepwise method seemed to suggest that the inclusion of variables of Model C would lead to multicollinearity. There were some discrepancies between the solutions of the two stepwise methods but they did share many statistically significant predictors. This was also applied to the case of the solutions of the default Enter method. Most of the statistically significant predictors generated by the Enter method were also found in the Forward stepwise solutions. This has provided the rationale to use the results of these solutions in Table 6.23 for further discussions. However, some solutions in the later steps of the stepwise methods did not show multicollinearity, but still included negative standardised coefficients. These solutions, like those in Table 6.14 and in the last two steps in Table 6.15, were considered to be implausible and rejected for the difficulty to interpret these results. Finally, either Lesson 31 or Lesson 64 or both were found as outliers (i.e., their standardised residuals were beyond the +/- 2.5 limits) in most solutions, regardless of the method of estimations. There were not more than two lessons found as outliners.

Scatter plots were checked to ensure the outliers did not pose any challenge to the current solutions.

Table 6.23: **Relative strengths between ISTOF and QoT in terms of explanatory power and predictability on the overall quality of teaching in Forward stepwise regression**

Comparison	Dependent: Overall quality of teaching	B	SE B	β	Unique variance explained	*Adj. R^2*
CVCP (ISTOF) vs CVCP (QoT)	**Step 2 of 2** ***(F(1,74)=205.66, p<.001)***					.77
	(Constant)	.01	.20			
	Structured teaching skills (QoT)	1.12	.10	.74****	**42.4%**	
	Effective class/lesson planning (QoT)	.50	.13	.24****	4.3%	
(Backward)	Integrated class management and climate (QoT)	.38	.18	.25*	1.3%	
CVCP (ISTOF) vs QoT Model B	**Step 4 of 4** ***(F(4,71)=85.42, p<.001)***					.82
	(Constant)	-.56	.21			
	Effective classroom organisation (B)	.64	.11	.51****	**8.6%**	
	Strategies to enhance learning and lesson focus (ISTOF)	.35	.11	.21***	**2.5%**	
	Clear objectives (B)	.32	.10	.19***	2.3%	
	Safe and orderly school climate (B)	.19	.09	.18*	1.2%	
(Backward)	Student engagement (B)	.24	.10	.22*	1.3%	
CVCP (ISTOF) vs QoT Model C	**Step 3 of 3** ***(F(3,72)=102.85, p<.001)***					.80
	(Constant)	-.27	.20			
	Efficient classroom management (C)	.75	.14	.50****	**8.7%**	
	Strategies to enhance learning and lesson focus (ISTOF)	.50	.10	.30****	**6.7%**	
	Safe and stimulating learning climate (C)	.31	.13	.23*	1.6%	
CVCP (ISTOF) vs Models B & C	**Step 4 of 4** ***(F(4,71)=85.92, p<.001)***					.82
	(Constant)	-.51	.21			
	Efficient classroom management (C)	.73	.13	.49****	**7.3%**	
	Strategies to enhance learning and lesson focus (ISTOF)	.36	.11	.22***	**2.6%**	
	Clear objectives (B)	.28	.10	.17*	1.8%	
	Safe and stimulating learning climate (C)	.28	.12	.21*	1.3%	
(Backward)	Student engagement (B)	.24	.10	.22*	1.3%	
CVCP (ISTOF) vs All QoT Models	**Step 2 of 4** ***(F(2,73)=132.26, p<.001)***					.78
	(Constant)	-.13	.22			
	Structured teaching skills (QoT)	1.12	.09	.74****	**42.4%**	
	Clear objectives (B)	.42	.10	.25****	4.8%	

Note: N=76; *p<.05, **p<.01, ***p<.005, ****p<.001; ISTOF = CVCP (ISTOF); QoT= CVCP (QoT), B= QoT Model B; C=QoT Model C; Backward: Backward stepwise method.

Table 6.23 compares the unstandardised and standardised coefficients, unique variance explained of each predictor of the regression model of each comparison as well as the adjusted *R* square of that model. Each regression model was generated using Forward stepwise method. In the 1st, 2nd and 4th comparisons, the Backward stepwise method also generated solutions with an extra statistically significant predictor in addition to those of the Forward stepwise method and the details of these predictors are also included in the table for reference. The criterion *Student engagement* of Model B appeared twice as statistically significant predictors but only in the Backward stepwise solutions. The dimension *Integrated class management and climate* appeared once, but only in the Backward stepwise solution. It is not clear whether these results were biased by the method of estimation.

Except in the last comparison, the solution of the last step of regression generated by the Forward stepwise method generally showed no serious multicollinearity. However, the collinearity statistics and diagnostics indicated the solutions in the 3rd and 4th steps were subject to potential bias by multicollinearity. The relative significance of the predictors changed dramatically (e.g., the dimension *Structured teaching skills* was statistically insignificant in the 4th step) and multicollinearity increased with steps.

There are three key findings in these comparisons. First, all the comparisons in Table 6.23 show a dominance of the QoT predictors, regardless of the QoT model to which these predictors belong. Among all QoT predictors, the dimension *Structured teaching skills* is most prominent, even after taken into account of the possibility of an inflated unique of variance by the Forward stepwise estimation (i.e., 3 times of that using the Enter method, see Section 6.3.3.). However, it should be noted that the indicators of teaching effectiveness used in this chapter are both QoT indicators. Therefore, a dominance of the QoT predictors may reflect only the relative strengths of QoT models in association with the overall judgement of teaching quality. It is likely that the indicators of a high-inference instrument like QoT tend to be more closely related, perhaps because of a halo effect. In other words, the fact that QoT models tend to predict overall judgment of teaching quality better may be because they are constructed to focus on this aspect due to its importance to inspectorate models on which it is based.

Second, the presence of *Strategies to enhance learning and lesson focus* as the only ISTOF dimension in the models of three comparisons suggests its relative importance. Extracted from Table 4.6, Table 6.24 shows the items of this dimension in the EFA analysis. Items 4, 37, and 11 were retained in the CFA analysis in the CVCP study (see Section 4.4 and Table 4.14), and Items 4 and 11 in the CFA analysis in the ECP study (see Table 4.23).

Table 6.24: **Factor formation of Strategies to enhance learning and lesson focus**

Factor Name	Item No.	Item Description
Strategies to Enhance Learning and Lesson Focus (LrnGoal)	**Item 4**	The teacher explains how ***assignments are aligned to the learning goals of the lesson***.
	Item 37	The teacher praises children for ***effort towards realizing their potential***.
	Item 3	***Assignments*** given by the teacher are clearly ***related to what Ss learned***.
	Item 12	The teacher asks Ss to identify the reasons ***why specific activities take place in the lesson***.
	Item 11	The teacher ***clarifies the lesson objectives at the start of the lesson***.

The CFA results also seem to suggest that the goals or lesson objectives might have not been received due attention as a distinctive dimension of effective classroom practices in the development of ISTOF. The criterion *Clear objectives* in the original theoretical QoT model (Model B) refers to teachers' behaviours similar to those referred to by the items of *Strategies to enhance learning and lesson focus.* The unique variance attributable to the dimension *Strategies to enhance learning and lesson focus* in the last comparison is almost the same as it is in the 2nd comparison. However, comparing the 3rd comparison with the fourth comparison, the presence of the criterion *Clear objectives* seems to lower the relative predictability of the criterion *Strategies to enhance learning and lesson focus*, which means some of its explanatory power has been taken up by the criterion *Clear objectives*. The distinctiveness of the criterion *Clear objectives* is lost in the new theoretical QoT model (Model C) because its indicators are combined with those of *Clear instruction* and *Activating pupils* to form a new factor of seven indicators under the label *Clear instruction* (see Section 5.6.2 and Table 5.19). Thus, *Clear objectives* is only loosely correlated with the factor *Clear instruction* of the new theoretical model (r=.48). Further evidence to support the view that the criteria *Strategies to enhance learning and lesson focus* and *Clear objectives* of the original theoretical QoT model are similar can be shown in Appendix XV. These two predictors are most often loosely associated with other predictors with most of the correlation coefficients below .50 (highlighted in red smaller font).

Third, it is rather interesting to see that the significant contribution of the factors of the new theoretical QoT model (Model C). That is, when only the original and the new theoretical models are included in the prediction, predictors of Model C tended to have a stronger impact. However, the opposite is found, when the CVCP (QoT) Model is included in the prediction as it is in the last comparison. Thus, Model C seems to have stronger predictive power, although it may be weaker in terms of convergent and discriminant validity as it has been discussed in the Section 5.6.2. Moreover, the predictors of the 2nd, 3rd and 4th comparisons look almost identical if the models of these predictors are disregarded. This means both variables *Effective classroom organization (or Efficient classroom management)* and *Strategies to enhance learning and lesson focus* are important predictors, but their significant contributions may be taken up by the dimension *Structured teaching skills* in the last comparison.

6.4.2 Relative strengths of the two instruments in predicting the good individual involvement of the pupils

As in the prediction of the overall teaching quality by multiple regression, similar comparisons between ISTOF and QoT were made for predicting the good individual involvement of the pupils and the preliminary results are summarised in Table 6.25 (see also Footnote no. 82 on page 241 for presentation arrangements):

Table 6.25: **A summary of the comparisons of models regarding their explanatory power and predictability on the good individual involvement of the pupils using different methods of estimations**

Comparison	Method of estimations	Statistically significant predictors identified	Presence of negative beta, multicollinearity and outlier(s)
CVCP (ISTOF) vs CVCP (QoT)	*Enter*	9 variables, 3 sig. predictors*:* • *Structured teaching skills (Recomputed) (QoT);* • *Differentiation and support (ISTOF);* • *Effective class/lesson planning (QoT)*	• *3 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *Two lessons as outliers*
	Backward	5th of 5 steps, 5 predictors 4 sig.: • *Structured teaching skills (Recomputed) (QoT);* • *Effective class/lesson planning (QoT);* • *Differentiation and support (ISTOF);* • *Integrated class management & climate (QoT)*	• *2 predictors with a negative beta* • *The final solution showed some multicollinearity* • *Two lessons as outliers*
	Forward	**2nd of 4 steps, 2 sig. predictors:** • **Structured teaching skills (Recomputed) (QoT);** • **Integrated class management and climate (QoT)**	• Only the 3rd and 4th solutions had a predictor with a negative beta • Current solution had no apparent multicollinearity problems • Two lessons as outliers

Comparison	Method of estimations	Statistically significant predictors identified	Presence of negative beta, multicollinearity and outlier(s)
CVCP (ISTOF) vs QoT Original Theoretical Model (Model B)	*Enter*	14 variables, 4 sig. predictors: • *Differentiation and support (ISTOF);* • *Teaching learning strategies (B);* • *Clear objectives (B);* • *Stimulating learning climate (Recomputed) (B);* • *Safe and orderly school climate(B);*	• *5 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *Two lesson as outliers*
	Backward	9th of 9 steps, 6 predictors 6 sig.: • *Teaching learning strategies (B); ;* • *Clear objectives (B);* • *Differentiation and support (ISTOF);* • *Clear instruction(B); ;* • *Stimulating learning climate (Recomputed) (B);* • *Safe and orderly school climate(B);*	• *2 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *One lessons as outlier*
	Forward	**4th of 4 steps, 4 sig. predictors:** • **Teaching learning strategies (B);** • **Clear instruction (B);** • **Activating pupils (B);** • **Differentiation and support (ISTOF);**	• No negative beta was found • No apparent multicollinearity problems; • Two lessons as outliers
CVCP (ISTOF) vs QoT New Theoretical Model (Model C)	*Enter*	11 variables, 3 sig. predictors • *Teaching learning strategies (C);* • *Clear instruction (C);* • *Differentiation and support (ISTOF)*	• *4 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *Two lessons as outliers*
	Backward	8th of 8 steps, 4 predictors 4 sig.: • **Teaching learning strategies (C);** • **Clear instruction (C);** • Adaptation of teaching (C); • Differentiation and support (ISTOF)	• *One predictor with a negative beta* • *The final solution showed no apparent multicollinearity problems* • *Two lessons as outliers*
	Forward	• **2nd of 2 steps, 2 sig. predictors:** • **Teaching learning strategies (C);** • **Clear instruction (C)**	• No negative beta was found • No apparent multicollinearity problems • Two lessons as outliers
CVCP (ISTOF) vs Models B & C	*Enter*	14 of 19 variables, 4 sig. predictors: • *Differentiation and support (C);* • *Teaching learning strategies (C);* • *Clear objectives (B);* • *Stimulating learning climate (Recomputed) (B)*	• *5 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *Two lessons as outliers*
	Backward	9th of 9 steps, 6 predictors 6 sig.: • *Teaching learning strategies (C);* • *Stimulating learning climate (Recomputed) (B);* • *Clear instruction (B);* • *Differentiation and support (ISTOF);* • *Safe and orderly school climate (B);* • *Clear objectives (B)*	• *2 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *One lesson as outlier*
	Forward	**2nd of 4 steps, 2 sig. predictors:** • **Teaching learning strategies (B);** • **Clear instruction (C)**	• Only the 3rd and 4th solutions had one predictor with a negative beta • No apparent multicollinearity problems • Two lessons as outliers

Comparison	Method of estimations	Statistically significant predictors identified	Presence of negative beta, multicollinearity and outlier(s)
CVCP (ISTOF) vs All QoT Models	*Enter*	16 of 22 variables, 1 sig. predictor: • *Differentiation and support (ISTOF)*	• *8 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of extreme multicollinearity* • *Two lessons as outliers*
	Backward	11th of 11 steps, 6 predictors 6 sig.: • *Structured teaching skills (Recomputed) (QoT);* • *Effective classroom organisation (B)* • *Differentiation and support (ISTOF)* • *Stimulating learning climate (Recomputed) (B);* • *Adaptation of teaching (B);* • *Effective class/lesson planning (QoT)*	• *This final solution had 3 predictors with a negative beta* • *Both collinearity statistics and diagnostics were indicative of multicollinearity* • *Two lessons as outliers*
	Forward	**3rd of 5 steps, 3 sig. predictors:** • **Structured teaching skills (Recomputed) (QoT);** • **Activating pupils(B);** • **Differentiation and support (ISTOF)**	• Only the 4th and 5th solutions had a predictor with a negative beta • Current solution without apparent multicollinearity problems • Two lessons as outliers

Again, all multiple regression models generated by the default Enter method were more likely to show multicollinearity, suggesting the estimations of this method might be biased. More predictors with a negative standardised coefficient beta and extreme collinearity statistics and diagnostics (e.g., VIF and condition index over 50) were found when the number of independent variables increased, suggesting multicollinearity increased with the number of variables added into the predictions. Similar to the comparisons for predicting overall teaching quality, those solutions which did not show multicollinearity, but included negative standardised coefficients, were regarded as implausible and rejected because variables were not expected to be negatively correlated.

Contrary to the case for predicting the overall quality of teaching in Table 6.22, the solutions of the Backward stepwise method seemed to suggest that the inclusion of variables of Model B, rather than C, would lead to multicollinearity. The discrepancies between the solutions of the two stepwise methods were more apparent when the Backward stepwise models were biased by multicollinearity. For the comparison between the two CVCP models, the solutions of three methods were the same, but the Forward solution would allow for selecting the solution of the first step to avoid the difficulty to interpret the negative effect (i.e., a negative standardised beta) of the dimension *Effective class/lesson planning.* Interestingly, the dimension *Integrated class management and climate* of the CVCP (QoT) model was no longer statistically significant in the last solution at the fourth step of the Forward stepwise method. Instead, the dimension *Differentiation and support* became statistically significant as it was in the solutions of the Enter method and the Backward stepwise method. However, given that the Forward stepwise solutions can provide more interpretable results, they are summarised and compared in Table 6.26*.* Finally, Lessons 9 and 71 were generally found in all solutions and they appeared to the same lessons (see also Section 7.2.4 for the account of Lesson 9). An examination of the various scatter plots suggested that these outliers had not biased the current results strongly.

Table 6.26: **Relative strengths between ISTOF and QoT in terms of their explanatory power and predictability on the good individual involvement of the pupils (N=76)**

Comparison	Dependent: Good individual involvement of the pupils	*B*	*SE B*	*β*	Unique variance explained	*Adj. R^2*
CVCP (ISTOF) vs CVCP (QoT)	**Step 2 of 4** ***(F(2,73)=172.57, p<.001)***					.82
	(Constant)	-.47	.19			
	Structured teaching skills (Recomputed) (QoT)	1.20	.20	.63****	**8.9%**	
	Integrated class management and climate (QoT)	.56	.19	.30***	*2.0%*	
CVCP (ISTOF) vs QoT Model B	**Step 4 of 4** ***(F(4,71)=110.04, p<.001)***					.85
	(Constant)	-.51	.21			
	Teaching learning strategies (B)	.50	.14	.37****	**2.5%**	
	Clear instruction (B)	.40	.15	.23**	1.5%	
	Activating pupils (B)	.32	.13	.22*	1.2%	
	Differentiation and support (ISTOF)	.23	.10	.18*	1.0%	
CVCP (ISTOF) vs QoT Model C	**Step 2 of 2** ***(F(2,73)=193.46, p<.001)***					.84
	(Constant)	-.69	.24			
	Teaching learning strategies (C)	.83	.15	.52****	**7.1%**	
	Clear instruction (C)	1.18	.24	.44****	**5.1%**	
(Backward)	Adaptation of teaching (C)	-.29	.14	-.19*	0.89%	
	Differentiation and support (C)	.23	.11	.18*	0.89%	
CVCP (ISTOF) vs Models B & C	**Step 2 of 4** ***(F(2,73)= 193.68, p<.001)***					.84
	(Constant)	-.69	.24			
	Teaching learning strategies (B)	.70	.12	.52****	**7.2%**	
	Clear instruction (C)	1.18	.24	.44****	**5.1%**	
CVCP (ISTOF) vs All QoT Models	**Step 3 of 5** ***(F(3,72)=130.26, p<.001)***					.84
	(Constant)	-.33	.17			
	Structured teaching skills (Recomputed) (QoT)	.87	.21	.46****	**3.6%**	
	Activating pupils(B)	.41	.13	.28***	2.1%	
	Differentiation and support (ISTOF)	.29	.10	.23***	1.8%	

Note: *N=76, ****p<.001*; ISTOF = CVCP (ISTOF); QoT= CVCP (QoT); B= QoT Model B; C= QoT Model C; Backward: Backward stepwise method.

Like Table 6.23, Table 6.26 compares the unstandardised and standardised coefficients, unique variance explained of each predictor of the regression model of each comparison as well as the adjusted *R* square of that model. The regression model of each comparison was generated using Forward stepwise method after using the default and Backward stepwise methods in SPSS 16.

Again, the results show a stronger dominance of the QoT models in the prediction because only the dimension *Differentiation and support* of the CVCP (ISTOF) model was found statistically significant in the various comparisons. The results of the first and the third comparisons in Table 6.26 were the same as those results found for the QoT models in Table 6.12 and Table 6.19 and summarised in Table 6.21 above. The results of the second comparison were almost as those in Table 6.15, except the inclusion of the dimension *Differentiation and support* in this statistically significant solution. Given the reasons explained in the last section, these findings become unsurprising. Moreover, despite a higher likelihood of having multicollinearity when the predictors of the Model B were included in the Backward stepwise solutions, they showed relatively stronger predictability over those predictors of Model C, except the factor *Clear instruction*. This was evident in the last two comparisons where two predictors of Model B were found statistically significant. Predictors of Model C, as shown in Table 6.23, played a strong role in predicting the overall quality of teaching, but their roles in predicting the pupil involvement were overshadowed by the predictors of Model B.

The results of the last comparison in Table 6.26 have clearly revealed the relative strength of the dimension *Structured teaching skills (Recomputed)* over other predictors in their associations with the good individual involvement of the pupils. Thus, it was strongest in the predictions of both general indicators of lesson quality, in terms of the overall teaching quality and the overall individual involvement of pupils. The unique variance of this dimension shown in the first comparison in Table 6.26 was the highest, but its unique contribution reduced to 3.6%, when more variables were used in the prediction in the last comparison, which involved all the variables of the various models. It should be noted again that indicators of the dimension *Structured teaching skills (Recomputed)* are mainly indicators of the criteria *Teaching learning strategies* and *Effective classroom organisation* of the original theoretical model (in Table 5.12 in Section 5.6.1). This may also explain why the criterion/factor *Teaching learning strategies* did not appear in the solution of the final comparison. It would be redundant if they were to be included in the prediction. The new composite factor *Clear instruction* of Model C also seemed to be strongly associated with this indicator. These results suggest that enhancing pupils' individual participation may be contingent on teachers' abilities to deliver clear instruction, activate pupils, and facilitate with learning strategies.

6.4.3 Concluding notes on the multiple regression results

The comparisons made in this section were not intended to establish the superiority of either classroom observation instrument. On the contrary, from the results seen in Table 6.23 and Table 6.26, one can only conclude that different sets of predictors seem to be associated with different indicators of teaching effectiveness. Effective classroom management or organisation is a strong predictor of overall teaching quality for the current sample, much stronger than the results found by Hattie (2009, p.102) for its association with teaching effectiveness in his recent report of meta-analyses of educational effectiveness. Moreover, Hattie's (2009) finding seems to be consistent with the current results on predicting the student participation because those observed teachers' behaviours associated with the dimension *Structured teaching skills*, even when IND24 was excluded. These behaviours include those ensure knowledge transfer and interactive activities and those concern lesson structure and maximisation of learning time/opportunity in Carroll's (1963) learning model (see Section 2.4.1) and subsequently in Creemers' (1994) classroom effectiveness model (see Section 2.4.4). It is clear that the two global indicators of overall teaching effectiveness (r=.77, $ps.$<0.001) are associated, but the current study cannot establish how they may be related to academic results (i.e., the traditional indicators of educational effectiveness).

The relatively high correlations amongst the factors discussed in Section 6.2 (also see the complete correlation table in Appendix XV) raised some doubts about the validity as discussed in Chapters Four and Five. They also point to the likely existence of an underlying construct of overall generic teacher effectiveness that could well be closely related to the judgement of overall teaching quality. The present regression solutions generally indicated an R square high above .80, but individual factors only accounted for a small unique amount of variance in overall teaching quality. This again suggest strongly that most of the variance is shared among the factors and thus all can be seen as contributing to the overall teacher effectiveness. These ideas are further discussed in Chapter 7 in relation to the generic and differentiated theories of individual teacher effectiveness.

6.5 Relative validity of the models in the CVCP and ECP samples

In the previous sections, models have been compared using the same sample in the CVCP study. The data were based on the ratings of two classroom observation instruments collected on the same occasions in 76 lessons of four Hong Kong EFL teachers. The unit of comparison or analysis is based on the lesson and thus the subsequent model generation and model comparisons were all based on the lesson as the unit. In contrast, the data of the ECP study was based on 158 lessons of 79 teachers as the data of the two instruments were collected on different occasions. The teachers in that sample were selected purposively to represent more effective teachers in England. The unit of analysis in that study was the teacher. Thus, when the models of these two different studies are to be compared, we have to assume that the unit of analysis is comparable. It is easier to assume that each case in the ECP study represented a lesson of a teacher who had been chosen as likely to be as relatively effective in terms of measures of pupil outcomes. This follows that any conclusion to be drawn on the cross-validation results would be about the lessons. Cross-validation here also does not warrant any conclusion to be made about the relative effectiveness between the two samples. The teacher effectiveness of the ECP sample was confirmed independently and their sample selection strategy reflected that. The current sample selection strategy discussed in Section 3.5.2 did not treat the teachers in this sample as a direct comparison group of their English counterparts. However, it is still interesting to know how similar or different the dimensions of teaching of the two samples were.

The purpose of cross-validation is also not to make any conclusion about the lessons, because this can be easily done by comparing the means and the frequency distribution of the ratings as it has been done in Section 4.2 and Section 5.2. It is also more than comparing the underlying dimensions found in the data as this has been done in Section 4.7 and Section 5.7. Cross-validation analysis is intended “to select the ‘best’ model among a set of alternative models” (Diamantopoulos & Siguaw, 2000, p.129), because “the model that fits best in a given sample is not necessarily the model with the best cross-validity, especially when sample size is not large” (MacCallum et al., 1994, p.28).

As the sample sizes of both studies were not large, it seems that the purpose of cross-validation is justified. According to Dianmantopoulos and Siguaw (2000), when cross-validation analysis is done with samples of different populations, it can be either a case of validity extension or a case of validity generalisation, dependent on whether a single model or several models are used. As in Section 6.3.6, two ISTOF models and three QoT models can be used for cross-validation of ISTOF results using the multi-sample analysis in LISREL 8.72. Thus, the following sections present results of a case of validity generalisation performed on the ISTOF and QoT data of the ECP and the CVCP samples. In each case of validation, the model of interest was fitted onto the data of the ECP sample using a tight replication strategy in which all parameters were constrained as they were estimated originally for the CVCP sample (i.e., calibration sample).

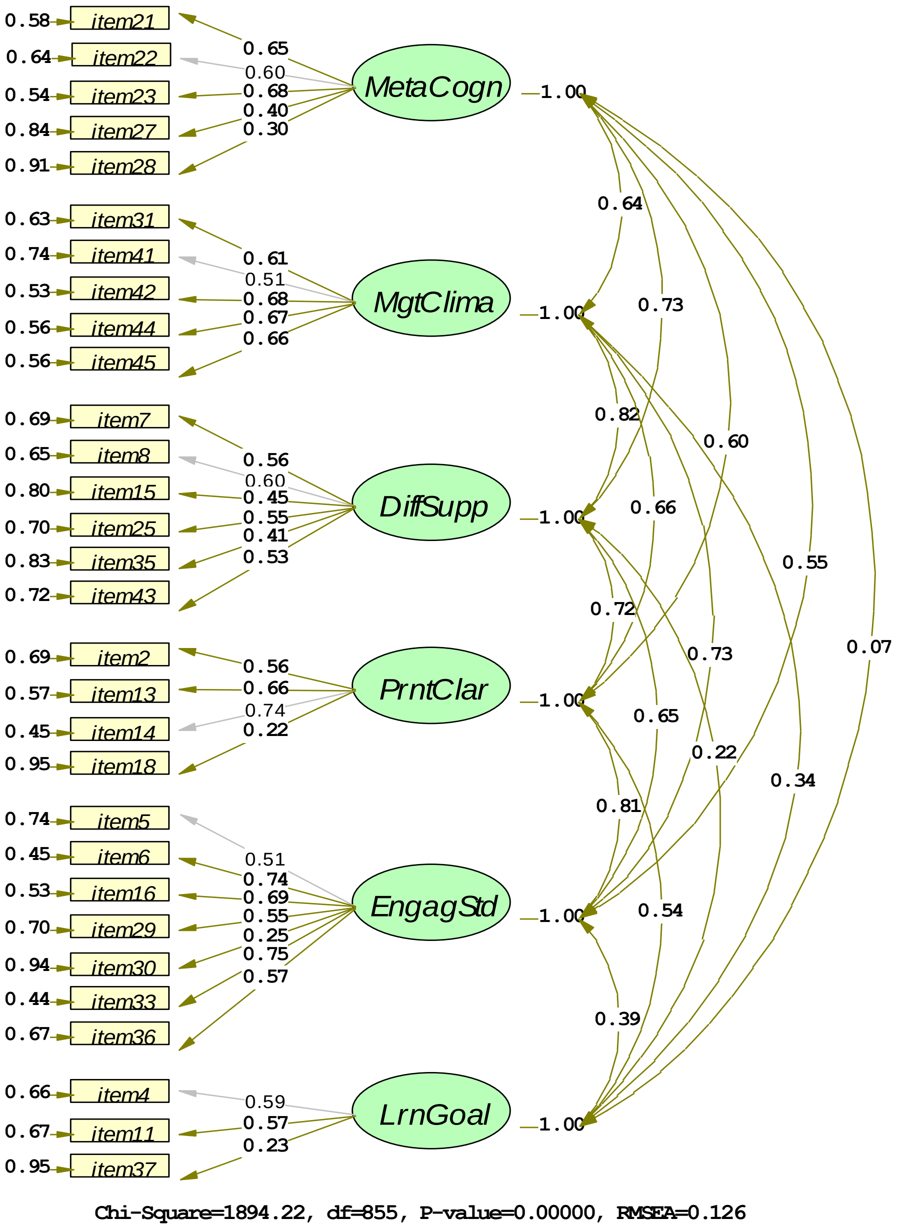
6.5.1 Validity generalisation of two ISTOF models in the ECP sample

The CVCP (ISTOF) Model

As seen in Figure 6.1, the solution of fitting the CVCP (ISTOF) Model to the ECP sample for validation does not produce encouraging results[83]

83 Because of the non-positive definite PHI-matrix, this solution was obtained when the ridge option value was reset to 0.6. There were other factors that might have affected the present estimation. First, the ECP data contained a considerable amount of imputed data. Second, most of the ECP data were significantly negatively skewed and platykurtic, which might affect estimation based on covariance matrix and maximum likelihood. In fact, in the ECP study, the analysis was based on asymptotic covariance matrix and diagonal weighted least squares.

Figure 6.1: **Cross-validation results using the CVCP (ISTOF) Model with standardised coefficients**



N=76 for the CVCP sample

N=79 for the ECP sample

<u>KEYS:</u>
MetaCogn: Meta-cognitive skills teaching

MgtClima: Classroom management and climate

DiffSupp: Differentiation and support

PrntClar: Clarity and logic of presentation

EngagStd: Student engagement

LrnGoal: Strategies to enhance learning and lesson focus.

Global goodness-of-fit evaluated using multiple fit indices suggested that the six-factor model did not fit the data (χ^2 =2188.75, df=855, p = 0.0; Population Discrepancy Function Value (F0) = 6.79, 90% CI for F0 = (5.99 ; 7.64); RMSEA=0.126 with 90%CI=(0.12; 0.13) and p-value for CFit= 0.0; NFI= 0.69; CFI=0.79; IFI=0.79; RFI=0.69; Critical N = 67.70). The group goodness-of-fit indices indicate that the CVCP sample is more influential than the ECP sample since it accounts for more than two-thirds of the overall model chi-square: Contribution to Chi-square of the CVCP sample = 1504.86, while that of the ECP sample = 683.89; Percentage contribution to chi-square of the CVCP sample = 68.75%, while that of the ECP sample = 31.25%; SRMR for the CVCP sample =0.21, while SRMR for the ECP sample =0.23.

An inspection of standardised residuals and modification indices revealed many localised points of ill-fit in the solution (e.g., the largest modification index= 55.30; largest fitted residual = -24.87, while largest standardised residual = -8.35). Standardised parameter estimates from this solution are presented in Figure 6.1. The majority of factor loading estimates ranged between 0.51 and 0.73, but Items 18, 28, 30, and 37 have loadings at 0.30 or below. These items also have the lowest squared multiple correlations at around 0.05 or 0.06. There are only about one-thirds of the squared multiple correlations are above 0.40, with some up to 0.56. All these results show that the CVCP (ISTOF) Model is ill-supported in the ECP sample. Because of the very different nature of the two samples, these results are perhaps unexpected. The ECP sample was found contribute only about 30% to the variances, the samples still shared a lot in common. The mean scores in the ECP sample are higher and the variance lower reflecting the attempt to select more effective teachers.

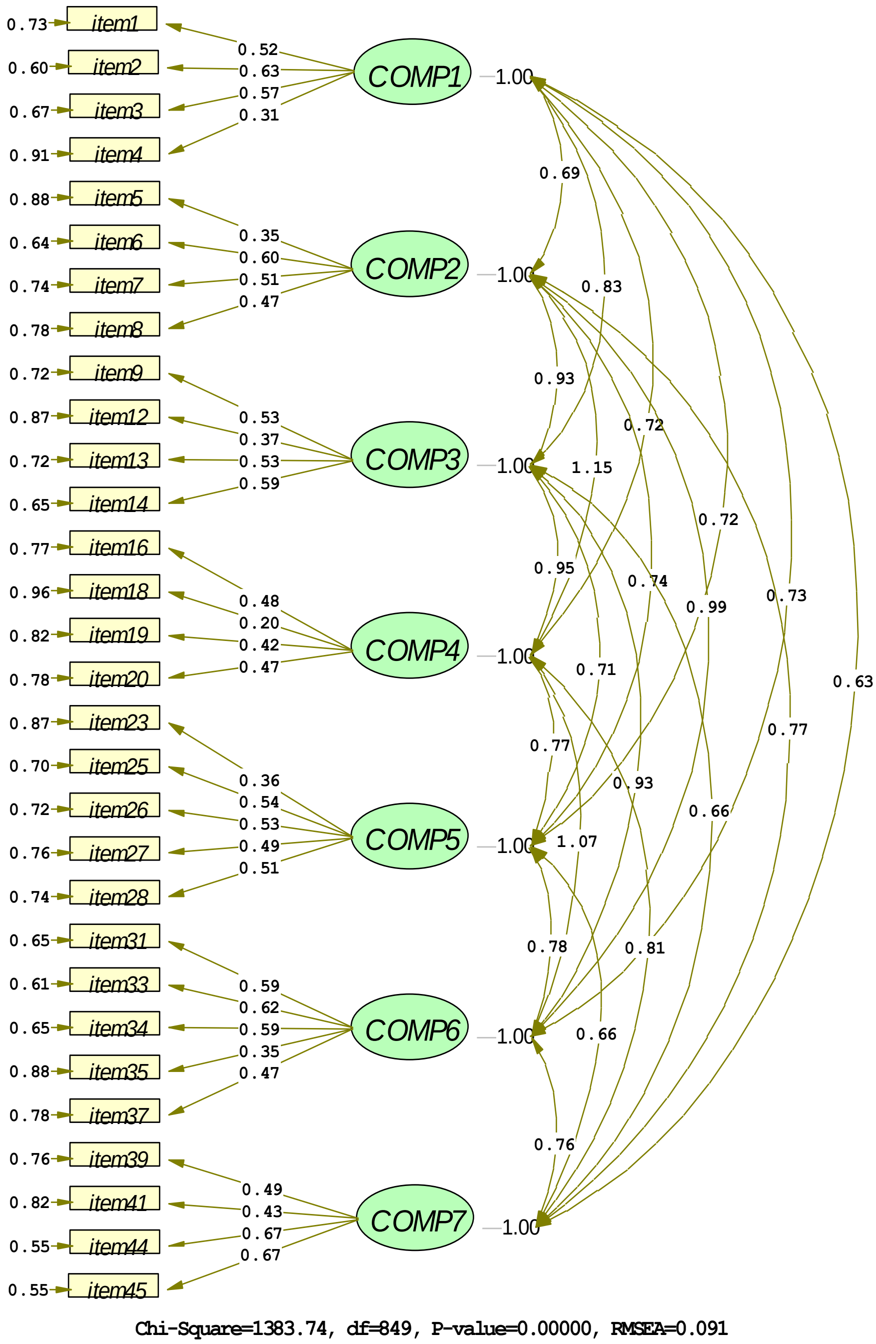
The theoretical ISTOF model

In contrast to the solution for the CVCP (ISTOF) Model, the solution of fitting the theoretical ISTOF model to the ECP sample for validation produces better results[84]. The results are plotted in Figure 6.2. The seven-factor model seemed fitting the data better than the CVCP (ISTOF) Model in the multiple fit indices showing the global goodness-of-fit (χ^2=1574.29, df=849, *p*=0.0; Population Discrepancy Function Value (F0) = 3.50, 90% CI for F0 = (2.85; 4.19); RMSEA=0.091 with 90%CI=(0.082; 0.099) and p-value for CFit= 0.0; NFI= 0.70; CFI=0.84; IFI=0.84; RFI=0.70; Critical N = 93.11). As in the CVCP (ISTOF) Model, the group goodness-of-fit indices also indicate that the CVCP sample is more influential than the ECP sample as it accounts for a bit less than two-thirds of the overall model chi-square: Contribution to Chi-square of the CVCP sample = 1006.13, while that of the ECP sample = 568.16; Percentage contribution to chi-square of the CVCP sample = 63.91%, while

84 This solution was obtained when the ridge option value was reset to 0.9 because the PHI-matrix was not positive definite.

that of the ECP sample = 36.09%; SRMR for the CVCP sample =0.18, while SRMR for the ECP sample =0.19.

Figure 6.2: **Cross-validation results using the theoretical ISTOF Model with standardised coefficients**



N=76 for the CVCP sample

N=79 for the ECP sample

KEYS:

Comp1: Assessment and Evaluation

Comp2: Differentiation and Inclusion

Comp3: Clarity of Instruction

Comp4: Instructional Skills

Comp5: Promoting active learning and developing meta-cognitive skills

Comp6: Classroom Climate

Comp7: Classroom Management

Standardised parameter estimates from this solution are presented in Figure 6.2. The majority of factor loading estimates ranged between 0.47 and 0.67, but Items 4 and 18 have loadings at 0.31 or below. Two-thirds of the squared multiple correlations are below 0.30, with the highest at 0.45 and the lowest at 0.04. An inspection of standardised residuals and modification indices indicated fewer localised points of ill fit in the solution (e.g., the largest modification index= 54.70 found in the ECP sample; largest fitted residual = -14.56, while largest standardised residual = -7.54). All these results show that the theoretical model is only poorly supported in the ECP sample. However, comparing to the CVCP (ISTOF) Model, the theoretical model still produces a rather better fitting solution. Its solution also shows an approximately 9% larger contribution of the ECP sample than it is in the solution of the CVCP (ISTOF) Model. Overall, these results suggest that the theoretical model has a better construct validity than the CVCP (ISTOF) Model.

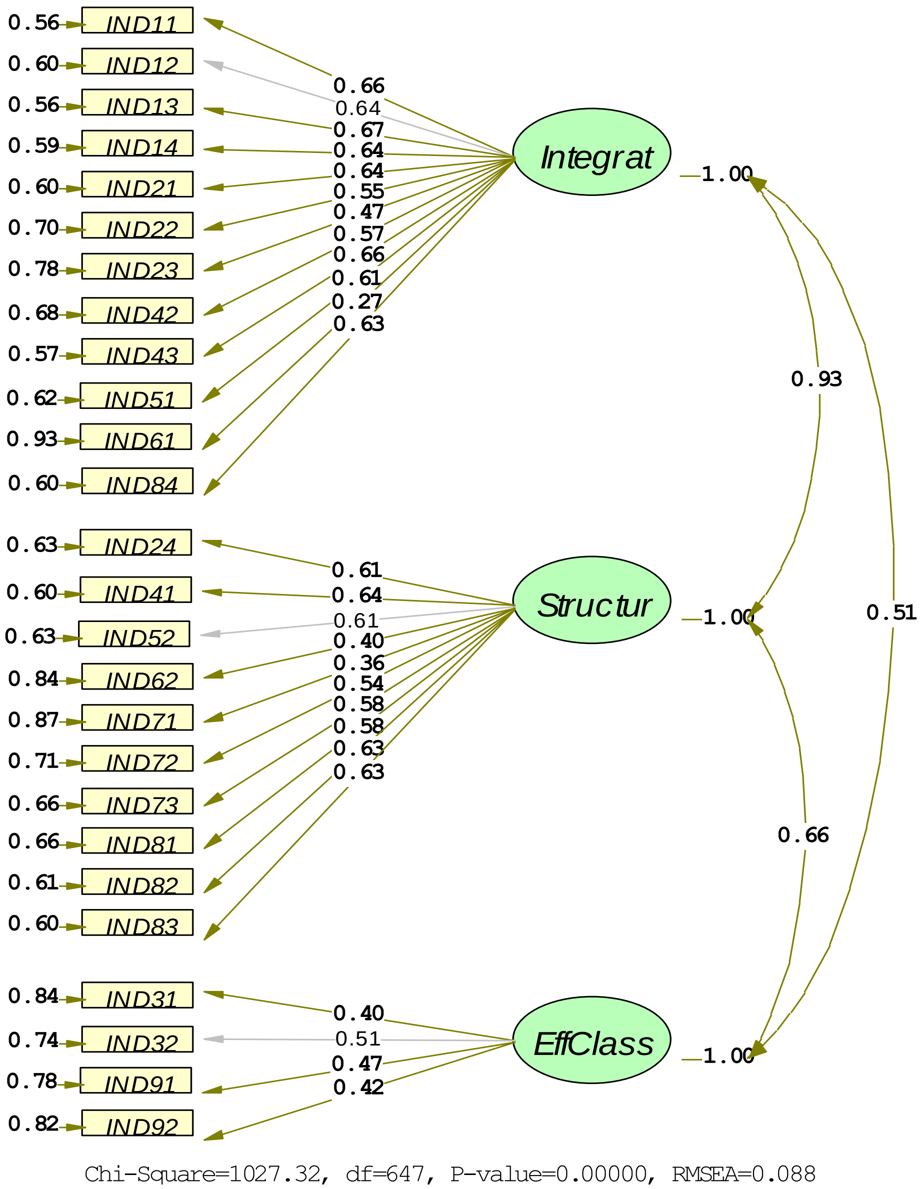
6.5.2 Validity generalisation of three QoT models in the ECP sample

The CVCP (QoT) Model

As shown in Figure 6.3, this three-factor CVCP (QoT) Model seemed to fit the ECP data better than both ISTOF models. The multiple fit indices indicate a slightly better the global goodness-of-fit[85] (χ^2=1574.29, df=647, *p*=0.0; Population Discrepancy Function Value (F0) =2.49, 90% CI for F0 = (1.94; 3.08); RMSEA=0.088 with 90%CI=(0.077; 0.098) and p-value for CFit= 0.0; NFI= 0.71; CFI=0.81; IFI=0.81; RFI=0.70; Critical N = 73.17). Compared to the ISTOF models, the group goodness-of-fit indices of this model show a stronger influence of the CVCP sample. This model accounts for over 70% of the overall model chi-square: Contribution to Chi-square of the CVCP sample = 1146.11, while that of the ECP sample = 409.08; percentage contribution to chi-square of the CVCP sample = 73.70%, while that of the ECP sample = 26.30%; SRMR for the CVCP sample =0.17, while SRMR for the ECP sample =0.19.

85 This solution was obtained when the ridge option value was reset to 1.0 because the PHI-matrix was not positive definite.

Figure 6.3: **Cross-validation results using the CVCP (QoT) Model with standardised coefficients**



N=76 for the CVCP sample

N=79 for the ECP sample

<u>KEYS:</u>

IntgrCls: Integrated Class Management & Climate

StrucTch: Structured Teaching Skills

EffClsPl: Effective Class/Lesson Planning

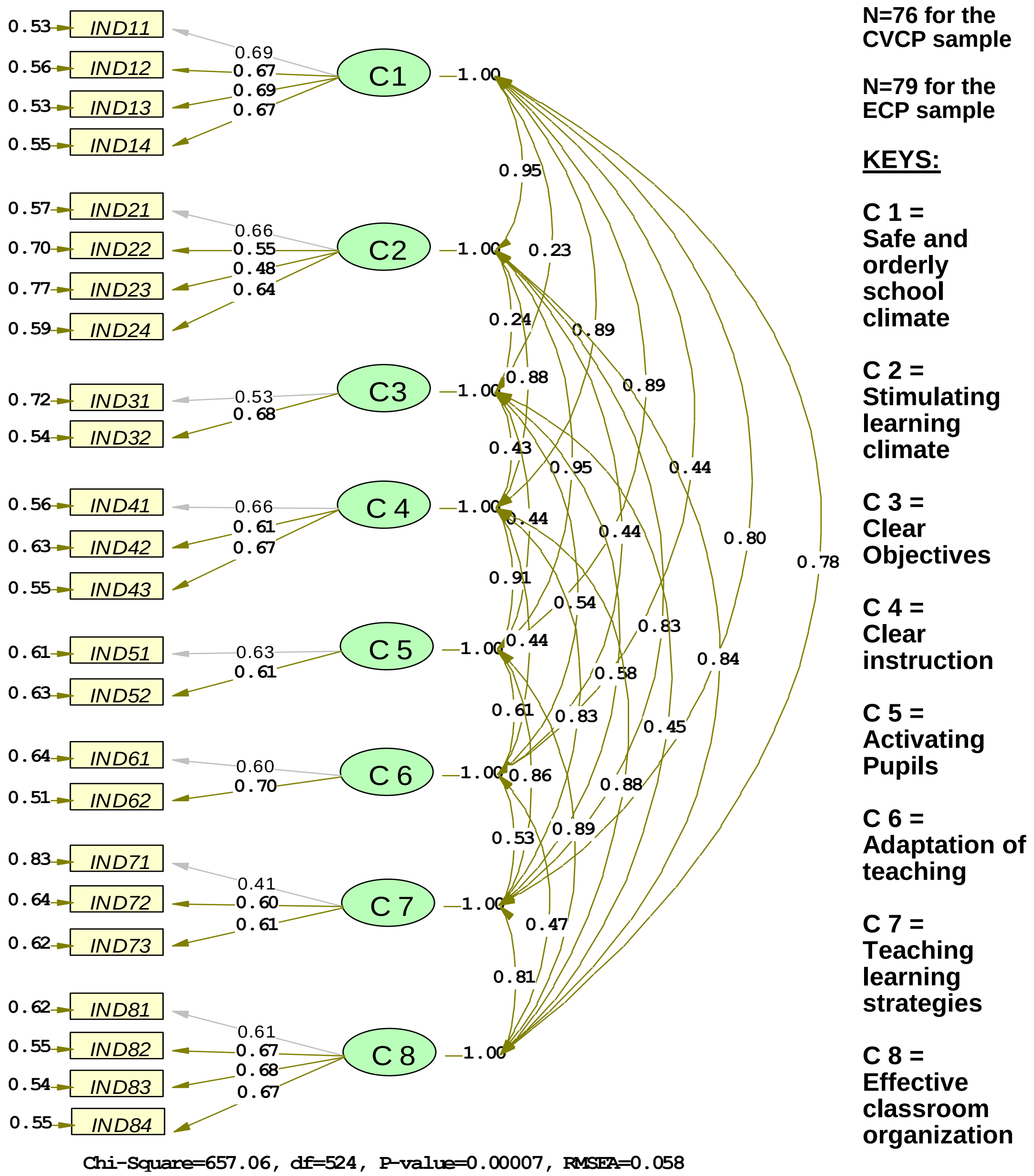
Standardised parameter estimates from this solution presented in Figure 6.3 show that the majority of factor loading estimates ranged between 0.47 and 0.67, but IND31, IND61, IND62 and IND71 have loadings at 0.40 or below. An inspection of standardised residuals and modification indices indicated a few localised points of ill fit in the solution (e.g., the largest modification index=65.30 in the ECP sample; largest fitted residual = -52.15, while largest standardised residual = -8.45). Half of the squared multiple correlations are below 0.30, with the lowest at 0.07 and 0.16 respectively found at IND61, and IND31 and IND62. All these results show that the CVCP

(QoT) Model is still poorly supported in the ECP sample, but its solution is slightly better than that found for both ISTOF models.

The original theoretical QoT Model (Model B)

Figure 6.4 shows the solution of fitting the eight-factor original theoretical QoT model (Model B) onto the ECP sample.

Figure 6.4: **Cross-validation results using the original theoretical QoT model (Model B) with standardised coefficients**



The resulted solution is superior to the solutions of the previous models discussed so far. Several multiple fit indices indicate a reasonable global goodness-of-fit within an acceptable range (χ^2=894.08, df=524, p=0.0;

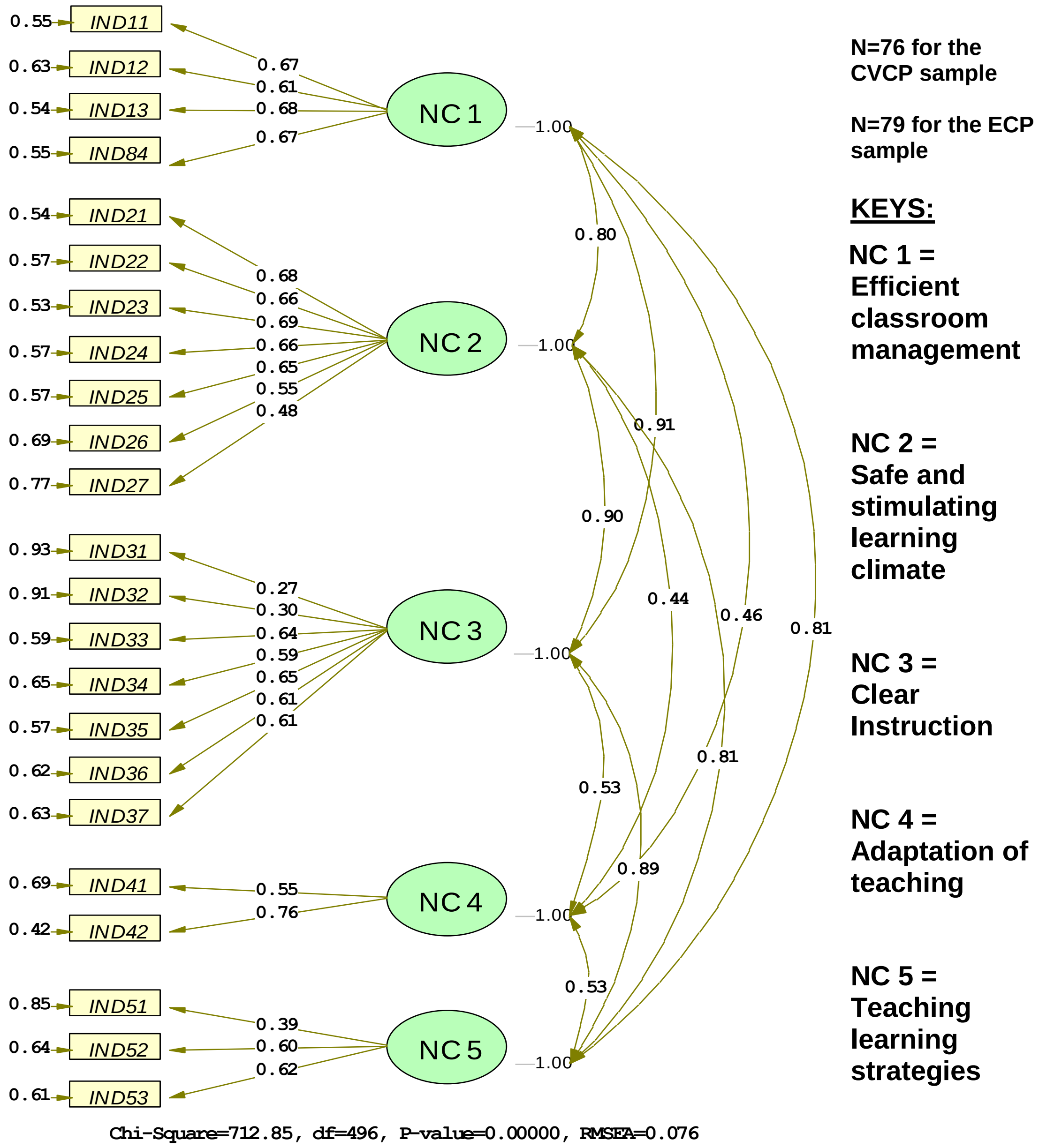
Population Discrepancy Function Value (F0) = 0.87, 90% CI for F0 = (0.47; 1.32); RMSEA=0.058 with 90%CI=(0.042; 0.071) and p-value for CFit= 0.19; NFI= 0.82; CFI=0.92; IFI=0.92; RFI=0.81; Critical N = 104.06). Compared to the CVCP (QoT) Model, the group goodness-of-fit indices of the original theoretical model show an increased influence of the ECP sample, but still falls short of that of the CVCP sample in its contribution to the overall model chi-square: Contribution to Chi-square of the CVCP sample = 612.14, while that of the ECP sample = 281.94; Percentage contribution to chi-square of the CVCP sample = 68.47%, while that of the ECP sample = 31.53%; SRMR for the CVCP sample =0.17, while SRMR for the ECP sample =0.17.

Standardised parameter estimates from this solution presented in Figure 6.4 show that except those of IND23 and IND71, all standardised factor loading estimates are all above 0.50. No statistically insignificant estimates have been identified. The loadings of IND23 and IND71 are 0.48 and 0.41, respectively, which are still within an acceptable and statistically significant range. Most of the squared multiple correlations are moderate, ranging between 0.30 and 0.49. The lowest one found at IND23 and IND71 are 0.23 and 0.17 respectively. Standardised residuals and modification indices have indicated a few localised points of ill fit in the solution (e.g., the largest modification index=57.50 in the ECP sample; largest fitted residual = -27.30, while largest standardised residual = -7.94). All these results show that except a few localised ill-fit estimates, the original theoretical QoT model is in general moderately supported in the ECP sample.

The new theoretical QoT Model (Model C)

Figure 6.5 shows the solution of the last QoT model onto the ECP sample. The resulted solution of this new theoretical model is not as good as that of the original theoretical one. Comparing to those of the original model, the multiple fit indices indicate an overall decline in the global goodness-of-fit (χ^2=917.21, df=496, p=0.0; Population Discrepancy Function Value (F0)=1.42, 90% CI for F0 = (0.98; 1.91); RMSEA=0.076 with 90%CI=(0.063; 0.088) and p-value for CFit= 0.00092; NFI= 0.80; CFI=0.90; IFI=0.90; RFI=0.80; Critical N = 96.45).

***Figure 6.5:* Cross-validation results using the new theoretical QoT model (Model C) with standardised coefficients**



However, the group goodness-of-fit indices of the new theoretical model are very similar to those of the original model, indicating a similar contribution of the ECP sample to the overall model chi-square: Contribution to Chi-square of the CVCP sample = 618.64, while that of the ECP sample = 298.57; Percentage contribution to chi-square of the CVCP sample = 67.45%, while that of the ECP sample = 32.55%; SRMR for the CVCP sample =0.17, while SRMR for the ECP sample =0.18.

Standardised parameter estimates from this solution presented in Figure 6.5 show that except four indicators (i.e., IND27, IND31, IND32, and

IND51), all standardised factor loading estimates are all above 0.50. The factor loadings of this new theoretical model are higher than those of the original model, as many are above 0.60 and one as high as 0.76. Again, no statistically insignificant estimates have been identified. Accordingly, except for a few indicators, most of the squared multiple correlations are moderate, ranging between 0.35 and 0.58. The lowest one found at IND31 and IND32 are 0.07 and 0.09 respectively. A few localised points of ill fit in the solution are identified in their standardised residuals and modification indices (e.g., the largest modification index=58.05 in the ECP sample; largest fitted residual = -27.30, while largest standardised residual = -7.94). All these results show that except a few localised ill-fit estimates, the new theoretical QoT model is also moderately supported but not as strong as the original model in the ECP sample.

6.6 Conclusion

In presenting their dynamic model of educational effectiveness, Creemers and Kyriakides (2008) explained how different factors may operate at different levels in school. They did present some empirical findings for those factors that may have influenced student achievement at the classroom level, but they studied those factors in isolation, rather than investigating their interactive impacts. Based on multiple lessons observed for a small number of EFL class teacher in a specific school in Hong Kong, the underlying dimensions identified in the confirmatory factor analyses discussed in the last two chapters provided a framework to explore the relationships among these dimensions together rather than in isolation. The bivariate correlations illustrate how different dimensions may be associated to one another in a model. The results clearly showed that while most dimensions were strongly correlated, the strengths of those correlations varied considerably across dimensions.

Importantly, correlations cannot identify the relative significance of these dimensions for teaching effectiveness. The two selected global indicators of overall teaching effectiveness in Section 6.3 of course may not be seen as substitutes for the actual academic performance of students as a measurement of teaching effectiveness, but they may be seen as important in their own right and as related indicators. The results presented show that the dimensions that may have relatively stronger impacts on the two indicators are different. It also turned out that dimensions of models that show similar goodness-of-fit in the data may vary in their strengths in their impacts on these indicators. The more parsimonious and/or updated models are not necessarily more predictive, because the original theoretical QoT model is found stronger in terms of predictability.

Regarding predictive validity, the results also tended to favour QoT more than ISTOF. Though both instruments may be high-inference by nature, they differ in an important aspect: ISTOF is based more on the frequency or occurrence of teachers' behaviours, while QoT is based more on the field researcher's judgement on the observed or inferred effectiveness of the observed teachers' behaviours. This probably would have enhanced the predictability of QoT as the rater attempt to evaluate the direct effect of the teachers' behaviours on students in the rating process. The dimension *Effective classroom organization* of QoT is not confined to monitoring students' behaviours or acting accordingly towards disruptive behaviours as the term classroom management traditionally implies. The indicators of this dimension seem to be referring to what Creemers (1994) termed as opportunity to learn. This is an important aspect that is not addressed in the ISTOF schedule, but has shown to be an important predictor associated strongly with both indicators of teaching effectiveness. Moreover, while teachers' behaviours specified in the dimension *Clear objectives* are more likely to affect teaching quality, those teachers' behaviours specified in the dimension *Teaching learning strategies* tend to predict students' individual involvement more directly.

Cross-validating different models in the ECP sample has provide evidence related to the external validity of different models. It may not be a surprise that the theoretical models of both instruments fitted better than the empirical models in the data for the ECP sample. However, this result holds nothing against the empirical models as they may just describe the CVCP sample better. The fact that the original theoretical QoT model also has stronger predictability than the empirical model may only suggest the instrument has strong external validity as it has been developed and tested in various countries (see van de Grift, 2007).

So far the analyses and results presented has been focused on *the lesson* as the unit of analysis. These results have laid the foundation for the next chapter to deal with in-depth case analyses, which involve *the teacher* as the unit of analysis.

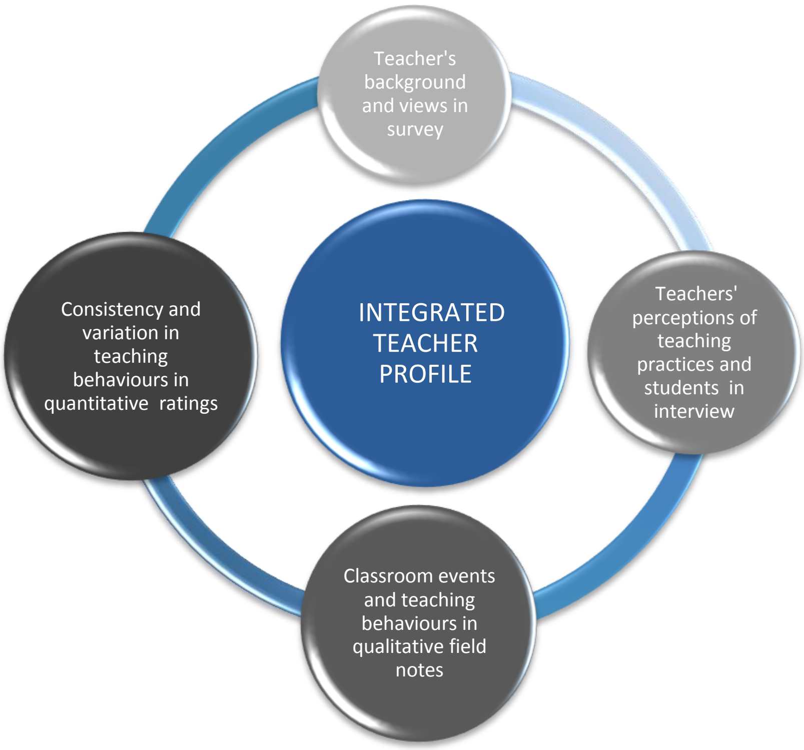
CHAPTER 7 : CONSISTENCY AND VARIATION OBSERVED IN THE FOUR EFL TEACHERS

7.1 Introduction

In the last three chapters, the analyses and results presented investigated the underlying dimensions of classroom practices identified in various empirical or theoretical models using ISTOF or QoT as the observation instrument with the unit of analysis focusing on *the lesson*. Due to this, the findings illustrate the similarities and differences between *observed practices in lessons,* rather than differences between *the four teachers* participating in the study. In other words, teacher and their teaching have been artificially separated in the quantitative analyses as the findings on the teaching in the lessons have drawn no reference to the teachers.

However, the focus of this chapter is on *the teacher*. Findings are thus presented to build up a case profile for each observed teacher. Figure 7.1 shows that multiple data are employed to enrich understanding of the teaching behaviours of the teachers.

Figure 7.1**: Different types of data as accumulative evidence for building the teacher profiles**



Note: The relative size of the circle indicates the relative amount of data available.

There were four types of evidence: 1) teachers' self-reported background and views in the returned teacher questionnaire; 2) their perceptions and perspectives on teaching practices, students, and other contextual influences presented in the interview; 3) excerpts of the classroom events in their lessons that characterise their teaching practices; and 4) the patterns in their scores of the underlying dimensions that indicate the consistency and variation in their teacher behaviours across lessons. The purpose of these case studies is not for teacher appraisal, but for providing a picture of the observed variation in teaching behaviours of the studied teachers.

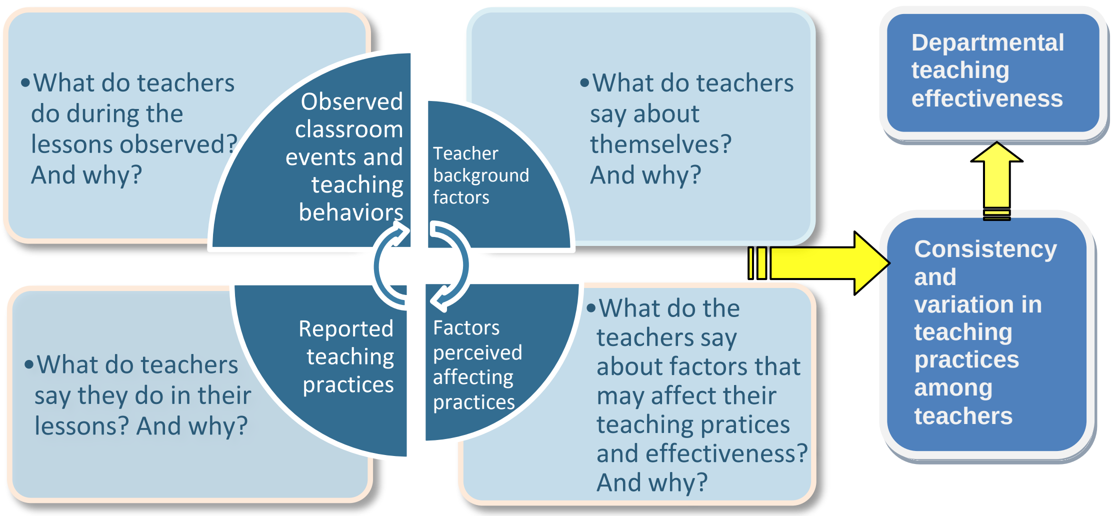
For building up each case profile, these diverse types of data that shed light on the perceived teacher profile are presented as the different steps of a thought-through process. In the first step, brief biographical information recorded their training and teaching background as well as the self-reported views collected in the teacher questionnaire (see Appendix III) on teaching, relationship with students, self-perceived effectiveness, and professional development in the school. The interview in the second step let the teachers freely express their perceptions on their teaching practices, the students and the learning atmosphere, the collegial support and professional development in the school.

For the third step, descriptive excerpts selected from the extensive account of classroom events of 15 to 23 lessons of a teacher. Excerpts are selected as snap shots of what the teachers did that might characterise some of some typical as well as distinctive features of their lessons. Excerpts tended to show more from lessons in which teaching behaviours were rated high, low or average. As mentioned in Section 3.5.4, details of the field notes were somewhat limited by their length of approximately 200 to 400 words.

In the final step, results presented in Chapters 4 and 5 are reorganised and presented on a teacher basis. The patterns in each teacher's scores in the underlying dimensions of observed teaching behaviours are presented in their ranks among all the lessons observed. Each score has been converted into one of the three levels, HIGH, AVERAGE or LOW. This treatment allows a view to look at variations broader than the actual variances. The main concern is to explore to what extent these patterns would support to the various theoretical models of teacher effectiveness discussed in Chapter 2. For example, would a teacher consistently score high in most underling dimensions most of the time as the generic model of teacher effectiveness may predict? Would a teacher score higher in some underlying dimensions more often than others and would his/her teaching effectiveness be differentiated in certain contexts, as the differentiated model of teacher effectiveness may predict?

It is expected that by the end of the end of this four-step thought-through process, a summary profile about each teacher and his/her teaching behaviours and practices and a set of factors that they perceived might affect their teaching practices and effectiveness positively and negatively could be identified. As depicted in Figure 7.2, the outputs of the case studies can further contribute to the understanding of consistency and variation in teaching practices among teachers in the department that can be associated to the departmental teaching effectiveness.

Figure 7.2: **The relationship between data of the case study and the cross-case analysis**



Note: The relative size of the quarter indicates the relative amount of data available.

7.2 The case of Charlie

7.2.1 Brief backgrounds and self-reported views

Charlie is in early thirties but already has ten years of teaching experience. He had worked in another school for two years before he started working in *Ming Tak Comprehensive*. He graduated from the Faculty of

Education of the University of Hong Kong and has a master degree in Education, specialised in English teaching. On average, he spent about 46-51 hours on his work. His motivation as a teacher and commitment to teaching was high but his job satisfaction was only moderate. Although he felt much stressed for his work as a teacher, he expressed no planning to change his profession in the next two years. However, he was less certain about whether his teaching role and teaching job would change in the next two years.

7.2.2 Perceptions on teaching and learning in school and factors affecting their teaching practices and teacher effectiveness

Charlie reported that he often introduced content through formal presentations in class, posed open-ended questions, and engaged the whole class in discussions. However, he promoted cooperative, investigative, or independent learning less often. In preparing his lesson, he would consider student's prior understanding and teach groups of heterogeneous abilities. He was very pleased with the management and climate in his classes and satisfied with his relationships with his students. The only thing that worried him was the little assistance he could give parents in helping their children to achieve in school. He was rather disappointed with the professional development in the school, for the lack of support and the lack of professional caliber in the faculty. As everyone was busy, he could not work with his colleagues to improve instructional strategies or conduct peer observations.

Charlie was interviewed twice, one for making comments as a teacher and one as the department head. Charlie was aware that his leadership might affect the teacher effectiveness and teaching practices of the department, because he was accountable for the performance of the whole department as the head as well as the performance of the senior forms as the only teacher of all four A-level classes. In addressing his role as the teacher in the interview, he said that this year he had two exceptional strategies to enhance student outcomes those senior classes. This reflected his anticipation of the examination requirements and his awareness of the relative strengths of the teachers as well as the needs of his students in facing the exams. First, though Charlie was experienced in teaching senior form classes, he indicated more effective in teaching speaking and writing skills. Therefore, he shared the teaching with another teacher in two classes observed this year, as each teacher could focus on teaching examination papers in which s/he was more specialised. Second, he also asked individual group of students to go to practise oral skills with the native speaker English teacher during the English lessons. More oral practices might help increase students' oral skills, but they might also find it difficult to catch up what they had left out in the lessons. While the effectiveness of these strategies was still uncertain, they reflected Charlie's open-mindedness and flexibility in teaching strategies.

When asked to evaluate his teaching performance in the observed lessons, Charlie expressed strong confidence in his teaching and thought that he had achieved his goal in teaching:

> Regarding my goals in teaching, my hope is that they would understand me and will be benefited in the exams…. I think I have achieved a large extent of my goal. … I like all my lessons, every lesson.
>
> Reflect after each lesson. I will change the styles and emphasis but not the notes or explanations.

In particular, he noted that his different approaches to junior and senior form students reflected his teaching practices were affected by the classroom climate and learning atmosphere and his strengths were in teaching senior forms:

> I think classroom climate and clarity of instruction are my strengths in senior forms, but classroom management for junior forms.
>
> Classroom climate refers to the learning atmosphere in the classroom. They [students] will ask you lots of questions if classroom climate is good… I am not saying that classroom climate is not important in junior forms. It is important but classroom management is more important to our [junior forms] students. Classroom management is more difficult in junior forms as it is difficult to maintain their attention.
>
> Too much teacher-led instruction and strict classroom management may perhaps make [junior form] students more passive, but that is a must to command their attention for our students.

Relatively speaking, Charlie enjoyed more when he could see more interactions in the classroom and he explained this was why he preferred teaching speaking and writing in the senior classes. His comments reflected the reciprocal influences between the teacher and the students in class:

> There are more interactions when I teach speaking and writing. Students may still be passive when I do the talking but students would have a lot of peer interaction after I teach them sentences patterns. I treasure this kind of interaction. I would like to them to show joint effort in developing ideas.

Sharing similar views with the school principal (see Section 8.2), Charlie argued for his differentiated approach to students of junior and senior forms:

> I like to see [senior form] students develop their ideas. It is one of their weakest areas. I like to see them raise different questions. But for junior forms I would try to use simple questions and not much high order thinking for junior forms because I need to make sure that they can get the answers; for example, some factual information in the reading comprehension. In senior forms, students have a higher expectation for me, so they would raise questions rather than just sit there. It is seldom that junior form students would ask me questions because they don't know how to ask.

As the head of the English department, Charlie did not perceive that there was much variation in the teaching practices among the English teachers and assumed that students would learn similarly as long as they were attentive:

> We believe that if we can control the class, the chalk and talk approach is fine. No matter what teaching styles you are using, they will be learning because Hong Kong students are passive. They will just sit down and listen to you.

However, Charlie did notice the differential effectiveness of teachers in relation to student outcomes and made a contradictory comment:

> There is no doubt the teachers' effect on students' outcomes. I can see some teachers can make great improvements even they are assigned to teach classes with academically less able students.

Therefore, he criticised the teacher-centred approach that attempts to minimise the amount of interactive activities in the classroom and the non-discriminatory approach to unruly behaviours that had minimised the learning activities of disruptive students:

Sometimes teachers would just stick to their own teaching style no matter what changes arise in the classroom contexts. Some teachers still think that interactive activities are irrelevant to learning English and prefer only direct instruction.

I guess there may be only one or two who are really disruptive all the time. So I have to ignore them in order to do interactive stuffs, even when the students are weak in English. They will engage in the activities if you ask them to. You see, many disruptive students are actually very brilliant as they can keep on answering my questions. I still have to pay attention to them. I think teaching practices have to vary with the circumstances.

When Charlie referred to the teaching practices of "some teachers," he tried to distinct himself from them and did not address how his role as the department head might affect them.

While acknowledging that external factors may affect both teaching practices and teacher effectiveness, Charlie stressed more on students' psychological well-being and the negative peer group pressures in the school that form an anti-academic ethos:

Students need stronger determination. Here the learning atmosphere is very low. When I asked the senior form students why they didn't go to the study room opened here after school, they said they didn't want to study in school because they couldn't feel the learning atmosphere. It is very serious in the junior forms. High achievers are often embarrassed by their good results because they will be teased by other low achievers. Sooner or later, these achievers will give up the hard work. It happens not just in English but in all subjects. With all these peers, you need strong determination.

Charlie showed his empathy for the students' difficulties and needs for accomplishment in learning English as second language learners:

I can understand their problems and difficulties as I am also a Chinese learner of English. I can see what kind of problems that Chinese ESL learners peculiarly have. We share similar learning experiences.

I think it is obvious that they can get the immediate satisfaction [in learning other subjects] and sense of achievement. It is not easy for them to perceive their accomplishments in learning English. In other subjects, they know how to express themselves in L1 but not in English.

He was particularly concerned with students' low self-concept when they were admitted to the school, but stressed that the role of effective teachers in helping students to overcome the psychological and the actual learning processes that students had to overcome in a Band 3 school:

> The self-concept of the students here is very, very low. It is understandable as it is a failure to be in our school. It is somewhat classified as a school for low achievers or losers in the area. Yes, but some of them may meet some teachers who can revive them such that they can go through all the processes and eventually get into the university.

7.2.3 Classroom events that characterise teaching practices

Charlie's attitudes and perceptions reflected his open-mindedness and optimism as a teacher. He expressed his confidence had been built on his perception of his students' outcomes and the very interactions and deep learning he had treasured in his classrooms. Table 7.1 summarises the relationship between the topics of the oral presentations, a distinctive section of all his lessons and the lesson focus.

Table 7.1: **Topics of oral presentations and lesson focus in Charlie's observed lessons**

Date	Class & Lesson No.	Topic of Oral Presentation	Lesson Focus
18/09/08	Form 6 Arts (Lesson 1)	**Diet and health**	**Group discussion: Eating habit and food**
18/09/08	Form 7 Arts (Lesson 3)	Mountaineering safety	Listening comprehension: Life-time employment
19/09/08	Form 6 Arts (Lesson 5)	Rights of the patient	Writing practice: The governmental responsibility on early childhood education
22/09/08	Form 6 Science (Lesson 7)	Rights of the patient	Writing practice: Smoking and cancer
22/09/08	Form 7 Arts (Lesson 9)	Mountaineering safety	Composition --Movies
22/09/08	Form 7 Science (Lesson 11)	Mountaineering safety	Composition --Movies
23/09/08	Form 7 Science (Lesson 13)	Mountaineering safety	Listening comprehension: Life-time employment
23/09/08	Form 6 Science (Lesson 15	Diet and health	Writing summary: Banning tobacco advertisement

Each oral presentation involved two students presenting a short speech on a selected topic, after which Charlie would ask other students to make comments on the presenters' performances. This routine looked more like a distinctive, rather than an integrated, part of a lesson[86], because, as shown in Table 7.1, except in Lesson 1, the topics were not related to the lesson focus.

86 This actually also reflected the washback effect of the public examinations on the English curriculum of senior and upper senior forms in Hong Kong. The lesson focus usually concentrates on practicing one of the language skills (i.e., writing, reading and grammar, listening, and speaking), rather than integrating different skills on the same theme. It is also a common practice that teachers use different textbooks for practising different skills, which do not have always have common topics and themes. There are always pros and cons about the extent to which teachers have to tailor materials for the curriculum.

However, according to Charlie, asking students to give oral presentation at the beginning of the lesson was his new teaching strategy this year.

In perceiving the diverse English proficiencies in his students, Charlie employed strategies to enhance learning and participation in the oral presentation, but the range of strategies seemed limited to taking notes for students who could not follow the presentation, naming students he wished to contribute, sought class consensus (shown in italics in the following excerpts):

Lessons 1, Form 6, Arts

8:51 ….One student comes out to give oral presentation while other students are watching. The teacher puts a timer on the desk and lets her start. While the student gives her presentation, *the teacher writes down points/things that she needs to pay attention to her presentation* (e.g., grammatical errors like **She think; *improve our healthy*) on the blackboard. *The teacher asks other students to comment on the student's performance*, but without any response. The teacher comments that the student didn't have enough eye contacts. Then another student makes his comment that maybe too many notes was taken and used in the presentation.

Lessons 13, Form 7, Science

8:48 A student starts his oral presentation, while The teacher stands on the left watching. *Again he writes down words mispronounced on the blackboard*: compass, crisis, risk, extra.. .. and grammatical errors: There *is some negative errors. Hong Kong's *economic is uncertain. *The teacher comments that the student read too much from his note card*. The teacher added that he was fluent but needed to pay attention eye contact and mispronounced words.

8:55 Another student does her presentation. *The teacher writes errors on the blackboard*: should *concluding, *will confused. A student asks what the difference between fuzzy & fussy is. The teacher continues explanations on some vocabulary.

Observing the low responses in students, Charlie also tried diverse strategies to promote peer interaction/feedback:

Lessons 1, Form 6, Arts

9:02 Another student comes out to give her oral presentation. The teacher writes on the blackboard again the grammatical errors and comments on the problem of reading too much from notes and the pace of giving presentation. The teacher engages other students with the points he listed on the blackboard. A few students get into details and ask the usage of some idiomatic expressions. 9:09 Oral presentation session ends.

Lessons 7, Form 6, Science

8:46 Class starts with a student's presentation*. The teacher observes the student by standing closely on her right*. He writes the topic and notes on the blackboard.

8:55 The teacher asks students to give comment on the student's performance. No student responds. The teacher has to change his question and asks about whether the introduction is appropriate or not. The teacher says the student can make use of different sentence structure, paraphrasing HK people disappointed → to their disappointment. Then the teacher comments that the oral presentation was very nice, clear and well organised. The teacher tries to provide evidence as justifications for his praises to provide a model of giving positive feedbacks.

9:00 Oral presentation session ends.

Lessons 9, Form 7, Arts

11:46 Lesson starts with a student's oral presentation. The teacher takes notes on the blackboard. The teacher signals the student to look at the class when she presents.

11:50 The teacher asks students to give comments, but without any responses from students. The teacher says comments needn't be negative. A student says she was fluent. The teacher says she was also quite clear, but too many words in simplified forms. Another student says there were a variety of sentence patterns used. The teacher asks him to give examples. The student says examples like those using which. The teacher clarifies his comment by referring to mean the usage of relative clauses.

12:04 Oral presentation session ends.

From the students' responses observed (underlined in the above excerpts), oral presentation seemed to be a routine that attracted few student responses. Other than the two presenters, many students participated very little in comparing to what they did in the other learning activities. Lower response rate might be expected when the presentations became very difficult to follow because of poor organisation, mispronunciations, grammatical and structural errors. Even an experienced oral examiner like Charlie had to concentrate on listening to the presenters and did not observe the class as often as he did at other times.

As Charlie always had a double-period lesson for his classes, he rarely stated the lesson theme and the skills in focus on the blackboard at the beginning of the lesson. Instead, he would write the lesson focus on the blackboard in the middle of the first lesson after the oral presentation routine. For example, in Lesson 9, he only wrote the lesson focus on the blackboard 15 minutes before the end of the lesson. As the students had a reading passage about movies from which they could know useful vocabulary in their contexts, Charlie could then engage his students in deep learning like posing open-ended questions and whole class discussions:

> 12: 04 The teacher writes Composition-Movies on the blackboard…
>
> 12:14 The teacher asks what the ideas of the passages are: e.g., Why did the writer distinguish different types of movies like documentary, thrillers, romances, etc? What sort of vocabulary you may need for writing a movie a documentary movie? The teacher discusses with the students how the writer discusses the movies and tells students that they need not mention a real movie in the exam as it is fine to fabricate a movie. The teacher goes on to talk about a recent movie about the local community and writes down the vocabulary related to the issues discussed in that movie.

Continuous questioning by students indicated their high motivation to learn. Charlie had to stop answering their questions before it was too late for them to work on their writing task:

> **12:20** The bell rings. The teacher tells students that he would let them write the first paragraph and check over it and give them immediate feedback. But he wants to make sure that they understand the passage and its vocabulary. Some students ask him words that they don't know: e.g., *fuel, harassment, documentary, collaborate, massacre, vowed, phenomenal,* etc. The teacher explains about 10 words.
>
> 12: 33 *The teacher wants to stop explaining words for students, but students keep on asking him*. The teacher explains words like *trafficking,* while distributing blank paper for them to write composition of 500 words. He reminds them to write the first two paragraphs first and let him give them comments.

Charlie did not sit in his own desk to wait for his students to finish their writing task, but walked around in the classroom to see whether any individual help was needed. As Charlie did not use microphone, he often moved around in the classroom, promoting a closer proximity to his students:

> 12:39 *The teacher moves around in the classroom to check students' work and help them individually*. Some students would seek help but most students work on their own. They whisper and discuss with students next to or around them.
>
> 12:49 The teacher has moved around once and starts the second round to help individual students.
>
> **12:55** Before the bell rings, the teacher has helped 11 students since 12:33, about 2 minutes per student.

As Charlie showed his care in his students, they were motivated to seek for help.

7.2.4 Patterns of consistency and variation in teaching practices

Due to an unexpected holiday for a typhoon, fewer lessons of Charlie were observed. Yet, some patterns are still clear in Table 7.2, which shows consistency and variation of Charlie's teaching behaviours across the fifteen lessons observed. Except for *Effective class/lesson planning*, Charlie's average score in each underlying dimension as measured in the ISTOF and QoT CFA models was above the sample average, suggesting Charlie's teaching behaviours in most lessons tended to above average. His strengths tended to conform to the prediction of a generic theory of teacher effectiveness as he showed strengths in many were many underlying dimensions. His performance in the more global dimension, *Integrated class management and climate,* was even better than that in the more specific dimension, *Classroom management and climate*. This probably reflected that Charlie's strength was integrating classroom management and climate with other teaching practices as he was also strong in another global dimension, *Structured teaching skills*. In more specific underlying dimensions, Charlie was often strong in developing meta-cognitive skills (*Meta-cognitive skills teaching*), differentiating support to cater for the needs of students (*Differentiation and Support*), and delivering a clear and logical presentation (*Clarity and logic of presentation*).

***Table 7.2:* Consistency and variation in Charlie's observed teaching behaviours across lessons and their relations to indicators of teaching effectiveness**

Lesson No.	Meta-cognitive skills teaching (ISTOF1)	Class-room manage ment & climate (ISTOF2)	Different -iation & support (ISTOF3)	Clarity & logic of presenta tion (ISTOF4)	Student engage-ment (ISTOF5)	Strateg-ies to enhance learning and lesson focus (ISTOF6)	Integra-ted class manage ment & climate (QoT1)	Structur-ed teaching skills (QoT2)	Effective class/ lesson planning (QoT3)	Overall judgement of teaching quality (IND100)	Good individual involve-ment of the pupils (IND24)
Sample Mean	2.16	2.05	1.80	1.97	1.96	1.71	1.99	1.91	1.55	2.93	2.92
Sample S.D.	0.45	0.39	0.46	0.58	0.49	0.46	0.49	0.49	0.35	0.74	0.91
Charlie's Mean	2.50	2.00	2.03	2.45	2.33	1.97	2.48	2.24	1.39	3.13	3.60
1	AVG	HIGH	HIGH	AVG	LOW	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
2	AVG	HIGH	HIGH	HIGH	LOW	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
3	AVG	HIGH	AVG	HIGH	LOW	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
4	HIGH	HIGH	HIGH	HIGH	LOW	HIGH	HIGH	HIGH	LOW	M STRONG	STRONG
5	HIGH	HIGH	HIGH	HIGH	AVG	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
6	HIGH	HIGH	HIGH	HIGH	AVG	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
7	HIGH	HIGH	AVG	AVG	LOW	AVG	HIGH	HIGH	LOW	M STRONG	M STRONG
8	HIGH	HIGH	HIGH	LOW	LOW	AVG	HIGH	HIGH	LOW	M STRONG	STRONG
9	HIGH	HIGH	AVG	AVG	AVG	LOW	AVG	AVG	LOW	M STRONG	M WEAK
10	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	M STRONG	STRONG
11	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
12	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
13	AVG	AVG	AVG	AVG	LOW	AVG	HIGH	LOW	AVG	M STRONG	M STRONG
14	LOW	AVG	LOW	AVG	AVG	AVG	HIGH	AVG	AVG	M STRONG	M STRONG
15	HIGH	AVG	AVG	AVG	HIGH	AVG	HIGH	AVG	AVG	M STRONG	M STRONG

Note: For tables hereafter, AVG (=average, the middle 33%), LOW (=the lowest 33%) and HIGH (=the highest 33%) are defined in terms of the percentile of the corresponding factor score; STRONG=predominately strong; WEAK= predominately strong; M STRONG= more strengths than weaknesses; M WEAK=more weaknesses than strengths.

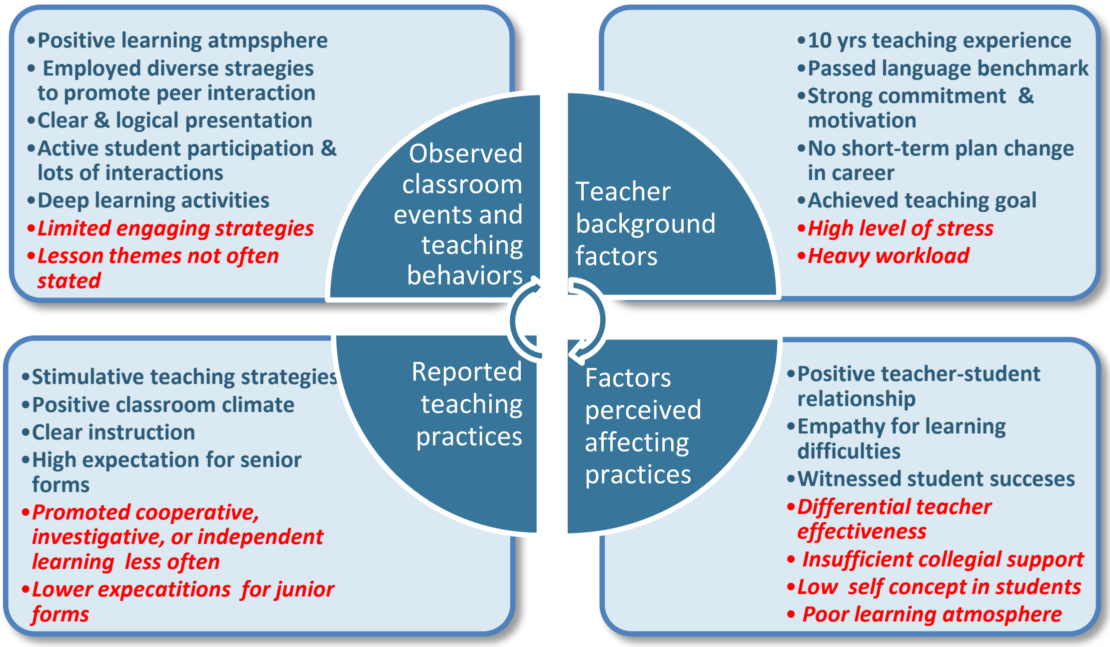
However, there was evidence that indicated Charlie did not receive high ratings consistently in all aspects as a generic theory of teacher effectiveness would predict. In almost half of the lessons, Charles was rated low in teaching behaviours related to *Student engagement,* though Charles' average score in this dimension was above average (2.33>1.96) and this seemed to affect the overall individual involvement of the students little. Charles also often scored low in *Effective class/lesson planning,* most of which were found in the first period of a double-period lesson, suggesting that this might be related to his practice to state the lesson focus late in the middle of the lesson. As Charlie has low scores on this dimension, this may have affected the number of lessons in which he received the highest rating for the overall teaching quality.

Regarding the stability of teaching behaviours, Charlie's performance seemed to be rather consistent. Lessons 10, 11 and 12 were exceptional lessons in which Charlie scored high in the scores of every underlying dimension. Charlie had more average scores in three lessons: Lessons 9, 13, and 14. Lesson 9 was the only lesson in which the overall individual involvement by the students tended to be lower than Charlie's average lessons, although his teaching behaviours related to *Student engagement* were not particularly low in that lesson. Although Charlie scored average in most of the underlying dimensions in Lessons 13 and 14, this only had moderate effects on both indicators of overall teaching effectiveness. This may suggest that it is easier to maintain overall teaching quality and student involvement as usual, when students are generally motivated and when the teacher can consistently maintain the quality of his teaching most of the time.

7.2.5 Case summary

Figure 7.3 summaries the findings in Charlie's profile and organizes them as factors that may positively or negatively affect the departmental teaching effectiveness.

Figure 7.3: **Summary findings that characterise Charlie's profile**



7.3 The case of Lucy

7.3.1 Brief backgrounds [87]

Lucy was an experienced teacher in early thirty. She taught the same school with Linus before. It was a school with similar class composition and socioeconomic background as the Ming Tak Comprehensive. She taught four

87 This section is short for Lucy as she did not return her teacher survey.

classes at different forms and of different academic abilities and English proficiency.

7.3.2 Perceptions on teaching and learning in school and factors affecting their teaching practices and teacher effectiveness

Lucy commented that the Form 1 class had given her a hard time before the streaming. It had a large class size with lots of disciplinary problems. It was difficult for her as she regarded *classroom management* and *catering for individual needs* were areas in which she felt less confident. She frankly doubted the practicality of inclusion policy introduced by the government and regarded it would just create more difficulties for Band 3 schools. On the one hand, the teacher might overlook the need of the more capable students:

> … it is really hard to pay attention to the lower [ability] ones and weaker ones in the class and you will neglect the strong ones. So for example, even in the class I just have 2DE with me on the third and fourth lessons ...I still have ... we still got some weaker ones, for example, except Peter who sits at the corner right? Hmm ...even when I was teaching comparatives, which is some very simple concept in the primary school, I can see Peter knows how to do it. But for those other 90 percent [students], they had to practise how to write the sentences. For the weaker ones, it is really hard...I had to ask their neighbours to help them to finish their exercises. Otherwise, I would leave all other students behind, the bright ones and those who had already picked up what they learnt ...

On the other hand, she felt the inclusion policy would not work, because the students were perhaps just too young to be considerate for others:

> … I think the naughty ones or the naughty classes, even when they are more capable, I don't think, will help the less capable one. I think they may not be naughty, but for example, Form 1 or Form 2 students, they are only just twelve or thirteen alright. I don't think they have such attitude to help others.

Lucy thought her strengths were "*clarity of instructions and questioning and presentations*" and she reflected on her presentation skills and questioning skills, mentioning how she improved presenting materials on the blackboard earlier as an example.

Lucy found collegiality affected her teaching effectiveness and teaching practices, despite minor disagreements with her colleagues:

> Because we work together to work out the worksheets or to work out the practice on how to teach. So that is why it is so good to compromise with the general practices in our panel.

She preferred harmony to raising criticisms on others on trivial matters and she was pleased with the kind of flexibility to maintain the ownership that she had over her teaching practices without voicing out the differences:

> I think if you have disagreements in certain ways with the others... I think for the teachers you adjust on your own.... simply you know we won't disagree on the overall themes or the curriculum or the planning of...of the scheme of work that sort of things ... the macro ones but the micro one is how you carry out your teaching, how you put your message across... that I think is not fixed... if you do not disagree, you have to use your own ways. So it is not ... I think you are just on your own. You don't need to shout out or voice out, "Oh, I don't think so .... such kind of teaching is rubbish... no need
>
> … I think we got much flexibility in the panel within the form meeting we have the flexibilities for the teachers to carry out ......how to carry out the lessons. I think that is important.

Regarding enhancing student outcomes, Lucy thought that the new curriculum requirements have resulted in changes in teaching practices:

> I think it is the change of [exam curriculum] requirements. For example, maybe in the old days we were not so emphasized on speaking or whatever, then you won't put so much effort in speaking, but maybe in reading, grammar ... grammar exercises, that sort of things. You see in the new curriculum or the new requirements for the examinations such kind of backwash [i.e., washback] effect is quite serious.

Although Lucy regarded that a good teacher-pupil relationship affected her teaching practices, she found it more difficult for her in junior forms as she was also anxious to maintain an authoritative figure image:

> Yes, teacher-student relationship [affected teaching practices most]. For example, for upper forms you use a more friendly approach with them, but for lower forms, you have an authority role. So this is different…

She acknowledged her different approaches to senior and junior forms resulted in a better teacher-student relationship with the senior forms:

> [I have better relationships] with the upper forms. You have to be friendly to get along with the upper forms, but for lower forms, especially Form 1, you need to maintain a more authoritative image to keep the order.

Her belief that an authoritative image is the key to successful classroom management might have prevented her from developing a relaxed classroom climate for learning, especially in the junior forms. However, Lucy explained that her approach was really dependent on class composition and the student reactions. Lucy declared that teaching practices are results of group dynamics between the teacher and the students:

> I think the 1C students are really good, even for the 1D I was very friendly at the beginning ….but later I found I couldn't use that friendly way anymore… Yes, the new mixed class was quite different after the streaming. The student composition changed. However, if you come back to the 2DE class, you can see that I was also very friendly to them… You see because XXX [a student's name] was not there. … teaching is affected by the interaction, the group dynamics. The teacher is also a member of the group though she has a special role comparing with the other members. So I strongly believe the group dynamics.

That may explain why she preferred less able but behaved students to more able but disruptive ones:

> Of course, you would choose [teaching] the low ability ones. Of course, I thought most teachers would prefer low ability ones [to disruptive ones].

Lucy thought that catering the individual differences in ability would be difficult but it would be easier to handle than unruly students. She rejected the idea to take those disruptive students as remedial groups, as it would be unfair to the teacher who would teach these groups.

When asked about the extent of impacts of external policies on her teaching practices, Lucy cited the medium of instruction policy as an example:

> We understand the intention of the MOI policy is good as students need inputs, but in practice, we can't do it in every class. There are problems like foundation of the students and short attention span. Both prevent them [the students] to listen to English for the whole lesson.

Lucy argued against the MOI policy for it imposed a one-size-fit-all pedagogy for students.

7.3.3 Classroom events that characterise teaching practices

Lucy used the microphone extensively and taught in mixed code mainly in L1 with L2 terms most of the time. As shown in her writing and selected colour chalk, she organized materials on the blackboard with care. In a double-period lesson (Lessons 3 and 4) planned for a Form 5 (equivalent to Year 11 in England) Science class, Lucy gradually failed to engage her students as she used activities not clearly relevant to develop the target skills:

> …. Lucy analyses the first paragraph by showing them the key topic sentence and the clause that contains the main point. Lucy asks them to underline the main point. Lucy points out that the topic sentence is usually the first sentence of a paragraph. Lucy says that the topic sentence is supposed to summarise the points/content of the paragraph.
>
> At 11:03, Lucy asks students to analyse the second paragraph. Lucy explains in L1: Unlike in *writing* (L2), for *listening* (L2), they only have to write simple sentences.
>
> Lucy writes on the blackboard: The purpose of Health Awareness Week is to raise students' awareness on healthy living.
>
> …..At 11:14, Lucy speaks with a microphone in mixed code. Her instruction is basically in L1 with some key terms in L2 (email, topic sentence).
>
> At 11:19, Lucy tries to set the TV. Lucy underlines the key words in the passage and links the pronouns with their preceding references. Lucy wants to show how the writer structure may the paragraph by using pronouns.
>
> At 11: 25, Lucy scolds two students for not doing any work just before the bell rings. Lucy continues to show on the LCD projects how to get the main points from the paragraphs. Lucy teaches for 4 minutes more, rushing what she wants to say before ending the lesson. Students get impatient and distracted with other students passing by outside the corridor.

Lucy tried to attempt many things in the lesson but the students seemed to be disoriented. Her reliance on Cantonese (L1) and microphone had not improved the clarity of presentation or understanding. This was a lesson which was supposed to practise listening skills, but there was no listening practice until the end of the double-period lesson. It became a lesson on analysing paragraphs instead. Lucy did not establish a clear link between the seemingly incompatible learning activities with the lesson focus to create a sense of purpose for her students. If the above excerpt did not include what she said at 11:03, one probably would mistake the lesson focus as something about reading or writing. Lucy might want to show students how to get the main points from a listening passage through reading, rather than listening, but she did not provide independent practices for them to see how that might work.

From the field notes on Lessons 6 and 7, it shows that Lucy relied on negative feedback, complained and scolded at the Form 1 students almost every 5 minutes. Managing these forty students of two classes was tricky, so Lucy might find justifications for stopping students' chatting and dozing, but her behaviours showed that she seemed to be losing her temper rather than addressing the rules. If her reactions were really effective, the students would have behaved accordingly, but they did not. Her excessive use of L1 in the lesson helped her little to manage the class. All these behaviours probably distracted the students more than she expected. When students were not engaged in activities that required them to listen, speak, read, or write in English, they tended to lose the sense of purpose. However, what Lucy did in these two lessons was not present in Lessons 10 and 11 for her Form 3 (equivalent to Year 9 in England) class. The following excerpt of Lesson 10 shows how Lucy prepared these students with lively, dynamic, experiential learning activities before doing the listening practice:

> …..At 14: 23, Lucy distributes handouts and tells students that they are notes on feeling in different alphabetic order and wants students to act out the facial expressions in accordance with those adjectives. The objective of the lesson is well-defined.
>
> At 14.25, Lucy divides the class into two halves, one on the left and the other on the right. Each group takes turn to do the facial expression for the adjectives expressing feeling. Each group has one representative to give facial expression to his/her group to guess what the adjective is. The activity is not difficult and provides some fun to all students. It is a lively element because students are asked to participate and to take some control. From her smiles, Lucy looks relaxed and enjoys as much as her students. The activity probably can enhance memory as the adjectives do not just carry cognitive meanings, but also will be associated with meaningful learning experience in class.
>
> At 14:42, the whole class goes through all adjectives, while each group acts 4 times.
>
> At 14:44, Lucy asks students to read Letter A in p.9 in the Listening book. Lucy ask them to think about the attitude of the writer and to think about the adjectives they have just learnt.

In these lessons, Lucy became a teacher completely different from what described above. There was no complaining and no yelling, only patience and approving smiles. By the end of Lesson 10, all students had learnt the meanings and pronunciations of the adjectives related to feelings that they needed for the listening practice in Lesson 11. Lesson 10 showed a big contrast with Lessons 3 and 4 in the strong sense of purpose of its learning activities and with Lessons 6 and 7 in its relaxed and supportive learning atmosphere.

Lucy could teach effectively not only with those Form 3 students who knew her well, but also with the Form 1 students who knew her only for a month. In Lessons 14 and 15, Lucy's students were almost the same as those in Lessons 6 and 7, except those less able and most disruptive students were then assigned to other two small, remedial classes. Lucy's class was still a mixed group and its size was still forty, but its composition was different. It roughly took Lucy 15 minutes in Lesson 14 to sort out things and maintain order and discipline, but it was well spent for later activities in Lesson 15. Lucy reinstated the order in class and lecture students on codes and rules as if they first met. No sooner than Lucy started teaching, her speech was clearly different from that in other lessons except Lessons 10 and 11. The frequency of mixed code in her instruction sharply declined with the amount of L2 increased. Lucy's change in the MOI reflected her expectation of the new student composition. Lucy demonstrated her strengths in organise her teaching through full utilisation of the blackboard as she routinely stated the date and the lesson objective (Adverbs of frequency) on the blackboard before showing her examples and notes. By Lesson 15, Lucy showed her strengths in arranging learning activities as she did with the Form 3 students in Lessons 10 and 11. Her high expectation of her students paid off with students' accomplishments and enthusiasm:

> Sooner after the bell rings at 10:50, Lucy asks 5 students in total, all in L2 and students reply in L2 without problems. When Lucy asks a student the meaning of *bored* in one of the questions, the student replies in L1, but it seems to be an acceptable response as Lucy does not ask him to say it in L2.
>
> At 10:55, Lucy seems to be pleased with the performance of the students, so Lucy asks every student to find a partner, either next to him/her or someone at the back. They are expected to ask each other with questions in *How often* and jot down the replies. They have to take turn to do the task. They

can use the worksheet to help them in asking the questions and replying. Lucy walks around to check how students are doing.

In general, 1B students are doing it faster (students of the two classes are sitting in different columns). So at 11:03, Lucy asks 1B students to move away from their seats to ask students in their class five more questions and jot down their replies. 1B students are delighted by their achievements and show great enthusiasm.

By 11:08, 1C students claim that they have finished their task and ask for permission to leave their desks and ask other students like 1B students.

7.3.4 Patterns of consistency and variation in teaching practices

Despite moving around, students did it with discipline and order and hardly any student was not engaged with the activity. Students asked five classmates and changed the answers into statements, rather than simplified forms. That was a highly interactive, student-led activity that one might not expect Form 1 students could do it successfully. It was also surprising that no student was fooling around or taking the advantage of leaving their seats to chat with their classmates. The observations suggest that Lucy demonstrated her teaching abilities only in classes of well-behaved, motivated students, regardless of their academic year groups. Class composition affected her teacher effectiveness.

More lessons of Lucy than other teachers were observed in the five-day classroom observation period. Table 7.3 reveals her teaching behaviours observed across twenty-three lessons. As shown in italics in Table 7.3, Lucy's average score in every underlying dimension as measured in the ISTOF and QoT CFA models is below the sample average. Lucy was rated unfavourably in almost underlying dimensions. About half of the time, Lucy could still show more strength for the two indicators of overall teaching effectiveness. In many lesson, there were some underlying dimensions on which Lucy received better ratings (e.g., *Meta-cognitive skills teaching* in Lessons 18 and 19; *Student engagement* in Lessons 16-20). In these lessons, the two indicators of teaching effectiveness also seemed to be incompatible as the overall teaching quality was rated better than the overall student participation.

Table 7.3: **Consistency and variation in Lucy's observed teaching behaviours across lessons and their relations to indicators of teaching effectiveness**

Lesson No.	Meta-cognitive skills teaching (ISTOF1)	Class-room manage ment & climate (ISTOF2)	Different-iation & support (ISTOF3)	Clarity & logic of presenta tion (ISTOF4)	Student engage-ment (ISTOF5)	Strateg-ies to enhance learning and lesson focus (ISTOF6)	Integra-ted class manage ment & climate (QoT1)	Structur-ed teaching skills (QoT2)	Effective class/ lesson planning (QoT3)	Overall judgement of teaching quality (IND100)	Good individual involve-ment of the pupils (IND24)
Sample Mean	2.16	2.05	1.80	1.97	1.96	1.71	1.99	1.91	1.55	2.93	2.92
Sample S.D.	0.45	0.39	0.46	0.58	0.49	0.46	0.49	0.49	0.35	0.74	0.91
Lucy's Mean	*1.94*	*1.75*	*1.44*	*1.49*	*1.58*	*1.31*	*1.68*	*1.60*	*1.48*	*2.70*	*2.30*
1	AVG	LOW	AVG	LOW	LOW	LOW	LOW	LOW	HIGH	M WEAK	M WEAK
2	LOW	LOW	AVG	LOW	LOW	LOW	LOW	LOW	AVG	M WEAK	M WEAK
3	AVG	AVG	LOW	LOW	AVG	AVG	AVG	LOW	HIGH	M STRONG	M STRONG
4	AVG	LOW	LOW	LOW	LOW	AVG	LOW	LOW	AVG	M STRONG	M WEAK
5	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	M WEAK	WEAK
6	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	WEAK	WEAK
7	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	WEAK	WEAK
8	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	M WEAK	WEAK
9	LOW	LOW	LOW	LOW	LOW	AVG	LOW	AVG	LOW	M STRONG	M STRONG
10	AVG	HIGH	HIGH	HIGH	AVG	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
11	HIGH	HIGH	HIGH	HIGH	AVG	HIGH	HIGH	HIGH	AVG	STRONG	M STRONG
12	LOW	LOW	LOW	AVG	LOW	LOW	AVG	LOW	LOW	M WEAK	M WEAK
13	LOW	LOW	LOW	AVG	AVG	LOW	AVG	LOW	LOW	M STRONG	M WEAK
14	LOW	LOW	AVG	AVG	LOW	AVG	LOW	AVG	AVG	M STRONG	M STRONG
15	AVG	LOW	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
16	LOW	AVG	LOW	LOW	AVG	LOW	LOW	LOW	AVG	M STRONG	WEAK
17	LOW	LOW	LOW	LOW	AVG	LOW	LOW	LOW	AVG	M STRONG	M WEAK
18	HIGH	LOW	LOW	LOW	AVG	LOW	LOW	LOW	AVG	M STRONG	M STRONG
19	HIGH	LOW	LOW	LOW	AVG	LOW	LOW	LOW	LOW	M STRONG	M WEAK
20	AVG	AVG	LOW	AVG	HIGH	AVG	AVG	AVG	LOW	M STRONG	M STONG
21	AVG	AVG	LOW	AVG	LOW	AVG	AVG	AVG	LOW	M STRONG	M STRONG
22	AVG	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	M WEAK	M WEAK
23	AVG	AVG	AVG	AVG	LOW	AVG	LOW	AVG	AVG	M STRONG	M STRONG

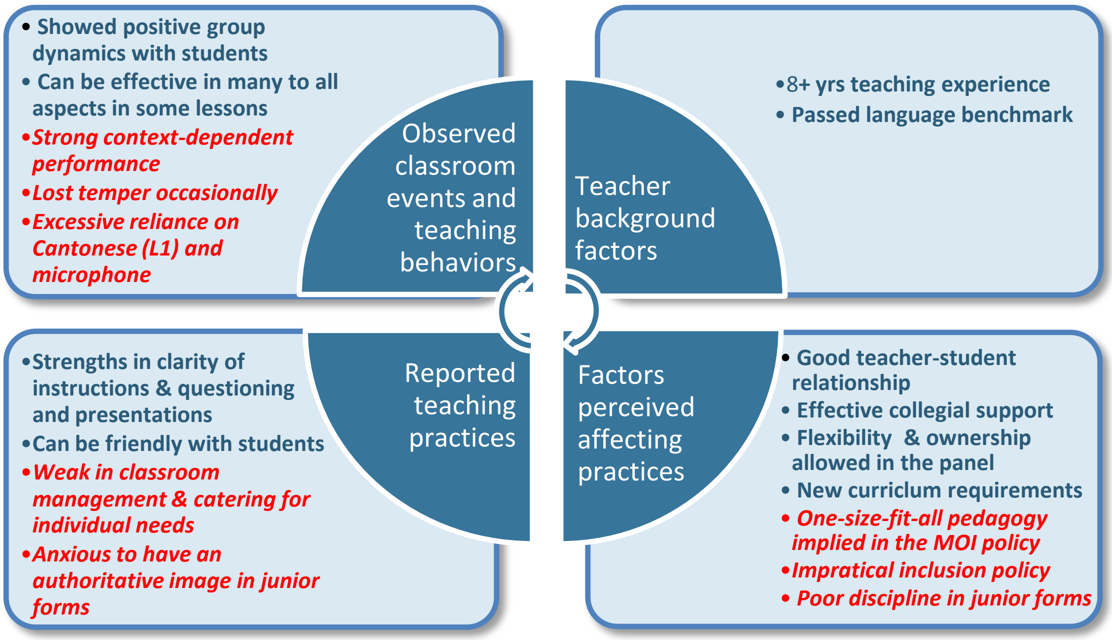
However, the main contrasts were found in Lessons 10, 11 and 15. Lucy scored higher in most dimensions when she was teaching the Form 3 class (i.e., Lessons 10 and 11). She also scored high in Lesson 15 when she taught the newly formed group of Form 1 students after streaming. In Lucy's Lessons 6 and 7, the results of the indicators of teaching effectiveness truly reflected her low score in every underlying dimension of classroom practices.

The seemingly conflicting results suggested that Lucy's teaching effectiveness might be strongly affected by some specific contexts. The variation in her ratings lend support for the theory of differentiated teacher effectiveness, which denies that teacher effectiveness is a generic characteristic of a teacher, but suggest it is likely to vary with different contexts when student grouping changed or when classroom management improved. For example, the abrupt changes in ratings in Lesson 14 and 15 reflected that Lucy's teaching performance improved when she succeeded in reinstating the order after lecturing the newly formed mixed Form 1 class on codes and rules as if they first met in class. Certainly, one may argue that to be effective, a teacher has to be effective in all contexts, but this is what a generic theory of teacher effectiveness would suggest.

7.3.5 Summary

Figure 7.4 summaries findings in Lucy's profile and organizes them as factors that may positively or negatively affect the departmental teaching effectiveness.

Figure 7.4: **Summary findings that characterise Lucy's profile**



7.4 The case of Sally

7.4.1 Brief backgrounds and self-reported views

Sally was a new teacher in *Ming Tak Comprehensive* and sitting in a smaller staff room, a bit isolated from the crowd. She was in her early thirties and had about ten years of teaching experience. She came to *Ming Tak Comprehensive* as her previous school required a lot of travelling. Despite feeling stressful for her work, her motivation as a teacher, commitment to teaching and feeling of job satisfaction were all high and was more conservative about her ability to make a difference to her students' learning. She reported that she did direct instruction more often, but she also stressed cooperative learning and often asked her students to explain concepts to one another. Yet, she rarely engaged them in investigative learning or project-based learning.

7.4.2 Perceptions on teaching and learning in school and factors affecting their teaching practices and teacher effectiveness

Sally reported that she was pleased with her relationships with students as she did not have classroom management problems and learning atmosphere was usually good in her classes. Yet, she declared that she was dissatisfied in her helping students to value learning and motivate their interest in schoolwork, even though she somehow succeeded to make her students believe that they could do well in schoolwork. She was weakly confident in her ability to get parents to become more involved in school activities as parents were generally not involved in the teaching and learning in school.

Sally expressed strong confidence in having a good lesson with Class 2BC in Lessons 7 and 8. She also recognised her strengths in classroom management and teaching strategies. Sally's account of the relationship between lesson planning, reflection, and implementation of lesson activities suggested that she began to command the kind of expertise teaching that guided more by intuition than conscious planning (see Tsui, 2003 on expertise teaching):

> … you have a plan but you also have a plan in your head that you have to draw on the situation to do something that may fit their [the students'] *needs*. For example, like the statistics exercise, I have to change it right away.
>
> However, reflection or planning is not everything because it depends on what happens in the class the level of ability of students , the level of concentration, so even you have prepared a lot of things, it doesn't mean that you can successfully implemented them in the class.

Despite her skills in classroom management and rendering diverse teaching strategies, Sally expressed views about herself which sounded like the characteristics of teachers in their early years of teaching. For example, she said she was still not confident enough in her abilities in catering for the individual differences and her questioning and presentational skills and she sometimes felt exhausted by the demand of finding interesting activities to arouse students' interests:

> Yes, we are very tired mentally, so we will soon run out of tricks to make the lessons interesting for the students. We need new insights in teaching strategies all the time.

Sally was rather enjoying junior form teaching because it was less exam-oriented and students were more responsive:

> … for junior forms, they are more energetic and don't mind presenting their ideas. They are more willing to present their ideas than the senior form students. So it is important to give them chance to voice out their opinions and say some silly things. There may be some differences between students in different classes in terms of levels of abilities, but in general, I love more interacting with the students. Interactions may not be always in English; sometimes I may just tell them some silly stories to keep them interested in the classroom.

Sally's teaching preference was consistent with her goal in teaching that expressed a non-exam-oriented, student-centered expectation.

> I just want the students to love English, even they may not learn a lot in my lessons but by the time they leave school, they may still have the ability and interest in English.

Sally emphasised more on enjoyment in learning and interactions with students. She created an atmosphere that students can feel the respect when expressing themselves, which is particularly important for teenagers. She is not just teaching, but she is showing her care and hope in and outside the classroom to build up a good teach-pupil relationship that is crucial for a relaxed and supportive learning atmosphere:

> I would like to talk sometimes after the lessons to build up a relationship with the students. Teenagers are not rational enough, so once they have a good relationship with you, they will listen to you. So I really have to spend time to talk to them, expressing my hope to them. I used the same strategy while teaching in the other school.

Therefore, what Sally wanted to build on the top of her insistence of discipline and order in the classroom might be a mutual trust and positive emotional relationship. Sally's approach was a clear contrast with the approach to maintain discipline just by maintaining an authoritative figure. It was not just a matter of different teaching styles. In effect, both approaches might look similar because classroom management and classroom climate occur concurrently, but they differed in essence and in practice. Sally did it by establishing her classroom management skills in a supportive classroom climate grounded on solid and positive teacher-pupil relationship, rather than on classroom management.

Sally recognised that her teaching in the senior form students was still exam-oriented, but she also expressed the frustration for failing to motivate the senior form students, as in Lessons 12 and 13:

> Form 4 they are refusing to contribute their time to their work, for example, I feel like that I am a walking dictionary and I feel very tired. If am not going to give them the translation, they will just sit there and fall asleep. It was really exhausting that I do all the translation and explanation all the times. Take 2BC as an example, when I told them to prepare for their work, they would do it, but not for form 4. They would look up the dictionary and make the translation and the lesson smoother.

However, Sally attributed the indifferent attitudes and poor motivation in students to their learned helplessness due to their lack of experiences in success and senses of accomplishment in the past:

> I deeply believe that every student wants to excel in their learning, but over the years they have learnt helplessness instead. They have learned to cope with failures. If you ask any one of the junior form students, I think s/he will want academic achievement. They lust for success. Sooner or later, these students who are at the bottom of the class when they were in the primary students will repeat the pattern in the primary school and become losers again. It is really helpless. They are just too far behind in the race that they don't have the time and space for them to start all over again.

Under the SSPA system, most of the intakes of *Ming Tak Comprehensive* were the low achievers in the primary schools who lacked academic accomplishments. According to Sally, these students would soon learn helplessness again in the secondary school if they failed to become achievers in their junior form years:

> Honestly, what parents and teachers care is how well you are doing academically. That is how we define a good student. The recognition is not there [for or students]. When I talk to them, most of them realise that their life chances is very much dependent on their education. Even those junior form students they will have a strong sense of efficacy, like 2BC or 2D students would do things better when you are giving them just a sticker or a stamp for their achievements. They will get excited for that

> kind of little recognition. We really shouldn't underestimate the power of recognition. What we have to think is whether their experience of success is sufficient enough for them to make them face the past and future experience of failure.
>
> … It is just too bad that we evaluate a person by academic achievements and these students are going to fail much easier and we don't give enough time for them to stand up again. How many students who can have a high self-concept even though s/he is academically weak? Very few.

What seems to have contributed to Sally's success in building a solid teacher-pupil relationship was her recognition of her students' need of accomplishments to sustain their senses of efficacy, which is particularly important in a society which overemphasises academic achievement. It seemed that Sally has a deeper understanding of the needs of the students and the cause of their lack of motivation than other teachers, including Charlie.

Yet Sally still had her anxiety in the extra demand in subject knowledge in facing the recent change in academic structure. She worried that her teaching effectiveness and teaching practices would be undermined as she might not be well-prepared for the change:

> I worried that I may not have sufficient knowledge in teaching stuffs that I am expected to teach, like poetry for example. We learn some at the university, but I am not sure whether it is sufficient for me to teach. English teachers have huge responsibility as students' exam results affect their future.
>
> We don't have the training in teaching poetry or drama, which native speaker teachers may be good at.

Sally was also uncertain about good teaching practices are sufficient to overcome the lack of learning support in *Ming Tak Comprehensive* because the learning atmosphere in school and at home was not desirable and the lack of reading abilities in students:

> You see, the learning atmosphere is not good with all that noise in the classroom, lack of motivated peers, and they don't mind listening to English songs, but most just don't like reading. They said they would have headaches when they read… I'm not sure whether they meant some physical handicaps or underdeveloped reading habit.

Sally also worried that their reflective practices were rather limited in scope and might be insufficient to enhance teaching effectiveness:

> We certainly reflect on our lessons, but we rarely do so on the modules and their direction. This may not be enough.

Sally thought that it was against the students' benefits to adjust the current MOI policy despite its fringe benefit to English learning:

> It is irony that students may be more benefited in learning English if they learn other subjects in English, but it is rather selfish for us to ask them to do.

7.4.3 Classroom events that characterise teaching practices

From the above accounts, Sally's teaching practices can be characterised by her student-centred approach that may have contributed to her teaching effectiveness. Sally managed her classes with a business-like manner, but she was flexible and responsive. She could share her own life stories when discussing pocket money with students or let her students to listen to a song after finishing their task before the lesson ended. She did what one would expect a teacher should do as well as what students expect a teacher would do top surprise them with joy. Sally was good at bringing lively elements into the classroom. She had high expectations of her students as reflected in her predominantly L2 teaching in all classes and her insistence on their obedience to rules and codes.

In Lesson 7, Sally showed her strengths in adapting materials and activities to the unexpected needs arose in the immediate classroom context:

> At 9:08, Sally notices that the worksheet is not formatted as it would allow students to write all the answers because there are only 3 columns for the answers of 3 groups, while she wants all the five groups to give their answers. She makes up her mind to make immediate changes:
>
> Let's change and use the G.E. exercise book to do the same exercise for the whole class. (She explains her reason to make the change to the class).
>
> At 9:10, Sally draws a new table on the blackboard and says with her microphone: *Put down the table in your G.E. book first.*
>
> … At 9:33, the class finishes Q5 (*How often do you go shopping?* There are 5 options for the answer). Sally complains about the noises that the students are making and orders: *Just jot down the results and do Q5*.
>
> At 9:36, Sally explains how to do a bar chart with the answers on the blackboard.

> By 9:44, Sally asks the students to put their chairs back to their original settings, signalling the group activity is finished. The class finishes earlier, so Sally has a causal chat with her students on how much pocket money she had when she was a teenager and how she spent it.

The excerpt showed how Sally made necessary and immediate adaptations for the group activities. On the one hand, it seemed that she had not anticipated this during her lesson preparation, but she managed to make the necessary adjustments in class and instructed her class to follow her instructions carefully and smoothly step by step. As a result, the lesson finished earlier than expected but it was full of activities that most students seemed to enjoy with their achievements. Her teaching on bar charts seemed to be an extra activity as it was not shown in the worksheet. It is useful, however, as it is relevant to build up their survey presentation skills.

In Lessons 12 and 13, Sally spent much of Lesson 12 on the vocabulary test and on explaining the reading passage paragraph by paragraph, leaving not sufficient time for teacher-pupil interaction and independent learning activities by the students. As was seen in other lessons of Form 5 classes led by the other teachers, Sally did a lot of drills and explanations of materials that made the learning activities rather homogenous and dull.

7.3.4 Patterns of consistency and variation in teaching practices

Table 7.4 below shows the patterns of observed teaching behaviours of Sally. Sally probably can be considered an exemplar that supports a generic theory of teacher effectiveness. She was hardly weak in any dimension of teaching behaviours and capable of scoring high in any of these dimensions. Sally's performance showed little variability across different dimensions and across lessons. Thus, her teaching behaviours conformed to the predictions of a generic theory of teacher effectiveness.

However, this does not mean that Sally's performance necessarily contradicted the prediction of the differentiated theory of teacher effectiveness. For example, occasionally, Sally might be rated low in many underlying dimensions in a lesson like Lessons 12 and 13, but these weaker dimensions did not seem reflect in her scores in the indicators of teaching effectiveness. That is, her lower scores seemed to exist only in statistical calculations, but unnoticeable in the overall impression of teaching quality or in the overall participation of students.

Table 7.4: **Consistency and variation in Sally's observed teaching behaviours across lessons and their relations to indicators of teaching effectiveness**

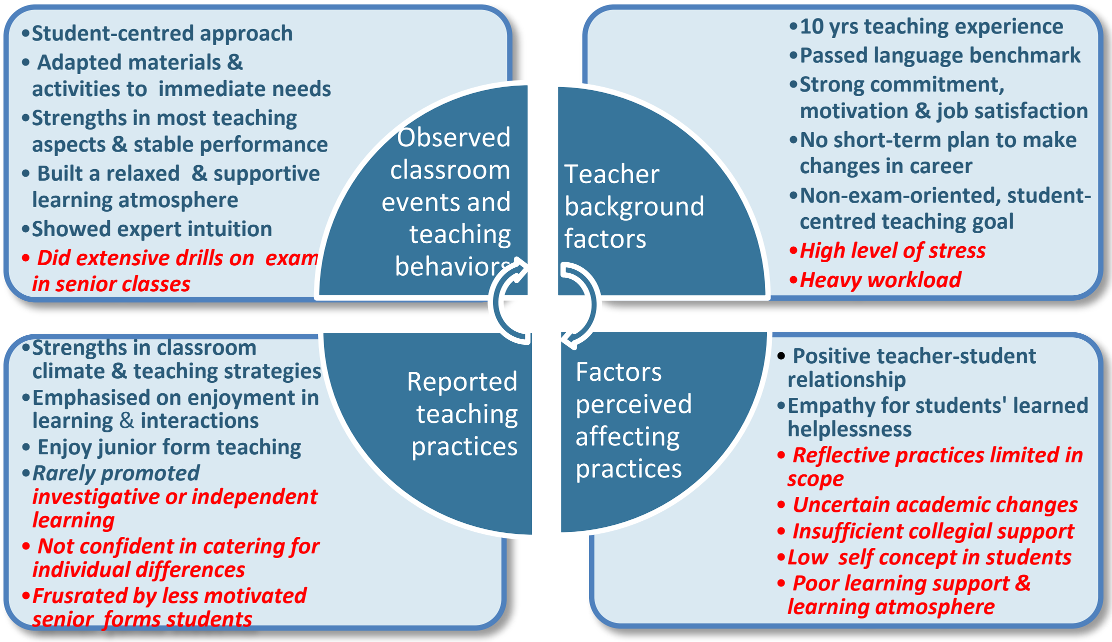
Lesson No.	Meta-cognitive skills teaching (ISTOF1)	Class-room manage ment & climate (ISTOF2)	Different -iation & support (ISTOF3)	Clarity & logic of presenta tion (ISTOF4)	Student engage-ment (ISTOF5)	Strateg-ies to enhance learning and lesson focus (ISTOF6)	Integra-ted class manage ment & climate (QoT1)	Structur-ed teaching skills (QoT2)	Effective class/ lesson planning (QoT3)	Overall judgement of teaching quality (IND100)	Good individual involve-ment of the pupils (IND24)
Sample Mean	2.16	2.05	1.80	1.97	1.96	1.71	1.99	1.91	1.55	2.93	2.92
Sample S.D.	0.45	0.39	0.46	0.58	0.49	0.46	0.49	0.49	0.35	0.74	0.91
Sally's Mean	2.39	2.43	2.19	2.41	2.33	2.13	2.24	2.27	1.76	3.45	3.50
1	AVG	HIGH	AVG	HIGH	AVG	HIGH	AVG	HIGH	HIGH	M STRONG	STRONG
2	AVG	AVG	AVG	HIGH	AVG	HIGH	AVG	HIGH	AVG	M STRONG	STRONG
3	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
4	AVG	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
5	AVG	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
6	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
7	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
8	AVG	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	M STRONG
9	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	STRONG	STRONG
10	AVG	HIGH	AVG	AVG	AVG	AVG	AVG	HIGH	AVG	M STRONG	M STRONG
11	LOW	HIGH	AVG	AVG	HIGH	AVG	AVG	HIGH	HIGH	M STRONG	M STRONG
12	LOW	HIGH	LOW	LOW	AVG	AVG	AVG	AVG	AVG	M STRONG	M STRONG
13	LOW	AVG	LOW	LOW	LOW	AVG	LOW	AVG	LOW	M STRONG	M STRONG
14	AVG	HIGH	AVG	AVG	AVG	HIGH	AVG	AVG	AVG	M STRONG	M STRONG
15	HIGH	HIGH	HIGH	AVG	AVG	HIGH	AVG	AVG	AVG	M STRONG	M STRONG
16	HIGH	AVG	HIGH	HIGH	HIGH	HIGH	AVG	AVG	AVG	M STRONG	M STRONG
17	HIGH	HIGH	HIGH	AVG	HIGH	HIGH	AVG	AVG	HIGH	M STRONG	M STRONG
18	HIGH	HIGH	HIGH	AVG	HIGH	HIGH	AVG	AVG	HIGH	M STRONG	M STRONG
19	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	AVG	HIGH	AVG	STRONG	STRONG
20	HIGH	HIGH	HIGH	HIGH	HIGH	HIGH	AVG	HIGH	HIGH	STRONG	STRONG

Although Sally only received an average score for *Integrated classroom management and climate* in both Lessons 19 and 20 and an average score for *Effective class/lesson planning* in Lesson 20, Sally still scored predominantly strong in the two indicators of teaching effectiveness. None of Sally's lessons was found weak in *Structured teaching skills*. The results of Lessons 12 and 13 indicated that an average score in this dimension, but not any other dimensions, seemed to be sufficient for Sally to score more strengths than areas needed improvement in the two overall indicators. These results were consistent with the findings that showed the strong association between the dimension *Structured teaching skills* and the indicators of teaching effectiveness discussed earlier in Section 6.4.

7.4.5 Summary

Figure 7.5 summaries findings in Sally's profile and organizes them as factors that may positively or negatively affect the departmental teaching effectiveness.

Figure 7.5: **Summary findings that characterise Sally's profile**



7.5 The case of Linus

7.5.1 Brief backgrounds and self-reported views

There was not anything distinctive in Linus' reported biographical backgrounds, except his extra role as a SEN teacher. Linus has taken up his new role as a SEN teacher this year because of previous teaching experience in Band 3 schools. For this reason, he was assigned to teach remedial classes for SEN students and an academically least able class in the senior form. To his colleagues, as well as to the researcher, he was remarkably cooperative and enthusiastic. In other words, Linus' behaviours conformed very much to his role and other people's expectations of his role. As he joined *Ming Tak Comprehensive* only for a year, he showed no intention to job or teaching role in two years time.

7.5.2 Perceptions on teaching and learning in school and factors affecting their teaching practices and teacher effectiveness

In the teacher survey, he reported strong motivation as a teacher, commitment to teaching and feeling of job satisfaction, though he expressed uncertainty about his ability to make a difference to students' learning. He reported having good relationships with students and held positive views about his teaching and effectiveness. He was satisfied with his efficacy in promoting learning environments for students but was not pleased with his strategies in assessing and motivating his students to learn. He was very dissatisfied with the familial support that his students could get.

When asked to comment on his teaching during the observation period, Linus considered that his teaching in Lessons 7 and 8 with a mixed Form 1 class was the best among all the eighteen observed lessons. He was rather confident that the role-play based on the poem he taught in class was impressive and enjoyable to the students. He thought it was a successful activity as most students were engaged in the task. In contrast, Linus recognised that Lessons 15 and 16 with Form 5 were not satisfactory:

> Basically, I enjoyed all the classes here, but 5B I think they seem to lack motivation to learn....[The researcher interrupted and clarified: you meant the last lessons with 5B] Yes, the lessons yesterday with 5B. I asked them to write a letter but not many students interested in writing and their learning was so passive.

It seemed that Linus attributed the lesson quality to the lack of motivation of the Form 5 class. He was also not confident in achieving at the same time the goal to cater for the diverse individual needs of the students of this class who were more heterogeneous in ability and the goal to prepare them for the public exam:

> I found myself still have some rooms for improvements in catering for individual differences, especially in a class of mixed abilities. It's difficult to handle when students have different levels. Another area [needs improvement] is assessment and evaluation. Especially during the examination time we need to set up questions for the exams, how can cater for students' different levels of abilities in one [exam] paper?

Linus perceived a conflict between his role and his sensibility. On the one hand, his role was to use his professional knowledge to set up questions of different levels of difficulties that would inform the students their different levels of abilities. On the other hand, Linus' sensibility made him sorry for those students who were going to fail in the exams and felt obliged to support them.

Linus raised two paradoxes here. First, there was a paradox between two goals of education: a goal that advocates for a student-centred approach of teaching to enable students to enjoy learning and another goal that emphasises the selective function of schooling and examination that eventually creates winners and losers. Despite the publicity of the first goal in the public policy documents (EC, 1999, 2000), Linus, as a teacher, was less confident about the assessment strategies available to him. However, as a SEN teacher, Linus did show some unique sensibility and perspectives in students' psychological needs by relating them to the negative impacts of the assessment system:

> I am less confident to be consistent in assessing students. It seems that our expectations for formative and summative assessments are rather different. In the summative assessment like form test and term exams. The public exam is a must for them. They have to face the reality. The junior forms are the same. Every term they have examinations. They can't escape. There will be failures, but they still have to face the reality. They have to be compared with the majority….When we are doing the formative assessment, we may just assess what they've learnt but for the summative ones, we distinguish their levels of abilities as well as and their abilities to transfer their knowledge, something our students are not good at, especially in English…. Although many will be doing averages, some students will be very disappointed after the exam, after they are compared with other students.

Linus attributed the cause of students' lack of experiences in success in the school and the public assessment systems. In Linus' opinion, most of the students in *Ming Tak Comprehensive* were the "losers" in the standardised assessments in schools or in public, which are valued more highly in the Hong Kong society. Students became disillusioned when they had to take the public exam that was beyond their current abilities. The reality is that students are not compared when they are ready but in the same time frame set for all:

> When I gave 3A and 5B the same test at the beginning of the term, the passing rate was much higher in 3A than in 5B. They are just called Form 5; it is just a label. The students themselves know that they haven't reached the Form 5 level. For example, when these students came to ask me for help, they would ask me, "I've just got the primary level, how can I

> compete with others in HKCE?" I still have to give them encouragements when they come to seek help.
>
> Their achievements are on average low, but their levels of motivation depend on the class. 1D is OK, but very poor for 5B because they have no confidence; their self-image is very low because they are aware of their inadequacy.

Based on their past grades (i.e., the best predictor of student outcomes with an effect size of 1.44, according to Hattie, 2009, p.297), the Form 5 classes understands their incompetence, and their low motivation reflects their learned helplessness.

The second paradox that Linus raised concerned the inclusive policy. Although Linus said he was less confident in providing differentiation and support in classes with mixed abilities, he was appointed as a SEN teacher. However, according to Linus, his role as a SEN teacher for English did not particularly require him to possess exceptional abilities in rendering differentiated support to different students in class. This because the school's strategy was exclusive, rather than inclusive, in the sense that SEN students were taken out of the normal classrooms as a small but homogenous group for remedial teaching as he did in Lessons 17 and 18 with six students.

Linus also thought that his strengths were classroom management and questioning and presentation skills. He explained that he deliberately ignored Dennis' disruptive behaviours as *the guidance team thought he was a SEN student and suggested not to provoke him*. He admitted that his new role might have affected his extra tolerance to disruptive behaviours of junior form students. Linus also believed that peer group pressure could be powerful in maintaining discipline:

> I've tried to use peer group pressure to keep discipline. It might be sometimes taking longer to make it work. The disruptive students are not always the weakest in learning, you see. They know that they are going to interfere other's learning and they aren't welcome. So when they are punished later, nobody would take their side. I always find peer group powerful in this way.

Therefore, Linus chose to engage the other students and ignore the disruptions made by Dennis as far as possible instead. Linus' explanation indicated that he should not take all the blame for not stopping the unruly

Dennis. The paradox was that the requirements of inclusion and classroom management were sometimes unclear and reflected the school policies, rather than a pure decision of the frontline teacher, especially when the teacher did not have the professional knowledge to determine what to do and when to draw the line. This also explains Lucy's frustration about the inclusive policy of the government because, according to the teachers, the officials sent in SEN students but, for the sake of the students' privacy, they would not let the school to know who these students were. As the government cannot provide the adequate training to all teachers, interested teachers are expected to anticipate what training they would need for this new policy and enroll themselves in the relevant courses.

In reflecting on the burden and limited success of the teachers to raise student outcomes, Linus said he once puzzled this for a while but came up with an unsurprising answer—the family:

> They [students] don't want to pay effort. I don't want to use the word lazy, but I just don't understand why. I ask them but they can't tell me an answer. I think the family has a major influence on them. In the parents' night, a parent of a Form 1 student asked us, "How come he needs to do his homework at home? Isn't he supposed to finish it in school? That's your [teachers'/school's] responsibility to teach him at school. Students should enjoy their lives at home and watch TV. It's a rare case but you can see how the parents might have influenced the kind of effort that the students would put in.

However, in reviewing things he did in successfully raising student outcomes, Linus stressed the importance of homework:

> I always reflect whether I have given the students enough homework. It may be true that not all of them will do it, but they need sufficient homework to reflect on what they have learnt.

He was also stressed reading, vocabulary and grammar:

> I am better at teaching reading and skimming, even for the junior forms. It's important to provide them the guidance to think how to get the meanings from the text. They also need vocabulary. For example, why they lack the confidence? Because they lack the vocabulary. I think teaching grammar is a must as we need to tell them the rules how to make up a sentence.

Linus was concerned about the importance of reading, vocabulary and grammar as the basic blocks and inputs to build up the foundation for other language skills.

Linus regarded the external impacts could enhance his teaching effectiveness through professional development:

> The new change [i.e., the new academic structure] is not necessary negative. We can learn more through courses and workshops. Our practices can be enhanced in way, but this happens only provided that you can be admitted to those courses. They are not always available, you know.

Therefore, Linus thought collegial support and collaboration affected his teaching effectiveness more, but he also stressed that organizational policy affects teachers' effectiveness in general:

> Collaboration between colleagues usually give us strong power and support, you know…we need to work together as team spirit is very important. We are in the same boat.
>
> ……Organizational characteristic or policy is important as it comes before the implementation. The decision makers are important here.
>
> I am looking at the issue in a more macro perspective. For example, once the school decide to stream the students, we don't have mixed abilities that much. We have five classes for Form 1, but for Class B, C and D, we divided them into four groups for Chinese, but we divided into three groups in English. This may affect the overall effectiveness.
>
> I have the flexibility to decide which class I should give more support for my role. …. I have the flexibility. My role is more flexible and the flexibility of my role is assigned by the panel's discretion under the mutual understanding of other English teachers.
>
> …..For example, normally, I was not supposed to co-teach with Dianna, but she said I might help with the poem teaching. So we joined the two classes. I can be flexible.

With the flexibility to fulfil his role, Linus might feel that he was empowered. His role and work were not only recognized by Charlie but also by other teachers in the department. He attributed this kind of flexibility and empowerment to the organizational culture of the school. Linus found that the working culture in *Ming Tak Comprehensive* was very different from his previous school though both were Band 3 schools:

> The middle managers [in my previous school] would like to pass their responsibilities to the teachers but here is different. The first year I came here and I was shocked by the conflicts and clashes [between the senior management and middle managers] in the staff meeting, you know I have never attended this kind of staff meeting before. I didn't understand the conflicts here up to now. We never argued in my last school, and we, junior staff, did things just like what our parents tell us to do. But we may need arguments for reaching a goal in a rational way, except sometimes we may get too emotional. I think our goal is clearer than my previous school. No, actually, the goal in the previous school was also clear, but teaching was just the means for survival there. Here, we have the bargaining power with the senior management. We can argue on the basis of teaching and learning goal to make things more realistic. Actually, both parties have very clear pictures for themselves and for the students.

The middle managers in his school who shared their responsibilities with the junior teaching staff have made the junior staff feel that they were supported and empowered. Linus did not think that the apparent conflicts and clashes were necessarily negative as long as they were rationally based on arguments related to teaching and learning goals and on intentions to make implementations practical.

The impact of organisational policy on Linus' teaching practices was also found in a deliberate attempt to implement reflective learning in students:

> I tried to ensure that the students can reflect on what they have learned by highlighting the lesson focus, especially before the end of the lesson, a kind of closure. …. I thought homework is important. I manage to give them homework at least twice a week. They need to reflect on what they've learnt through the homework and I also need to know how much they've learnt.

7.5.3 Classroom events that characterise teaching practices

Linus stressed his implementation on the reflective learning policy. He was genuinely convinced that the policy was helpful to his students and he interpreted and implemented the policy in two practical terms: to revise the lesson focus before the end of the lesson and to assign sufficient amount of homework to consolidate learning.

In Lessons 1, 2, 7, 8, 13 and 14, Linus was co-teaching two Form 1 classes with another teacher. In terms of the cost of human resources, co-teaching was very high as Linus' partner was not a teaching assistant. The

English department was only trialling it in Form 1 at the beginning of the term before the streaming process of Form 1 classes was completed by the end of September. A few incidents in the field notes suggest the teachers did not act proactively towards disruptive behaviours and firmly promote students' respect towards class rules and classmates in Lessons 1 and 2:

> At 8:58, Dennis [the pseudo name of a student who disrupted the classes regularly], who sits at the corner near the window, continues to distribute the class, but both teachers ignore him. Dennis was originally in Lucy's class in earlier observations. He disturbs all the time, but rarely any teacher attempts to handle him.
>
> At 9:05, Dennis disturbs again, but both teachers deliberately ignore him, but Teacher 2 stands in front of him like a firewall to separate him from the other students.
>
> At 9:24 in Lesson 2, Dennis comes out and walks across to the other side of the classroom, while the whole class is reciting the poem. Teacher 2 wakes up a sleepy student.
>
> Shortly before 9:37, Dennis asks the student who is eating gum to give him some and he distributes it to other students. It seems that Dennis is better than the teachers in spotting misbehaviours in class. Teacher 2 is not helpful to stop disruption in class.
>
> At 9:42, Dennis asks his classmate to give him some gum in L1 with obscene words that rhyme with "oral sex". Dennis's misbehaviour is still ignored and he asks for going to the toilet again after 3 minutes.
>
> Linus continues his teaching in the foreground with Dennis disturbing at the background. When he asks the class what the owl and Pussy cat did in the 3rd stanza. Dennis replies in L1: *copulation. Most romantic is to have sex*. Other boys start to yell and burst into laughers. Linus tries to stop the boys and asks them not to talk about sex in class. Girls sitting near the boys are embarrassed as they nodded their heads down to avoid looking up to the whole scene. Linus manages to cool down the class but the teaching is interrupted and most students appear to have reduced concentration.

Despite the disruptive behaviours of Dennis, Linus was successful in engaging other students in simple questions and limited mixed codes:

> At 9:11, Linus asks students what other two words rhyme in the last stanza. A student answers: *love; lovely*. Another student corrects him: *love; above*. Linus writes on the blackboard: *love; above*. (Students are able to express themselves with simple answers. Their comprehension is better than their speech though.)

At 9:13, Linus asks: *If you were the Pussy cat what would you do?* Some students say they are happy as they brought money with them. Linus hints that the answer might be in the next stanza.

At 9:15, Linus writes and then asks students in L2: *What is a Piggy–wig?* Then in L2: *What does it mean?* And then answers in mixed code: *because* (L1) *owl looks like a pig* (L2).

At **9:20** in Lesson 2, Linus writes on the blackboard: *wood/wu:d and stood/stu:d; long u sound*. (The bell rings). Then Linus asks which two vowels rhyme in the poem. Many students identify the following pairs and reply: *sing/ring; owl/fowl; word/stood; where/there; away/a day*. Linus asks in L1: *Are there any repeated lines here?* Students answer: *With a ring at the end of his nose. His nose. His nose.*

Linus engaged students with the materials like the poem he prepared and group activities like the repeated poem recitation and the role-play. Linus had successfully made the poem recitation in Lessons 1 and 2 more interesting by changing it into a role-play group activity in Lessons 7 and 8. However, the other teacher's role in the class was not clear. For example, the other teacher did not helpful much in keeping the discipline and order, nor shared any teaching part in the role play such that it was carried out in so much noise and laughter that students sitting at the back found it hard to follow. Her contribution was so little that it raises doubts about the purpose and effectiveness of co-teaching as well as the kind of collaboration that Linus treasured much in the interview. Linus had to stop and threaten the class many times with the possibility of detention in an attempt to keep the noise level down. The role-play engaged the students to make responses as a group like a chorus without any subsequent related independent work. Linus did not evaluate how much individual students had actually learned through questioning.

7.5.4 Patterns of consistency and variation in teaching practices

Teaching is also a live performance that a teacher must react instantly to unexpected incidents. For example, Linus had to decide when he should stop Dennis' unruly behaviours. He chose to stop him only after he disturbed the class to get more attention. Although Linus did not seem to be unable to manage his students, he often just yelled loudly at them. In doing so, Linus might look like losing his temper, rather than reinforcing the rules. Neither

Linus nor the other teacher did something to restore their authority and order in class before the problem gradually became out of control. As he said in the interview, his decisions in classroom management were affected by his role as a SEN teacher and the school policy on handling students with behavioural problems.

Table 7.5 below shows an overwhelming number of low scores found in various dimensions of teaching behaviours observed in Linus' eighteen lessons.

Table 7.5: **Consistency and variation in Linus' observed teaching behaviours across lessons and their relations to indicators of teaching effectiveness**

Lesson No.	Meta-cognitive skills teaching (ISTOF1)	Class-room manage ment & climate (ISTOF2)	Different -iation & support (ISTOF3)	Clarity & logic of presenta tion (ISTOF4)	Student engage-ment (ISTOF5)	Strateg-ies to enhance learning and lesson focus (ISTOF6)	Integra-ted class manage ment & climate (QoT1)	Structur-ed teaching skills (QoT2)	Effective class/ lesson planning (QoT3)	Overall judgement of teaching quality (IND100)	Good individual involve-ment of the pupils (IND24)
Sample Mean	2.16	2.05	1.80	1.97	1.96	1.71	1.99	1.91	1.55	2.93	2.92
Sample S.D.	0.45	0.39	0.46	0.58	0.49	0.46	0.49	0.49	0.35	0.74	0.91
Linus' Mean	*1.92*	2.05	*1.61*	*1.69*	*1.74*	*1.54*	*1.71*	*1.65*	*1.54*	*2.50*	*2.50*
1	AVG	LOW	HIGH	AVG	HIGH	LOW	LOW	AVG	HIGH	M STRONG	M STRONG
2	LOW	LOW	HIGH	LOW	HIGH	LOW	LOW	LOW	AVG	M WEAK	M WEAK
3	LOW	LOW	AVG	LOW	AVG	LOW	LOW	AVG	AVG	M STRONG	M WEAK
4	AVG	AVG	AVG	AVG	AVG	AVG	AVG	AVG	HIGH	M STRONG	M STRONG
5	AVG	AVG	AVG	AVG	HIGH	AVG	AVG	AVG	HIGH	M STRONG	M STRONG
6	AVG	AVG	AVG	LOW	HIGH	AVG	LOW	LOW	HIGH	M WEAK	M WEAK
7	LOW	LOW	AVG	AVG	HIGH	AVG	AVG	AVG	HIGH	M STRONG	M STRONG
8	LOW	LOW	AVG	LOW	HIGH	LOW	LOW	LOW	AVG	M WEAK	M WEAK
9	HIGH	AVG	AVG	AVG	HIGH	LOW	AVG	AVG	HIGH	M STRONG	M STRONG
10	HIGH	AVG	AVG	AVG	HIGH	LOW	AVG	AVG	HIGH	M STRONG	M STRONG
11	AVG	AVG	AVG	HIGH	HIGH	AVG	AVG	AVG	HIGH	M STRONG	M STRONG
12	AVG	AVG	AVG	HIGH	AVG	AVG	AVG	AVG	HIGH	M STRONG	M STRONG
13	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	WEAK	M WEAK
14	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	WEAK	WEAK
15	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	M WEAK	M WEAK
16	LOW	LOW	LOW	LOW	LOW	LOW	LOW	LOW	AVG	M WEAK	M WEAK
17	LOW	AVG	LOW	AVG	HIGH	HIGH	HIGH	LOW	AVG	M STRONG	M STRONG
18	LOW	AVG	LOW	AVG	HIGH	HIGH	HIGH	LOW	AVG	M STRONG	M STRONG

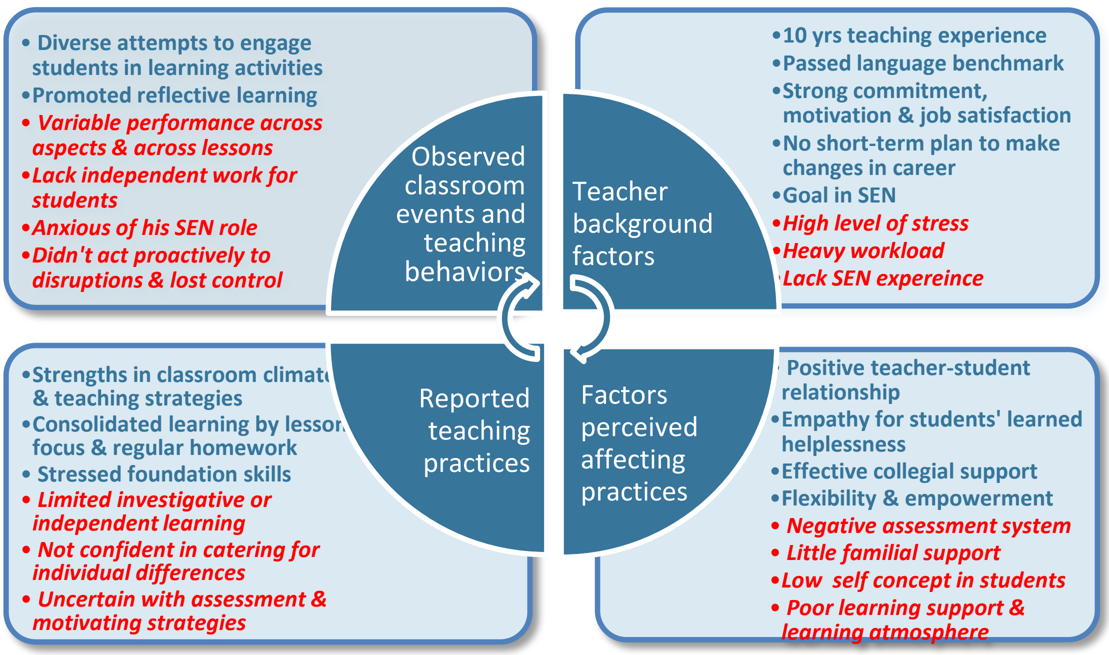
Note: The lessons in which Linus collaborated with another teacher are labelled with the lesson number bracketed in a square.

Linus appeared to be a typical example of teachers whose teaching shows most variability, not only across lessons but also across dimensions within a lesson. Linus' perception about his strengths in classroom management is incompatible with the ratings in Table 7.5. Rather, Linus scored high in his abilities to engage students (*Student engagement*) and lesson planning (*Effective class/lesson planning*) and these strengths were truly reflected in lessons where Linus scored in these two dimensions as well as in the two indicators of overall teaching effectiveness. Although Linus highly regarded the role play in Lessons 7 and 8 as he thought that he could engage most of the students in the activities, these lessons were not rated higher than his other lessons like Lessons 9-12. Failing to receive a high score in *Structured teaching skills* suggests that Linus might have received only lower scores for the teaching practices associated with the two indicators of overall teaching quality and individual involvement by the students. As noted in Section 6.4.2, this dimension was strongly associated with these two indicators of overall teaching effectiveness. It should also be noted that Linus received low scores in all dimensions and the indicators of teaching effectiveness in Lessons 13-16, but these all happened in one day. Thus, to be fair, Linus could be a teacher whose performance might show great variability and perhaps more clearly the negative impacts of a “bad” day, where unanticipated factors might have played a part.

7.5.5 Summary

Figure 7.6 summaries findings in Linus’ profile and organizes them as factors that may positively or negatively affect the departmental teaching effectiveness.

Figure 7.6: **Summary findings that characterise Linus’s profile**



7.6 Discussion and conclusion

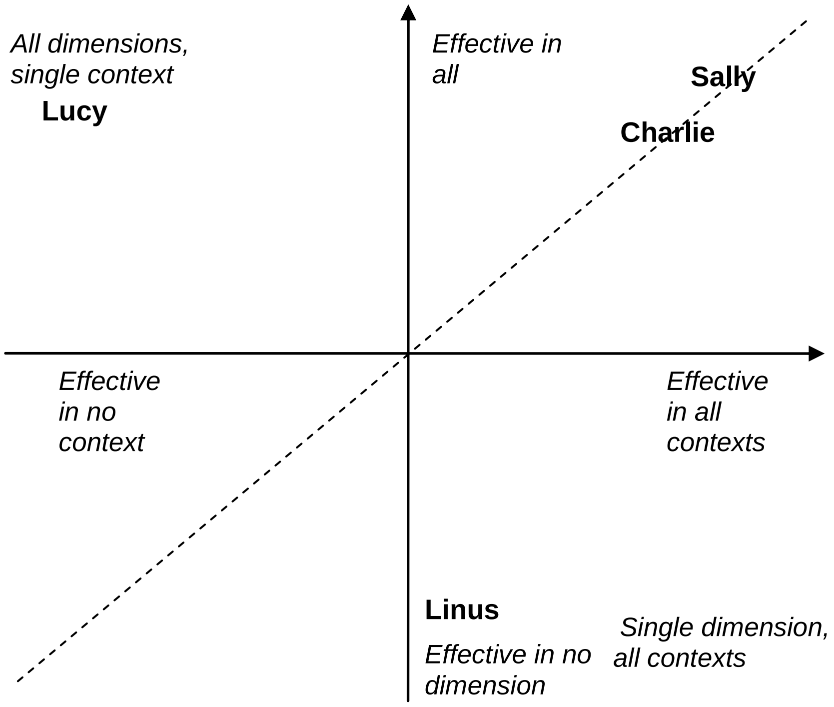
Both the classroom events and the quantitative results obtained from the classroom observations were consistent with findings in other research. First, teachers' behaviours as rated in classroom observation did not seem to be related much with teacher characteristics such as their sex, teaching experience, and qualification (Kyriakides, 2005; Muijs & Reynolds, 2000), because the four teachers had similar qualification and years of teaching experience. Second, effective teaching behaviours as rated by classroom observation schedules were found to be associated with the overall impression of teaching quality and students' individual participation (van de Grift, 2007). Third, Sally, and to a large extent Charlie as well, could be considered as effective teachers like those in the ECP study (Day et al., 2008; Sammons & Ko, 2008) as they showed higher consistence and effectiveness in most underlying dimensions of observed effective teaching behaviours. Fourth, effective teaching exists in different classroom contexts such as year groups.

These findings were consistent with the generic concept of teacher effectiveness (Creemers & Kyriakides, 2008). For example, two teachers, Charlie and Sally, were rated highly in most of the dimensions across most of their lessons. Sally was exceptional as she showed the least variability and thus she can be considered as an exemplar to illustrate that *an effective teacher would be effective in all dimensions of her teaching at all times*, *as a generic theory of teacher effectiveness (GTE) would predict*. None of Sally's lessons was found weak in the dimension *Structured teaching skills*. In the excerpt of Lesson 7, Sally also showed her exceptional strengths in *adapting materials and activities to the unexpected needs arose in the immediate classroom context,* which reflected "a speed of action, forward-directed solutions, accuracy, enriched representations, and rich elaborations of knowledge in terms of depth and organisational quality" (Leinhart, 1989, as cited in Tsui, 2003, p.56, as "expertise in teaching"). Charlie could be another exemplar if he had structured his lessons differently. He scored lower in lessons with the oral presentation routine, which became a distinguished, rather than an integrated, part of many observed lessons. The incompatibility of this routine and the lesson focus might have prevented effective student engagement and blurred its relation with the lesson focus.

However, there were also findings that support the theory of differentiated teacher effectiveness (DTE). First, there were days and lessons in which Linus' performance might be less satisfactory. Teaching is a stressful job that affects teachers' observed teaching behaviours at times. "Stability over time is a different, but equally important, form of consistency" (Campbell et al. 2004, p.75). Second, teachers reported that they differentiated their teaching with respect to the characteristics of the students. For example, Lucy were affected much by the class composition. Lucy used different amount of Cantonese (L1) or mixed code in different classes. Charlie stressed that there was a difference in junior and senior form teaching. For example, disruptive behaviours were observed more common in Form 1 classes. Third, teachers might vary in strengths in different underlying dimensions of observed teaching behaviours. For example, Linus could score highly in the dimension *Effective lesson planning* and sometimes in the dimension *Student engagement*, even when he scored average or below in most of the other dimensions. Fourth, the school contexts affect individual teacher effectiveness considerably in the grouping/setting policy, school culture in teacher collaboration, and/or school policies on inclusion and reflective learning. For example, Lucy was an interesting case as she had lower ratings in most underlying dimensions in most lessons, but she also scored highly in most underlying dimensions in lessons of some classes. These results indicated that Lucy's teaching effectiveness was strongly affected by some specific contexts. Lucy was more effective when she could effectively manage her Form 1 class when its class composition had changed. Her performances showed support for the DTE theory, which denies that teacher effectiveness is a generic characteristic of a teacher, but appears to vary with different contexts. However, class composition is not something that a teacher normally can bargain for as it depends on the senior management to determine which class a teacher would teach.

Using the two continuums defining consistency discussed in Section 2.5.4, the ratings of the observed teaching behaviours four teachers which varied in terms of their underlying dimensions and contexts can be plotted graphically as in Figure 7.7.

Figure 7.7: **Locating teachers observed on the two continuums defining consistency**



In this graph, only Sally and Charlie conform closer to the predictions of the GTE theory. Instead of locating at the lower left end along the dotted line as what the GTE theory may predict, both Lucy and Linus appear to locate toward the more extreme regions: all dimensions, single context versus some contexts, single dimension, respectively.

As noted earlier in Section 2.6.4, the DTE account is broader, and thus better, than the GTE one in describing these regions. For example, according to the GTE theory, an ineffective teacher would tend to be ineffective in all aspects of teaching. Linus' teaching behaviours could have conformed to this prediction of the GTE theory, if he had failed to receive high scores in all dimensions in a lesson. However, the GTE theory cannot explain why Linus sometimes also showed a relatively stronger strength in the dimension *Strategies to enhance learning and lesson focus*. In contrast, as the DTE theory considers that a teacher may vary in strength in different aspects of teaching, the strength of Linus seemed to reflect the current school policy to enhance reflective learning in students. This supports the view that a stronger focus on teaching and learning in school would enhance the overall teaching effectiveness in teachers (Mortimore et al., 1988; Muijs et al., 2004). Thus, the case studies support the view that the GTE and DTE are compatible and complementary to each other because the former can predict Sally and Charlie's teaching behaviours but the latter can account for Lucy and Linus' varied teaching behaviours better.

Based on these four case studies, it can be concluded that teachers vary considerably in the consistency of their observed teaching behaviours in this department and school. According to Marzano's (2003) probability model of educational effectiveness (PEE), their inconsistent teaching behaviours might contribute to the fluctuations in departmental teaching effectiveness as summarised in Figure 7.8.

***Figure 7.8:* Teachers' contributions to departmental teaching effectiveness**

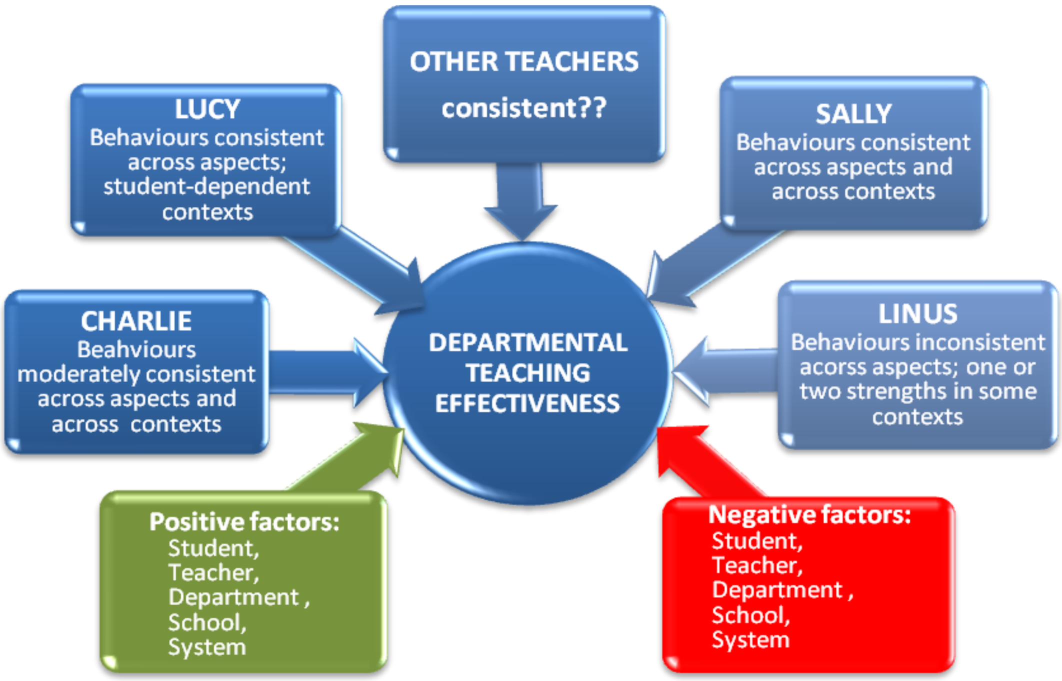


Figure 7.8 also shows that, as indicated in each case summary above, there are both some tentatively positive and negative factors in the teacher background, the teaching practices, the students, the department, the school and the system. All these factors may affect teachers' teaching practices and subsequently their teaching effectiveness. These factors are tentative in their influences, because whether a factor is positive or negative is sometimes hard to determine and may vary with time and context. For example, the

university training and language benchmark test may enhance the quality of the teacher. However, how relevant, as Sally also raised a similar doubt, can these be for assessing teachers' subject knowledge if the curriculum and academic structure in Hong Kong keep changing? Research on teacher knowledge and teacher quality as cited in Section 2.3.2 seemed to produce inconsistent and inconclusive results, suggesting that it might not be appropriate to regard teacher knowledge and qualification as strong predictors of student achievements.

Certainly, these teachers reported high commitment and motivation, but how long can they remain confident in their efficacy when they continue to teach in a school in challenging circumstances with poor learning atmosphere and ethos? Day et al. (2006) found that teachers working in the most challenging contexts were not necessarily those who were less resilient. Muijs et al. (2004) also indicated the research evidence did not support the view that teachers in the challenging contexts could not counter the negative circumstances to make a difference. Negative personal factors and external contexts can become unfavourable impacts on teaching practices and undermine teacher effectiveness. For example, Day et al. (2006) reported cases in which some experienced and once devoted teachers became disillusioned when they faced with aging, illness, family problems, micro-politics in schools, pressure generated by external policies and other problems. The beliefs and perceptions of these teachers about their students became negative, and this seemed to reduce their effectiveness and the quality of their teaching.

In the next chapter, the focus will shift to enrich the findings in Figure 7.8 by characterising the consistency and variation in departmental teaching effectiveness with special attention to the findings of three emerging research questions: (1) in what ways do the observed teaching practices and the lessons of the teachers differ? (2) In what ways do the junior and senior form teachings differ? (3) In what ways do more effective and less effective lessons differ? In addition, interviews with the head of the English department and the principal were used to explore their views regarding factors that affect the teaching practices and teaching effectiveness in the department.

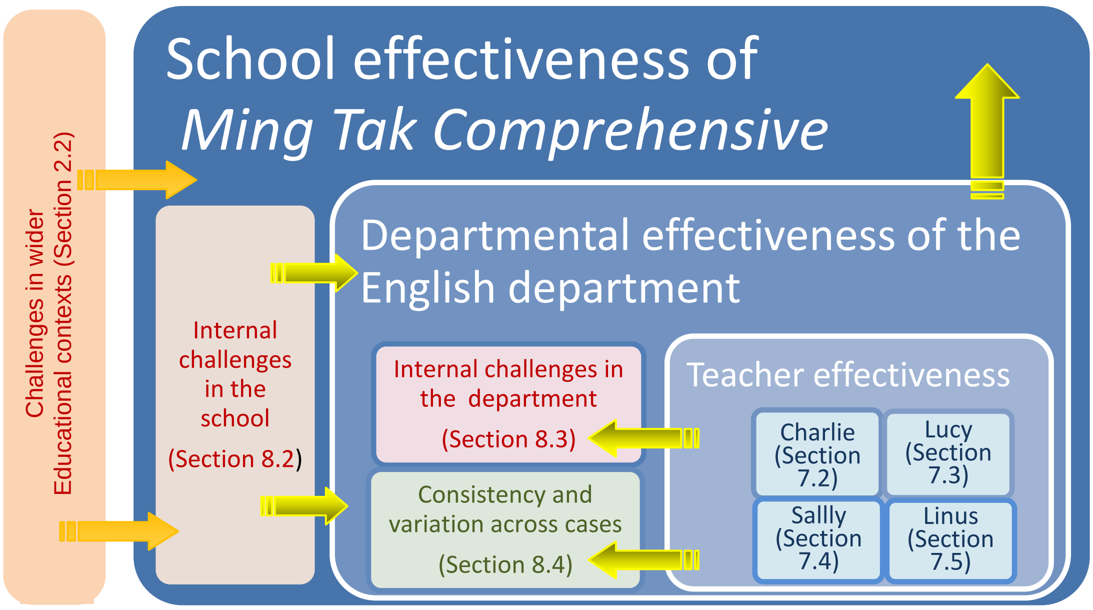
These views increase our understanding of the departmental teaching effectiveness through a focus on issues at the department and school levels.

CHAPTER 8 : CHALLENGES IN THE SCHOOL AND IN THE DEPARTMENT AND THEIR RELATIONS TO CONSISTENCY AND VARIATION IN CLASSROOM PRACTICES IN TEACHERS

8.1 Introduction

In the last chapter, the practices of the four participant case study teachers were examined in four steps of accumulation of evidence. Based on observation and interviews, these case studies examined the unique characteristics in the teaching behaviours of the teachers and their differential teaching effectiveness that may have contributed to the variation in the departmental teaching effectiveness. This chapter will examine the different sources of cross-case evidence to address the departmental effectiveness as depicted in Figure 8.1. As the participants were all ESL teachers of the English department of a school, their differential teaching effectiveness reduces the department's effectiveness and adds to those existing challenges of the school in maintaining its school effectiveness.

Figure 8.1: **Structure of Chapter 8 and relationships with sections in other chapters**



Section 2.2 discusses some of the systemic contextual factors regarded as potential negative impacts on teacher effectiveness (i.e., Hong Kong's unique secondary school places allocation system, the school reforms to narrow gaps between schools, the post-colonisation medium of

instruction policy, and the examination-oriented educational culture). These factors pose challenges to English teaching and learning in general and to Band 3 schools like *Ming Tak Comprehensive* in particular.

In Section 8.2, the school level challenges are explored in the general background of the school and through the semi-structured interview with the school principal. In Section 8.3, several sets of data are presented: the general background about the department, the summary of an unstructured focus group interview with the observed teachers, the semi-structured interview with the head of the English department, and the summary of two unstructured focus group interviews with students of one teacher, Sally. As the senior staff of the school, the principal and the department head reported what they had done in promoting teaching and learning in the school, what factors they perceived as important in affecting the teaching practices and teacher effectiveness and ultimately affecting the student outcomes. Six school-level challenges and three departmental level contradictions are identified respectively in the interview with the school principal and in the interview with the department head. The focus group interview with the teachers has highlighted the pressure of enhancing student outcomes, competitions from other subjects in the school and from external tutorial schools, and the difficulties to integrate cross-level learning targets. The focus group interviews with the students to illuminate their perspectives about learning in school and the English department in particular.

Interviewed teachers considered these questions were important for their teacher development strategies and for peer observations. They were expecting results that would have stronger implications to teacher development and school improvement. Their first concern was to what extent their lessons were comparable and whether the differences were just stylistic. They were not sure how the differences might affect the student outcomes. It was also argued that teaching behaviors had to vary in junior and senior forms because the respective curricula were so different. Finally, teachers were interested in how effective lessons and ineffective lessons could be characterised. Their concerns had brought this research a step forward in case comparison and case integration. The interview contributed to co-

construct knowledge with participants and showed the strength of mixing methods in generating new insights in the dynamic research process. The voice of the participants was not only heard, but also led to further investigations that were also meaningful to them.

In Section 8.4, after presenting the cross-case comparisons of the teachers' classroom events and of their perceptions of teaching and learning in school, issues raised in the unstructured focus group interviews are addressed in three additional research questions to explore the consistency and variation in teachers and their consequences:

- In what ways do the observed teaching practices and the lessons of the teachers differ?
- In what ways does the junior form teaching differ from the senior form one?
- In what ways do more effective and less effective lessons differ?

The quantitative results of these questions are presented to characterise the contrasts in the four teachers' lessons, in their junior and senior form lessons, and in their more effective and less effective lessons.

8.2 The challenges existed at the school level

8.2.1 Further background about *Ming Tak Comprehensive*

In addition to the general background about *Ming Tak Comprehensive* discussed in Section 3.5.2, the interviews with the teachers and the students revealed further challenges of the school. In the two focus group interviews with twelve students, none of them indicated that their parents could help them with their English work and none watched English TV programs at home. This is not surprising if most of the students have one or two parents immigrated to Hong Kong from China in their adulthood after the seventies. Like their academic achievements, students' family background may have an negative impact on their academic self concept.

According to the teaching staff, the proportion of *Ming Tak Comprehensive*'s students who are in the lowest Band in the intake has been increasing over the years. This is an alarming signal to the teachers and the school administrators. With more new-built schools in its catchment area in recent years, the school faces keener competitions from schools in the same district than before. *Ming Tak Comprehensive* has been a CMI school for more than forty years since its foundation. It suffered a major decline in the eighties as students and parents preferred an EMI school. Ironically, according to the school principal, the controversial MOI policy of the government in the past ten years was a blessing for the school because it turned many of its previous EMI competitors nearby into a CMI school like *Ming Tak Comprehensive.* However, starting from the next term, it has to face the challenge that the other CMI schools may try to introduce EMI into some of their curricula for some of their classes. This becomes a challenge to *Ming Tak Comprehensive* if it persists to be a CMI school at the cost of losing its competitiveness under the newly amended MOI policy.

Like many other Band 3 schools, *Ming Tak Comprehensive* has shown a spiral downturn in performance measured by its valued-added results in many subjects including English. These value-added results suggest that it is underperforming according to the Hong Kong average and the average of schools with similar contextual backgrounds[88]. Thus, the decline in intake does not explain its relatively poor performance. To tackle the decline in academic ability in the intakes, *Ming Tak Comprehensive* has been experimenting on different ways to enhance academic results over the years. However, setting and reducing class size remain the most popular strategies for the three major subjects (i.e., Chinese, English and Mathematics). Thus, in the junior years, students are divided into groups of similar abilities and sometimes the weaker groups are smaller, especially in English lessons. The smaller groups are usually called remedial classes, but they use the same textbooks, follow similar teaching schedules, share the same curriculum and take the same assessments as other classes do.

8.2.2 Challenges recognised by the school principal in the interview

In an interview with the school principal, six challenges have been identified that may affect and sustain the teaching effectiveness of the teachers observed in this research.

88 I have not obtained the actual value-added results in the past four years, but the general picture was confirmed in the meetings with the school principal and the English department head.

Challenge One: Building mutual trust between the management and the teaching staff

After spending about six of his eleven years of principalship in *Ming Tak Comprehensive*, Mr. Kwong (a pseudo name for the principal hereafter) has developed a global perspective about the challenges in his school and the place of his school in the education system. Recent research on principals who have been successful in enhancing student outcomes in England reported by Day, Sammons, Leithwood, et al. (2009) has drawn attention to the importance of trust in the relationships between leaders and other staff. Mr. Kwong regarded building mutual trust as his most challenging task in the school:

> [Regarding the question of which is my most challenging responsibility as a principal,] I think it is personnel matters or personnel management. It would be much easier if there are good relationships. And good relationships are built on trust. We need trust in order to make things go. I think it is difficult to establish good relationships and it takes time….Yes, [to build mutual trust] between the management and the teachers.

Mr. Kwong did not want to comment on how successfully he had been on this challenging task, despite his understanding of its importance.

Challenge Two: Sustaining teacher effectiveness through professional development and job satisfaction

Mr. Kwong emphasised the importance of both professional development and job satisfaction in enhancing and sustaining teacher effectiveness and making teachers resilience in face of the ever-challenging environments (see similar findings in Harris, Day & Hatfield, 2003; Harris et al., 2003; Day et al. 2006, 2007):

> Commitment is dependent on motivation and satisfaction. Teaching practices are dependent on professional development and both professional development and job satisfaction are fundamental for making teachers resilient. Motivation and satisfaction is not enough without professional development as it can't guarantee success and make success sustainable.

Mr. Kwong understood his roles in promoting professional development in teachers and offering teachers opportunities to work and take up responsibilities:

This goes back to what I talked about trust. I think teachers need to succeed. They need sense of achievement. I have to give them the opportunities to work -- the opportunities to succeed in their work and duties. Once they can succeed in their work, they will be motivated and committed to their work and seek opportunities to improve their teaching practices through professional development. Therefore, I also see my role particularly important in staff development because I put a lot of effort in the past to render opportunities for professional development through staff development. Of course, teachers also have to get involved in their professional development themselves. Things are changing now. There are no practices that can sustain very long.

His emphasis on the impact of professional development on teaching effectiveness is based on his observation that no practices can sustain very long with students of diverse backgrounds and abilities:

Nowadays schools have to stick with students with diverse abilities. Like our school, just in my first three years, before you left, we still had about half of the students from Band 2, but the majority of Secondary 1 students are from Band 3 now. The training that most teachers received in the old days would only work for better students and they need to learn new strategies in order to cope with the new situations. If they lack professional development, they will not be effective in their teaching and they won't feel they are succeeding. Without feeling success in their teaching, teachers will soon lose their commitment in it.

When asked to describe the key characteristics of his teaching staff and their work, Mr. Kwong mentioned three things. First, he praised his teachers for their willingness to do extra work:

There are disagreements but most teachers are committed to their work and I can see that teachers are willing to do extra to make a difference.

Then, he thought that cooperativeness in teachers had really made the lives of senior managers easier:

For the second characteristic, I would say it is the cooperation with the management. You see that there are disagreements around, but most would still do what the management say and there has not been any real opposition in action.

Finally, he stressed the importance of reflective teaching on improving teachers' work:

> The third would be teachers' self-reflection. I think teachers here would think about their teaching practices and do something about them all the times. Yes, teachers in this school would reflect on their own teaching.

Challenge Three: The depth of teacher-pupil interactions and student outcomes in senior form teaching

Despite all these good qualities, Mr. Kwong recognised that the lack of stimulating teach-pupil interaction may have undermined his teachers' effectiveness:

> I had been in some classroom observations before. The teacher was very good at promoting classroom climate. There were also different activities: direct instruction, videos, group discussions, etc. The classroom process was smooth and the instruction was clear. The teacher had a very good sense of humour and made the lesson very enjoyable. But when the lesson finished, I asked myself what the students were supposed to have learnt. My answer was rather negative as I couldn't recall learning much in the lesson if I were a student. There wasn't much teacher-pupil interaction. I always told the senior form teachers that they had to stimulate the thinking of the students more. We need to ask more challenging questions, not just simple yes or no questions or with ready answers. We have to stimulate students' thinking by asking to think about hypothetical questions and questions about what to do in different scenarios.

Mr. Kwong's comments about the classroom observations clearly revealed his priority regarding different aspects of teaching practices. He emphasised the importance of engaging students in deep learning. His distinction between surface learning and deep learning and his insistence on the latter's role in student outcomes is consistent with research by Biggs and his colleagues (e.g., Beattie, Collins, Briggs, 1997; Briggs & Collis, 1982; Collis & Biggs, 1979). He looked for teacher-student interactions revealed in the depth of questions and feedbacks between them:

> I have teachers that have very poor value-added results over the years, not just one year. So there must be something wrong in the teaching practices. When I did the classroom observations, I could see that there needed to be more teacher-pupil interactions. Teachers need to work on their question skills, for example.

Mr. Kwong expected deep learning in senior form students and thought it would eventually lead to better academic results. His thought was in line with research that stresses the value of interactive whole class teaching (Muijs & Reynolds, 2000). He seemed to be displeased with the academic results and teachers' views which attributed poor academic results to low motivation in students as he justified his claim by pointing to the fact that the senior form students were motivated by competitions in the public examinations and that they rarely engaged in disruptive behaviour:

> There won't be classroom management problems at senior forms. Senior form students are prepared to learn anything from you, if you can really have something solid to teach. They can tell the difference. We still have some students in Form 4 and 5 who have poor motivation, but motivation shouldn't be an issue for Form 6 and 7. You can't tell me that these students with good grades, A or B, when they come in and they get D or E in the A-level just because of poor motivation and nothing to do with teaching practices. Unlike the Form 1 students, we select these Form 6 and 7 students ourselves, so we won't have much excuse for their failures. But for junior forms, the situations are completely different.

Mr. Kwong made it clear that teachers of senior forms were accountable for the academic results of their students as they had more able and motivated students. Foe him, if a teacher has failed to demonstrate that s/he can engage students in deep learning, s/he is just doing an easy job by applying the same teaching practices for junior form students to senior form ones:

> But I have to look at student outcomes of senior forms more carefully. It's not just a matter of survivals. You see, teaching senior forms is supposed to be more difficult and demanding if you want to get good results. But it can also be an easy job if we don't ask for good learning outcomes at the same time. We have the rights and the reasons to do so, because teachers are teaching better students, in smaller class sizes, and that sort of things. I don't want to name the individual teachers, but there are teachers whose value-added results are consistently disappointing and they shouldn't have more rights to teach senior forms.

Challenge Four: The poor learning habits and student outcomes in junior form teaching

Mr. Kwong acknowledged that SEN students and low academic abilities amongst junior form students may make teaching difficult, but his view was that poor learning habits were the main cause of their weaknesses. Thus, these students needed to develop good learning habits in the classroom. Mr.

Kwong indicated that to develop good learning habit is what classroom management is supposed to be about:

> Well, we can't expect them [junior form students] to have challenging and independent thinking, for example. When most can't read and write, our job is to consolidate and fill in what is supposed to be there. To help them catch up. You see, they are behind as they have poor learning habits, not lower IQ that sort of things. There are some SENs, but we can't possibly have 50% SENs. I have read about books about excellence and I can't agree more that good habits breed excellence. Sometimes we have to go back to very basic things like asking students to learn paying attention, listening when others talk, raising their hands when they have questions, that sort of things.

Thus, in expressing his different expectations for teaching practices for junior forms and senior forms, Mr. Kwong was rather realistic and pragmatic:

> Yes, classroom management is most important for junior forms. You can't have proper teacher-pupil interaction without classroom management, right? I have different expectations for junior and senior forms. That's normal. For schools like ours, we have to have different expectations for different forms. We don't have much choice. I can't demand junior forms teachers to give me the kind of student outcomes as I ask for the teachers of senior forms.

Challenge Five: Conflicts in allocation of limited resources to achieve different goals for student outcomes

In dealing with the different needs of junior form and senior form students, Mr. Kwong revealed his preference for the latter in his allocation of human resources:

> When you were here, those panel heads kept coming into my room complaining about that all the best teachers in the school became the members of the Discipline team or the Guidance team or became the Class teachers of Form 1. They said that you had taken away the best workforce to the junior forms, to Form 1 and they couldn't get the best teachers to teach senior forms and produce the results I wanted. These teachers couldn't focus on senior forms as they could be because of their duties in junior forms…. It's a matter of school policy as well as politics. I just can't put all the best teachers for junior forms. It's not fair to these teachers to have the heavy burden to take care of the junior forms and let the less effective teachers to teach senior forms without satisfactory results and make their lives easy. You see, I am not against the idea to pay attention to Form 1 students and I believe that class teachers can make a difference. I had been a class teacher myself most of the time

> before I was a Principal. I know that it makes sense to manage Form 1 students right from the start and that it becomes too late when they are getting to Form 2 or 3. But we also need to have good student outcomes.

Mr. Kwong's comments also highlighted the perceived potential conflicts in achieving different goals with limited resources, among which human resources is the main one. As a school policy, there is nothing inherently inappropriate to put more emphasis on student outcomes of senior form students or those of junior form students. However, Mr. Kwong also recognised that it might be too late to change students' habits when they failed to develop good learning habits in their first year in the school. As a leader of a school, Mr. Kwong had to resolve the perceived conflicts when the limited human resources could not meet the needs of different departments and administrative teams in the school. Mr. Kwong seemed to suggest implicitly that teachers who taught senior forms be more effective and assumed that teaching the junior form students basic learning habit is a heavy burden but may not be more difficult than enhancing student outcomes in the senior forms. Mr. Kwong was not alone in holding this view, when other department heads complained about the strategy to put the best workforce on junior form students.

Challenge Six: Conflicts between the school mission and the competitive market culture in the system

Regarding the school's MOI policy and school mission, Mr. Kwong again expressed his pragmatic approach to the issue. He was clearly worrying about the marketing strategy of the school in its school profile. For him, CMI, despite its deep root in the school mission, was an out-dated strategy as it has lost its appeal to parents and the academically abler prospective students:

> Regarding the MOI issue, I raised the issue in the last staff meeting earlier. I asked the teachers to consider whether we had to be persistent in holding the initial mission of establishing this school, namely, to provide education through Chinese as the main medium of instruction. I think this was a good idea, as I also think that it is better to teach junior forms in mother tongue. However, a problem arises when the school profile is issued, I can't provide any class in English so that the parents of my students and prospective students, at least among

those more able students, would criticize me for that.

Regarding the MOI policy in the school, Mr. Kwong's comments sounded rather pessimistic. When nearby schools have changed flexibly, the pressure to react to the market force certainly is increasing, but switching to EMI also would require extra human resources which he cannot completely control. He needed the support from the senior management as well as from teachers:

> So I take this issue back to the School Management Board meeting as this is not a matter I can decide upon. We need to decide upon whether we insist on being a school that will continue to use Chinese as MOI despite the possibility of losing the appeal to some better students. We have to face the reality of getting more students…. So I would ask teachers whether they can teach in English, whether they are prepared to so. Of course, I can't expect everyone to do be able to do that, but there would be some who can do so. There must be *some* teachers who are able to do so, even though not every one of them can do so. In this way, can't I have *one* class using English as MOI? Or *one* subject in English? So I discussed with Mr. Siu [the Panel Head of Math] about the possibility of teaching Math in English as relatively more symbols are used and word problems are less required in learning Math…. So I raised these issues for him to think about. You see, it can't be true that I can't even have *one* class that can learn Math in English?

In the interview, Mr. Kwong could clearly identify his main role in the school and the strengths of his teaching staff. However, he did not try to account for why the school was obtaining poor value-added academic outcomes and had lost its competitiveness. Mr. Kwong's different expectations of junior form and senior form students suggested his different demands for teachers. This distinction may be misleading because teachers of junior forms and teachers of senior forms are not necessarily two groups of teachers. Many teachers, like the ones in the CVCP sample, teach both junior and senior forms. It is also questionable whether teaching junior forms is a less demanding job, especially when mixed ability in students is more common in junior forms and developing basic learning habits in these students requires teaching skills which probably only exist in more effective teachers. The MOI policy highlights the deep conflict in the goals of education in this school as well as in the Hong Kong education system as a whole. Streaming students by abilities or by MOI highlights different tensions at different levels: at the system level, there are tensions between schools in attracting more academically able students and between the school mission to cater for the needs of academically less able students and the need to obtain academic results for which the school is held accountable to survive in the education market place. At the school level, tensions occur between different subjects in maximising their shares in students' motivation and study time and between departments and teams in their shares in the school resources.

8.3 The challenges existed at the department level

8.3.1 General background about the English department and challenges recognised by the English teachers in the focus group interview

Currently, there are 10 teachers in the English department, including one native speaker English teacher. The turnover rate of the department has been high in recent years. Eight teachers have left the department since Charles became the department head. Teachers who taught other subjects apart from English before now teach English only. Each form has one coordinator to set the teaching schedule and divide the workload in making handouts and setting tests and examination papers for each term. Teachers are required to organise other activities to enhance English learning. These activities include preparing the reading materials for the 20-minute reading session in the morning, setting up extensive reading schemes for junior and senior forms students, running the daily business of the English centre, updating the materials on the two notice boards of the English departments, and preparing students for the annual interschool recital competition and other in-school extra-curricular activities. Informal dialogues on teaching are frequent as teachers' desks are grouped by subjects in the staffroom. There are about five formal departmental meetings annually. Other than the usual agenda, Charles would invite one or two teachers to do professional sharing to promote English teaching practices that teachers have learned in different workshops after each departmental meeting.

As students' academic results in English have been lower than the Hong Kong average most of the time, the principal has questioned the teaching effectiveness of the English teachers. According to the principal, English was found weak not only in the contextualised value-added results, but also weak in the attainment results of Form 3 (equivalent to Year 9 in England). The external circumstance for English teachers and students has become tougher because the students could no longer take a relatively easier paper for English in the HKCE when it was abolished in 2007. Therefore, the main challenge for the English department has always been to raise student outcomes. Since the value-added results by subject are available every year and are reported in the staff meeting, teachers or departments would compare results among themselves. There are always favourable and unfavourable perceptions and attributions about the departments afterwards. Comparing to other departments like Mathematics and Chinese, English department has been underperforming more often. Thus, there are some unfavourable perceptions about the English department among the teachers, questioning their teaching effectiveness.

Since the introduction of a benchmark test for language teachers by the government to enhance professionalism by raising teacher professional qualifications of the language teachers, all the English teachers have taken some professional development courses to help them to meet the new requirements to be an English teacher (see EDB, 2009). According to the English teachers in the focus group interview, enhanced collective teacher quality has positive impacts on teacher effectiveness. One of these impacts is evident in changes in the norms of English teaching practice. For example, it used to be a norm that English was *not* the main teaching subject of teachers who taught English before the benchmark test, but now nearly all English teachers teach English only. It is generally believed by the English teachers that professionalism in English teaching practice would enhance the teaching effectiveness of the English teachers.

According to the teachers, one of their major challenges is the low motivation to learn English in students. It is perceived to be easier for students to achieve better results in other subjects than English. Competitions between subjects exist in students' calculation of return for effort. Comparing to other subjects, English may not be appealing, because students are more motivated to work harder on other subjects in which they can achieve better results. Thus, despite its importance as a required subject for accessing higher education, English is rarely the subject on which students in *Ming Tak Comprehensive* would like to spend time. Among the academically more able students in the senior forms, many think that they should spend more effort on other subjects once they can meet the minimum standard or requirements for English. For those less able ones, the opportunity cost of doing well in English may be even higher because they feel hopeless as they are falling too far behind to have any chance to pass the public examinations. Therefore, they would prefer to spend their time and effort on other subjects that they still have some chance to get better results.

For the senior form English teachers, the pressure of accountability is clear, but they are more anxious to see that many senior form students are prepared to give up English when they compare the little return for their effort. Students expect return more immediate than the school management. They also compare their teaching with that of the private tutorial teachers. For example, very often the senior form teachers feel they have to comply with their demand to teach in L1, despite the official policy set forth by the EDB. They recognised that the ability and proficiency in senior form students vary more than the junior form students as some have just stopped making progress since Form 1. By the time these students are promoted to Form 4 or 5, it is too late for them to do anything. Therefore, the English teachers unanimously agreed that, the most challenging task for the English teachers is to raise the standard of junior form students such that they can meet the expected standard and progress set for them when they reach senior forms. Their comments were consistent with the research findings on the Matthew effect in less effective schools mentioned in Section 2.6.1, but seemed to be at odds with Mr. Kwong's priority of allocating more resources for senior forms.

8.3.2 Perceptions of deep contradictions in the interview with Charlie as the Head of Department

Charlie became the English department head four years ago, when the last department head resigned unexpectedly. Being the department head has been a challenge to Charlie because it was only his fourth year in the school and sixth year in his teaching when he was promoted. He felt he lacked the seniority as well as the experience to head a department, but he became the most qualified person in the department after several more senior and experienced teachers resigned or retired in the past few years. Charlie's interview reveals not just challenges, but some "*contradictions*" that indicate various things done by different stakeholders might be intended to enhance the teaching and learning, but they might eventually undermine learning outcomes instead.

Contradiction One: Mutual trust as an aspiration only

Despite Mr. Kwong's emphasis, mutual trust might exist only conceptually as aspirations in minutes and mission statements, because, as the head of the English department, Charlie sees his main responsibility to be a messenger that passes "orders" from the school management to other teachers in the department. Without trust from the top management, leadership is not shared and there would be no true distributed leadership for the middle management:

> I am the bridge between the school authority and the teachers….
>
> I received the order and bring it back to the teachers for implementation….
>
> My most challenging task is to lead other teachers to implement the annual school policies set by the school management. For example, this year their [the senior management's] main concern is *to promote students' reflection on learning and lesson preparation.* I have to convince other teachers to set this concern as a high priority in their work.

Nevertheless, Charlie is more than just a messenger because he is expected to implement the "orders" which are not readily understood in practical terms. Given his perception of receiving "orders" from above, Charlie does not seem to be a leader in its true sense. The worst is that Charlie is not always confident about their meanings and interpretations. Without mutual trust and ownership, policies are not easily understood:

> It's hard also because I need to interpret the school policy correctly for implementation…

> … I have to ensure that I have interpreted the school policies correctly and translated them into administrable or implementable terms.
>
> … Sometimes, the policy is not easily to implement. For example, the philosophy and objective behind the "Reflection on learning" policy may be understandable, but how can we make students to do that in a lesson? That's something supposed to be in their heads that we can't be sure that they are doing as we tell them to.

As an interpreter between the school authority and the teachers, who speak two different languages, Charlie can only rely on two main strategies:

> I have to set as an example myself [in carrying out the school policy], but mainly I just talk to the teachers.

Feeling without much authority and power in the school, Charlie cannot offer much support to his colleagues except setting himself as an example or keeping a dialogue with them. For the other more powerless individual teachers, who once realise that school policies are meant to be regarded as top-down orders, cooperativeness, as Mr. Kwong also confirmed, will become one of their key characteristics for survival:

> Teachers here are hard-working, cooperative and persistent.

However, cooperativeness may mean very different things to the school authority and a messenger or interpreter like Charlie. For him, it means bottom-up help and support, rather than top-down obligations:

> It [maintaining a dialogue] was not easy when I first became the department head four years ago with so many teachers senior to him around. I was rather fresh at that time. It becomes better now as the newly recruited teachers are experienced but more cooperative and helpful.

Moreover, in a Confucian society that prioritises harmony between agents in the social hierarchy, conflicts may be hidden:

> Teachers may not object to anything openly or in public. They may just do other things underneath, behind your back. So you can't be sure that what you want is implemented.

Without any bottom-up ownership, policies cannot be truly implemented and their effectiveness is undermined. Therefore, it is a contradiction that as a school supposedly practising school-based management, *Ming Tak Comprehensive* lacks the characteristic shift in its management, which moves "from the traditional concentration on maintenance and hierarchy to change, collegiality, teamwork, improvement and effectiveness" (EDB, 2010d). In Charlie's account, despite so much emphasis on mutual trust by the school authority, teachers in *Ming Tak Comprehensive* lacked the sense of efficacy that they can effect change in their school.

Contradiction Two: Professional development without true professionalism

Under the pressure of the accountability policy by the government and Mr. Kwong's emphasis on enhancing teacher effectiveness through professional development, English teachers have not lacked opportunities in professional development training and sharing:

> …. teachers also have to keep themselves update with the pedagogy expected by the EDB….
>
> We have the sharing meetings in which teachers would share what they've learnt in the workshops. There are many workshops around nowadays. We usually do the sharing after the panel meeting, so it would be about five to six times a year. We also have informal sharing during lunch… or other time… It's not a problem. Sharing is important as some workshops like those on the new academic structure are often overbooked.

However, it is not clear to what extent the training workshops, formal and informal professional sharing may have lasting positive impacts on teaching practices. Certainly, Charlie could witness some gradual changes in professional knowledge in his colleagues:

> The workshops on professional development also help these teachers gradually develop a different mindset.

However, he also noted that the limitation of profession development in comparing to the impacts deeply rooted in teachers' upbringing:

> There are still some teachers whose perceptions on teaching and the teacher's work are rather traditional. They are more skeptical of changes and thus less easy to get used to changes.
>
> because teachers do what they think will work for them. They tend to do things in the ways how they were brought up.

Professional development and training is a necessary but not the sufficient condition for enhancing teacher effectiveness:

> … but teachers tend to believe that they have to use stern methods like scolding to make them [junior form students] cooperate. I think this would hurt the teacher-pupil relationship BUT other teachers including the English teachers think that work for them. They aren't teaching at all. Sometimes I can hear what happen in the next classroom when I am teaching in 1A. Just scolding, scolding, scolding, punishing, punishing, punishing…. I think that doesn't work. Teachers use different pedagogy for different forms. I agree with you that teachers seemed to have two personalities like schizophrenics.

If teachers are not teaching but punishing students most of the time, one cannot expect much effectiveness in the junior forms teaching. Charlie suggested that it is a general teaching practice among many teachers, rather than a practice of some English teachers. It should not be seen just as a distinction between the traditional and the liberal approaches to teaching, since research suggest there needs to be a greater emphasis on praise rather than negative feedback (Mortimore et al., 1988) and students of low SES background in challenging contexts are more receptive to praise and recognition (Muijs et al., 2004). When respect and self-discipline are substituted by threat and punishment, praise and care are replaced by scold and indifference, and professional conduct and clients' trust are taken away from professionalism, teaching is detached from its moral base and ideal. There will be no true teacher professionalism if teaching practice is devoid of ethics and mission. It is a contradiction that though so much effort and professional development has been spent on enhancing teacher effectiveness, true professionalism seems to be at risk.

Contradiction Three: Contradictory expectations

One must wonder why some teachers may have developed such a "schizophrenic" teaching approach to their students as Charlie agreed *the same teachers developed different approaches to different forms*:

> I cannot see much difference between teachers but they have different practices for different forms. They adopt a friendlier approach to senior form students, but more authoritative approach to junior form students.

The truth may lie in the deep contradictory expectations for these students. Like Mr. Kwong said, he did not expect the same kinds of student outcomes for these two groups of students. For junior forms students, expectations are often limited to good discipline and learning habits, but for senior forms students, excellence in public examinations is the only indicator of effectiveness that matters for teachers and the only indicator of achievement that is valued for students. What Mr. Kwong's comment about holding teachers accountable for student outcomes is clearly something that the senior form teachers should not neglect:

> No matter what [teaching] methods they [teachers] use, we concern only their outcomes. So senior form teaching is exam-oriented, not much difference among them [the observed teachers]. But this year, I am experimenting with more interactive student-led activities in senior forms…
>
> Whatever method is not important, the only student outcomes they want is examination results both internal and external. The relationship between the two [internal and external exams] are high and consistent and predictive. So teachers need good outcomes in both.

Charlie is not very confident in the effectiveness of a student-centered approach for senior forms, because students are expecting the opposite, a teacher-led, exam-oriented teaching approach from him:

> Students are passive and prefer that way. They just want to listen and more teacher inputs [direct instruction], more teaching on sentence patterns, vocabulary that kinds of things, because they don't have a solid foundation.

In Charlie's view, this also explains why senior forms students go to private tutorial class where a passive exam-oriented approach is adopted, but he maintained an opened attitude toward the shadow system and positively regarded their impacts on the teaching practices on the English teachers:

> Students [in senior forms] have their own right to choose their own way to learn. Yes, they [private tutorial classes] are much more exam-oriented. Sometimes, this would affect our teaching as well. Students ask you to do a lot of exam-stuff, so we have to change.

Charlie thought different expectations exist in the aims put forward by the education authority as well as in our society:

> The EDB are pushing the contradictory aims. HK people don't need English for their lives for survival. The whole society is contradictory. They told students [in junior forms] that they need it but in fact they don't.

> Learning atmosphere here is poor. Students are passive, unmotivated, especially in junior forms…. I think English is the least interesting subject for them. Their perception is that they don't need English and it is not important in a CMI school. The junior form students generally don't realize the importance of English in the society. They just don't have that kind of concepts or perceptions. Motivation is stronger only in Form 5 and 7. Students in Form 4 or 6 just don't feel the urgency of the public examinations.
>
> Examination is the only motivator. They study for the examinations. It is part of the Hong Kong culture.

In an examination-oriented culture like Hong Kong's, students study subjects for the examinations when they feel the urgency to prepare for them. The whole society, including the EDB, is contradictory, in the sense that while its expectation for junior forms students' English proficiency is low, its expectation for these students' achievements in the same standardised public exams is high. For junior form students, English has no immediate survival value and public examinations are the realties far away. When students go to a CMI school, they receive a much lower expectation of the need to excel in English, as English is not needed in their daily lives and curricula. Thus, the contradiction is that though there is a wide gap between the low expectation for junior form students and the high expectation for senior form students, English teachers are expected to minimise it in a few years' time. Charlie expressed the frustration in teaching English in CMI schools:

> Students rely on L1 too much… request teachers to teach in L1, but exposure to English is much higher and richer in EMI schools. The [English learning] environment is harsh in CMI schools. Students may only have primary 1 level when they come in, so they may progress to primary six if not primary two when they graduated in Form 5, but they are expected to have a near Form 6 proficiency in the exams. These obstacles lower [teachers] satisfaction.

Surprisingly, it is *not* the accountability of senior form students' academic results that has troubled the teachers more, as it is still somewhat understood and expected. Rather, Charlie and other teachers' senses of efficacy in teaching junior form students are ironically lower after their trying everything to arouse their interests in English:

> We change our pedagogy and our curriculum, shorten the syllabus, simplify the materials, use simple textbooks, and amend the materials to enhance their motivation….. we do a lot of work to tailor-made the materials to suit their needs and language proficiency. We did a lot of things like what research said about motivation…. intrinsic, extrinsic, instrumental motivation… we try everything, but we just can't see absolutely how effective it is. They don't seem to treasure the encouragement… no use at all.
>
> … We add in more lively into our teaching like language arts, interactive activities…and we can see students [junior forms] like it. But not obvious in the results in task based curriculum and reading.
>
> Students cannot transfer their knowledge e.g. from one poem to another unseen poem.

The adjustments that the English teachers have do not seem to have narrowed the gap between the actual and the expected proficiencies in students and the Matthew effect in learning English seems to prevail in *Ming Tak Comprehensive*.

Charlie was also disappointed with the lack of sufficient support in the system and in *Ming Tak Comprehensive*. There is another frustrating contradiction lying in the high aims set forth for English teachers to achieve and the limited amount of support that they expect to receive:

> Regarding the new academic structure, we know very little because our applications for the workshop have been rejected many times. We are not sure about the requirements.
>
> Everyone in the school knows that English is very important to the students as they have to pass the public examinations, but promoting English learning has never been set in a priority as high as it should be. You see, the English department has not received the due support that one may find in other schools. In some schools, English teachers have no other administrative work, but never in ours. In terms of resources, English teachers perform the similar amount of administrative duties as other teachers and the similar amount of teaching workload as other subject teachers. The teaching workload has decreased over the years, but that for the English teachers is not exceptionally low.

Like the case for the MOI policy, the government has not provided sufficient information and support to the frontline teachers to overcome their anxieties over major policy changes. Although the government does provide some extra human resources to CMI schools, there is no follow up to see how individual schools are allocating these resources to ensure the English teachers and the students are benefited. Charlie has the rights to appeal, when he cannot see whether these resources are allocated to the English department and truly reflected in a lower amount of teaching and administrative workload for English teachers. Thus, the deep conflicts remain latent in the school and in the system.

8.3.3 Summary of focus group interviews with students of two classes of Sally

Several important findings emerged in the two unstructured focus group interviews with twelve Form 2 and 3 students. First, contrary to Charlie's negative perception, students expressed a high regards for the interactive approach of their teachers, but worried that they were not catching up fast enough to meet the public exam requirements.

Second, in contrast with the common view of the teachers interviewed, Form 3 students seemed particularly aware of their impacts on their life chances, though the public exams were still realties in distant, because most of them had siblings or relatives who were facing or went through these exams. All students interviewed were anxious about the school work but did not know how to cope with the pressures and the limited strategies available.

Third, the interviewed students were aware of the poor school image in the community, but found it somewhat misrepresenting the realty. For example, Form 3 students declared that they felt like they had just started learning English after they entered *Ming Tak Comprehensive*. They said their English learning in the primary school was full of dictation, copying, and fill-in-the blank exercises. They rarely had chances to speak and learn grammar as systematically and integratively as they were doing in the secondary school.

Fourth, regarding the distinction between effective and ineffective teaching, these students said they could tell the differences after being taught for so many years. These students unanimously agreed that they could distinguish effective teachers and effective teaching and would learn better in the lessons of these teachers. They said there were always effective and ineffective teachers in different subjects, but they had no idea about which subject department was better because they did not normally link teachers with their departments. For them, the teaching and approachability of a teacher were more important than the subject s/he taught.

Fifth, the interviewed students declared that they were more willing to work harder for teachers whom they liked. This is consistent with the notion that a strong teacher-pupil relationship would enhance learning. However, they generally liked mathematics more because there were always some definite answers and, more importantly, they knew how to get these answers even without the teachers' guidance.

Sixth, when asked the causes of their academic outcomes, most students generally attributed about 70% to their own abilities and efforts and 30% to teaching. This proportion seemed to be close to Marzano's (2003) estimation of variance in student academic outcomes that can be attributed to teacher effects[89]. This is also consistent with Gao and Watkins' (2001) claim that students tend to share the responsibility of teaching effectiveness. Therefore, the challenge for the English department is to build on these positive student motivation and perception and prepare these students fast enough to an exam-oriented curriculum that most of them have not yet prepared for.

8.4 Consistency and variation in teachers and their impacts

8.4.1 Cross-case comparison of the four teachers' observed teaching practices in classroom events

According to Marzano's (2003) probability model of educational effectiveness (PEE), large differential teacher effectiveness tends to undermine the departmental effectiveness and weak departmental effectiveness among different departments in turn weakens the overall school effectiveness. This is likely to indicate a lack of instructional leadership. As illustrated in Section 2.5.3, large differential teacher effectiveness in a department means that a student will be more likely to do poorly (i.e., in a lower percentile) in any standardised public exams. Based on the observed

89 According to Marzano (2003), the teacher effect was estimated to be about 40% in most western countries where the school effect was found much lower than Hong Kong ,as Lam et al. (2000) found.

classroom events, five aspects of teaching practices of the case study teachers were identified and compared below.

Use of L1 and mixed code

In general, Charlie and Sally instructed in L2 over 90% of the time and their common form of mixed code was L1 terms embedded in L2 speech. Mixed code was frequent in Lucy and Linus' lessons, taking up about 40-50% of the instructions. However, Linus usually repeated L2 instructions in L1 again, while Lucy spoke in L1 with L2 terms often, especially in senior form lessons. The amount of L1 and mixed code varied with contexts and did not seem to be related to the year group of the class. According to the teachers, they tended to use L1 or mixed code when they found students might not understand them in English. Teachers late expressed in the interviews the dilemma to choose between their intent to ensure better understanding and feedback by students and their obligation to use English exclusively in class in accordance with the government's MOI policy.

Use of teaching materials and facilities

All teachers used handouts to supplement materials in the textbooks. However, it was quite common that students forgot to get their handouts and textbooks ready before the lessons. Lucy was most organised in her presentation of materials on the blackboard. Charlie always tried to help other students in the oral presentation routine by taking notes on the blackboard. Lucy, Sally and Linus used the LCD projectors more often, but Sally used these more interactively and encountered fewer technical problems.

Use of praise and punishment

All teachers praised students, but negative feedbacks outnumbered the praises when students did not act appropriately in Lucy and Linus's classes. Sally often referred back to the rules she set and students had to stand up whenever they forgot to bring their textbooks or did not finish their assignments. Punishment only occurred in the junior forms and students were usually asked to stand up in their seats or at the back of the classroom. However, Lucy and Linus used reprimands as their main strategy of classroom management in junior form classes.

Teacher-student interaction

Sally did not overrun her lessons and let her students relax for a couple of minutes before the bell rang. In contrast, Lucy overran more often, especially in lessons before the recess and lunchtime. Sally and Charlie would pay attention to and interact with students sitting at different positions in the classroom, while Lucy and Linus focused more on students sitting around them. Lucy and Linus used a microphone all the time, so they often staged at the front of the class and hardly moved around in the classroom[90]. Charlie moved around most often and he did it not only because there was more individual and group work in his lessons, but also because he did not use the microphone. The voice presented through the microphone was audible but unpleasant and unnatural to hear, especially when the teacher used mixed code in their teaching or when s/he was giving negative feedbacks.

Senior Form teaching

Lucy, Sally and Linus all taught Form 5 classes, although their classes varied in terms of their academic abilities and aptitudes.[91] It should be noted that it was rather difficult to judge students' learning in the senior forms (i.e., Forms 4 and 5, equivalent to GCSE classes in England) because they appeared rather passive in class most of the time. Classmates tended to see one another as in-class competitors. Most students were not confident in their abilities and were afraid of losing face in public, so they were reluctant to ask or answer questions. The curriculum was exam-oriented as the school had to prepare the students in accordance with the exam syllabus. The general impression of senior forms teaching was consistent with the findings by Cheng (1998; 2005), who investigated extensively the washback effect of

---

90 The plug for microphone is located on the front wall. There are hand-free microphones available in the market, but none of the teachers observed used them.

91 These classes were streamed into arts, business, and science majors when they were in Form 4. They took different options in the Hong Kong Certificate Examination (a public exam equivalent to GCSE). Based on their Form 3 term results in traditional science subjects like Physics, Chemistry, and Biology and traditional humanities subjects like History and Chinese History students were streamed into classes where they showed stronger aptitude. In general, students with stronger academic abilities tended to opt for one of the two science classes but they would go to Arts class if they preferred. This year, Lucy was teaching the better Science class, Sally the Arts class, Linus the Business class. Although the entry requirements to the Business-related courses in the universities are keen, students majoring Business at the HKCEE level do not enjoy any advantage. Instead, they are usually academically the weakest in the senior forms.

the public exam and declared that positive washback effect were slow or would not appear if the expected pedagogy did not occur.

The following comparisons of Form 5 teachers indicate some typical variations existed in their teaching practices. For example, in Lucy's Lesson 13, the typical teaching pattern was an individual exercise after the teacher's formal presentation, but she rarely insisted on soliciting expected feedbacks from students:

> The teacher tells students that they have to grasp the mood of a poem in order to appreciate a poem. The teacher goes on to analyse the poem line by line. The teacher does ask some questions but does not expect students to be able to give correct answer because she gives the answers sooner than students' responses.
>
> 9:41 The teacher starts to teach students briefly how to tackle Qs in the exercise in L1 and students start working on their own. The teacher walks around to check students' progress. Meanwhile, students keep asking the teacher for help as it seems that the exercise is rather difficult for them.
>
> 9:48 The class is getting noisy as many students try to discuss the task with other students around.

At times, Lucy applied the similar classroom management strategies (i.e., giving negative feedbacks, complaining and scolding) to Form 1 classes when the students were not motivated and did not work accordingly as she expected. Despite these strategies, she could not engage these senior students in their work:

> 10:59 The teacher gets mad again and scolds students in L1: You know everything and don't need my teaching? I don't need to teach. It's all wasting time. We are wasting our time. I can live through it. I can give you the answer in print and you can bring it back home. I don't have to waste my time. You can get better marks and you have your responsibility and you will know (it's wrong) when you grow up.
>
> 11:02 The teacher continues to teach but 2 or 3 girls bury their heads on their desks. The teacher notices more students bury their heads on desks and yells at them in L1: *Have you learnt it already? Don't need to learn anymore? If you don't cooperate, we can do things properly.... all you do is just to make yourselves happy*.
>
> 11:05 The teacher asks students to turn their books to Ex 2 and starts to teach the difference between *have* and *has*. Soon the teacher jumps to the question: *why we need doesn't*

for the negative form and why we need to add do. A girl is still burying her head on the desk.

For the same topic in her Lesson 8, Sally's presentation was a little longer, slower in pace, more detailed and structured with students' prior understanding in mind:

> 12:04 Using the LCD projector, the teacher continues to explain keywords underlined. Some words are common like *mirror, look*, but some are less common in L2 learners' vocabulary like *scrawny, flaw, hassles, gruffly, brutal, drag*. The teacher does not give meanings directly but asks to guess from the poem. For example, The teacher asks students to guess the meaning of *hassles*. She uses simpler L2 equivalents for explanations like *problems, troubles unhappiness*.
>
> The teacher also tries to show how words are linked to show meanings like *movies, meals, work part-time, tutorials* all related to money, but that is not the main concern of a teen because it is love is what a teen concerns as in the line *All for a kiss*.
>
> The teacher shows how some adjectives can make description stronger like *awfully bad* to mean *very bad*.
>
> The teacher shows how words may derive from others; e.g., *rosy* from *rose* to indicate a bright future; *uni* as abbreviated form of *university*; *bores* as a noun to say someone who is *boring.*
>
> Till the end of the lesson (12:20), The teacher explains how words are used by the poet and asks students to indicate words may not be clear to them. Most students are quite quiet, but attentive most of the time. This is not easy for them to analyse a poem.

In Lesson 4, when Linus was teaching reading comprehension, he presented the materials by highlighting do's and don'ts but in a pace that most students were not catching up. For example, he gave them some tasks to work with, but like Lucy, Linus did not check their understanding through feedbacks before showing them answers:

> **10:50** The bell rings. The teacher asks students to read questions carefully in 5 minutes and then discuss with their neighbours.
>
> 10:56 Students start discussing this time, but soon the teacher shows the answers on the projector screen and explains the answers without checking their answers. Many students are whispering, so it becomes difficult to hear clearly what the teacher says at the back of the classroom. The teacher teaches an important skill, but some students do not pay enough

attention and three students fall asleep.

The teacher asks students to use their own words when they give their answers. For example, if the question asks for a noun as suggested by the question-word what, then they have to change the adjective in the text into a noun: *repetitive +motion → repetition*. The teacher tells students that many questions are not straightforward but require changing the parts of speech from words extracted from the text.

11:54 The teacher continues to show the answers on the screen. The teacher moves to Question 21, but the resolution of the LCD projector makes it difficult to read at the back of the classroom.

12:20 Towards the lesson end, the teacher reminds students it is important to follow exactly the requirement of the task. The teacher points out that it is wrong to write more than it is needed as marks are not given for extra work.

Though these teachers taught similar course materials, their teaching strategies and skills were different. Because the teaching quality in the department varies with differences in observed teaching effectiveness is expected, then some worse scenarios in Marzano's (2003) PEE model may occur.

8.4.2 Cross-case comparison of the four teachers' perceptions of teaching and learning in school

All four teachers showed strong ownership of their teaching. Charlie stressed his enjoyment in promoting interactions with his students, while Sally found the junior form students more responsive, energetic, and willing to voice their opinions. Their strong sensibility for their students were not limited to their abilities and difficulties in English learning, but also extended to their students' psychological well-being. While they recognised the low self-concepts, learned helplessness, and eagerness for recognition and praise in students, their teaching goals and strategies to help students were different. Regarding the main goal of teaching, Charlie attempted to ensure that students understand him and are benefited in the exams. Sally intended her teaching to make students love English and retain the ability and interest in English even after their graduation. In terms of teaching strategies, Sally relied on a relaxed and supportive learning atmosphere and a positive teacher-pupil relationship inside and outside classroom. In contrast, Linus stressed the importance of consolidation in vocabulary, reading, and grammar as well as reflective learning in homework.

According to Mr. Kwong, teachers in the school were cooperative and reflective. His perception was confirmed as Lucy noted the importance of collegial support and Linus emphasised the impact of a supportive school policy on him. However, Sally acknowledged their *reflective practices were rather limited in scope* and might not be sufficient to enhance teaching effectiveness. Rather, she thought they needed to think in more global terms and address the direction of the modules they taught, rather than specific teaching strategies of a lesson. What Sally suggested might be similar to what Marzano (2003) refers to as a guaranteed and viable curriculum. If the proficiency of the intake was low and the standard of the public exam was uncompromisingly high, teachers would then have to ensure the cross-year-group curricula were integrated and progressed accordingly. That is, teachers have to cover, sequence, and organise all the essential content in such a way that students have ample opportunity to learn it (Marzano, 2003). Similarly, Linus raised doubts about the *limited functionality of their assessments*. Charlie also noted that the *learning atmosphere in the school was not high*, particularly in the junior forms. Thus, it seemed that both assessment and learning goals in *Ming Tak Comprehensive* have not been playing their expected functions in enhancing student learning. According to Marzano (2003), challenging goal and effective feedback are complementary to each other such that students would have the pressure to achieve the established academic goals, while receiving adequate and timely feedback on specific knowledge and skills.

8.4.3 Characterising the four teachers by the ISTOF and QoT underlying dimensions and the global indicators of overall teaching effectiveness

In contrast to what Charlie stated, he believed in his interview comments. Section 8.3 has revealed that there may be clear and subtle variation between the consistency and variation patterns in the four teachers' observed teaching behaviours. Accordingly, planned ANOVA and post hoc comparisons were performed to explore similarities and contrasts between the four teachers' scores on the systematic observation schedules.

First, in the one-way between-subject ANOVA that compared the mean scores of the various underlying dimensions of teacher behaviours observed using the ISTOF and QoT instruments and the two global indicators of overall teaching effectiveness. As shown in Table 8.1, all the planned between-group comparisons were statistically significant. The adjusted *R* squared in the last column indicates how much variance can be explained by the corresponding underlying dimension or global indicator. Though consistent with the case analysis results in the last section, these results cannot illustrate subtle differences between the four teachers.

Table 8.1: **Differences between teachers as indicated in the underlying dimensions and the global indicators of overall teaching effectiveness**

Underlying dimension/Global indicator	F or Brown-Forsythe F	df1	df2	Sig.	Adj. R^2
Meta-cognitive skills teaching	11.97	3	72	***.000****	.31
Classroom management and climate*	33.69	3	43.19	***.000****	.55
Differentiation and support	15.98	3	72	***.000****	.38
Clarity and logic of presentation	13.77	3	72	***.000****	.34
Student engagement	9.42	3	72	***.000****	.25
Strategies to enhance learning and lesson focus	9.85	3	72	***.000****	.26
Integrated class management and climate*	23.66	3	53.67	***.000****	.45
Structured teaching skills	18.08	3	72	***.000****	.41
Effective class/lesson planning*	3.76	3	37.62	***.019****	.11
Overall judgment of quality of teaching*	9.28	3	61.45	***.000****	.23
Good individual involvement by the pupils*	18.19	3	64.48	***.000****	.39

Note: Brown-Forsythe F statistic was used instead of the normal F-test as Levene test indicated the variances of five dimensions/indicators shown with an asterisk were heterogeneous and the group sizes were unequal.

Thus, *post hoc* tests were performed to highlight contrasts between teachers in multiple comparisons and the results are summarized in Table 8.2 below.

Table 8.2: **Subtle contrasts evident in the multiple comparisons of teachers on different underlying dimensions and the global indicators of overall teacher effectiveness**

Underlying dimension/Global indicator	Contrasts found statistically significant at .05
Meta-cognitive skills teaching*	Lucy, Linus < Sally, Charlie
Classroom management and climate*	Lucy, Linus < Sally, Charlie
Differentiation and support	Lucy < Linus, Charlie < Sally
Clarity and logic of presentation*	Lucy, Linus < Sally, Charlie
Student engagement	Lucy, Linus < Charlie < Sally
Strategies to enhance learning and lesson focus	Lucy < Linus, Sally; Lucy, Charlie < Sally
Integrated class management and climate*	Lucy, Linus < Sally, Charlie
Structured teaching skills*	Lucy, Linus < Sally, Charlie
Effective class/lesson planning	Charlie < Lucy, Linus < Sally
Overall judgment of quality of teaching	Linus, Lucy < Charlie < Sally
Good individual involvement by the pupils*	Lucy, Linus < Sally, Charlie

Note: Given that the group sample sizes were unequal, Games-Howell was used instead of Tukey HSD or Gabriel.

These results show that in five underlying dimensions and the global judgment of overall teaching quality (indicated by an asterisk in the first column), Linus and Lucy are a pair who scored significantly lower than both Sally and Charlie. However, there are other dimensions and the global indicator of good individual involvement by the pupils that showed different statistically significant contrasts. For example, Linus had significantly better scores than Lucy in the two dimensions *Differentiation and support* and *Strategies to enhance learning and lesson focus*. Similarly, Sally was found to score significantly higher than Charlie in the dimensions *Differentiation and supports, Student engagement*, and *Overall judgment of teaching quality.* Charlie scored lower than other teachers in the dimension *Effective class/lesson planning*, as suggested in Section 8.3.1. These are subtle differences in the patterns of teachers' observed behaviours that may be not easily generalised in non-statistical analyses, although the qualitative observations have provided supportive evidence of such variation.

8.4.4 Characterising the lessons of the four teachers: grouping by similarities and contrasts

The similarities and contrasts between teachers can also be revealed in a simultaneous descriptive discriminant analysis, which compared the lessons of these teachers by a selected set of characteristics as independent variables that can be combined to explain the major differences among the teachers (Huberty, 1994). The discriminant function can be used to "describe the predictive relationship of the independent variables to the dependent variable" in a general model (Meyers, Gamst & Guarino, 2006, p.258). In the present case, the dependent variable refers to *the teacher* as a category with four groups representing the four teachers. The underlying dimensions of the observed teachers' behaviours using ISTOF and QoT as instruments were used as independent or predictive variables as they were in the multiple regression analyses in Section 7.5.

In order to compare the relative explanatory power of the underlying dimensions identified in the empirical model of each instrument, three sets of predictors were used: ISTOF underlying dimensions only, QoT dimensions only, and underlying dimensions of both instruments. By means of some discriminant functions defined by the underlying dimensions selected as a set of predictors, lessons classified or predicted to be as distinct groups could be compared with the actual counts of lessons by teachers. The more lessons a predictor set could correctly classify the lessons by teachers in their original groups, the better descriptive power it would have.

The classification results in Table 8.3 indicate that *the set combining the underlying dimensions of both instruments has the highest explanatory power* as 78.9% of original grouped cases could be correctly classified, followed by 69.7% by the ISTOF underlying dimension set, and 61.8% by the QoT underlying dimension set.

Table 8.3: **Relative successful classification rates of discriminant function analyses using different predictor sets**

Teacher	ISTOF underlying dimensions as predictors only		QoT underlying dimensions as predictors only		Both ISTOF & QoT underlying dimensions as predictors	
	Percent correctly classified	Predicted/ actual count	Percent correctly classified	Predicted/ actual count	Percent correctly classified	Predicted/ actual count
Charlie	60.0%	9/15	80.0%	12/15	**86.7%**	13/15
Lucy	73.9%	17/23	**82.6%**	19/23	69.6%	16/23
Sally	85.5%	17/20	82.6%	16/20	**100.0%**	20/20
Linus	55.6%	10/18	**0.0%**	0/18	**61.7%**	11/18
Total	69.7%	53/76	61.8%	57/76	**78.9%**	60/76

In particular, the discriminant functions defined by the ISTOF dimensions could classify Charlie and Linus' lessons not as accurately as it did for Lucy and Sally's lessons. Similarly, the discriminant functions defined by the QoT dimensions could correctly classify more of Lucy's lessons, but failed to classify any of Linus' lessons. In contrast, the combined dimension set can classify lessons of Charlie, Sally, and Linus at the highest successful rates. For each predictor set, three discriminant functions were identified, but the third discriminant function of both ISTOF and QoT predictor sets were found insignificant[92]. Therefore, only two discriminant functions are shown in for these two predictor sets in Table 8.4. As the discriminant functions are linear composite of the underlying dimensions, they can be understood as the latent variates that distinguish these underlying dimensions.

Table 8.4: **The major contrasts between teachers as defined by the discriminant functions formed by the underlying dimensions identified using the ISTOF and QoT as instruments**

Predictor Sets	Discrim-inant function	Underlying dimensions defining the discriminant function	Correlation coefficient	Standard-ised coefficient	Squared canonical correlation
ISTOF underlying dimensions only	**First**	Classroom management and climate	0.86	.88	0.63
		Clarity and logic of presentation	0.54	.53	
		Meta-cognitive skills teaching	0.53	.24	
		Student engagement	0.44	-.33	
	Second	Strategies to enhance learning and lesson focus	0.86	1.09	0.32
		Differentiation and support	0.58	.38	
QoT underlying dimensions only	**First**	Integrated class management and climate	0.83	.96	0.55
	Second	Effective class/lesson planning	0.76	.63	0.23
		Structured teaching skills	0.72	.12	
Combined set of both ISTOF & QoT underlying dimensions	**First**	Classroom management and climate	0.66	.31	0.68
		Integrated class management and climate	0.61	1.15	
		Structured teaching skills	0.50	.10	
		Meta-cognitive skills teaching	0.45	.39	
	Second	Differentiation and support	0.66	.74	0.51
		Clarity and logic of presentation	0.47	.15	
		Student engagement	0.45	.46	
		Effective class/lesson planning	0.40	.13	
	Third	Strategies to enhance learning and lesson focus	0.57	1.44	0.24

Note: For each discriminant function, there was a set of correlation coefficients and standardised coefficients. For limited space here, the correlation coefficients and standardised coefficients listed here are only those coefficients of the discriminant functions with which the underlying dimensions most strongly associated.

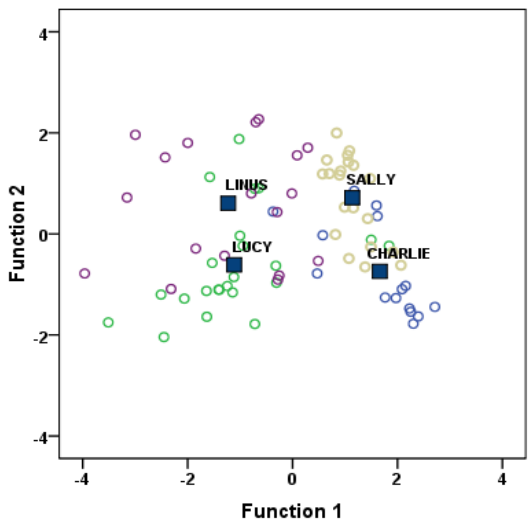
As shown in Figure 8.2 below, the discriminant functions of each predictor set have two illustrative functions: to group the lessons of each

92 Wilks'Lambda for each predictor set is summarised in the following table:

Predictor set	Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig. level
ISTOF underlying dimensions only	1 through 3	.22	105.45	18	.000*
	2 through 3	.60	35.75	10	.000*
	3	.88	8.88	4	.064
QoT underlying dimensions only	1 through 3	.34	76.36	9	.000*
	2 through 3	.77	18.67	4	.001*
	3	1.00	.02	1	.878
Combined set of both ISTOF & QoT underlying dimensions	1 through 3	.12	145.17	27	.000*
	2 through 3	.38	67.25	16	.000*
	3	.76	18.62	7	.009*

teacher around his/her centroid and to displace the lessons of different teachers apart from each other. Each centroid is the teacher's group average of the weighted linear composite making up the discriminant function (i.e., the latent variate in z-score), indicating the relative separation between the groups.

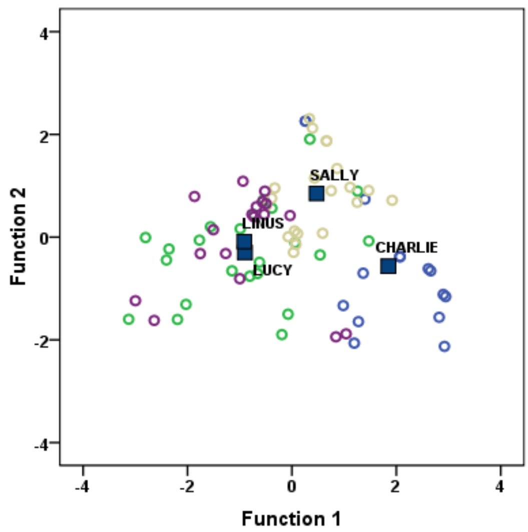
Figure 8.2: **Similarities among lessons of teachers grouped by the first two discriminant functions defined by each of the three predictor sets in the all-group scatter plot.**



Canonical discriminant functions of ISTOF underlying dimensions *only*

Function 1: *Classroom Management & Climate*, Clarity & Logic of Presentation, Meta-Cognitive Skills Teaching, and Student Engagement

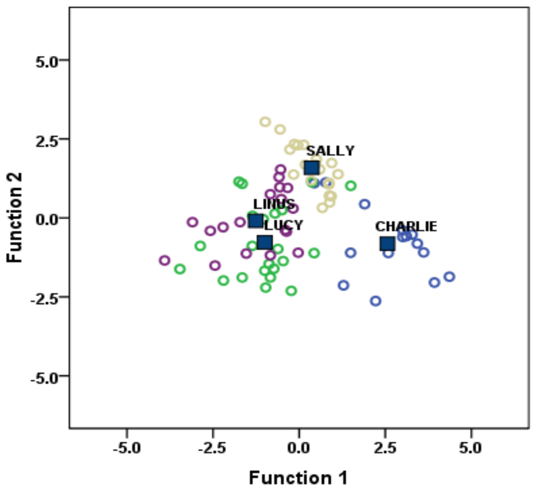
Function 2: *Strategies to enhance learning and lesson focus* and Differentiation & Support



Canonical discriminant functions QoT underlying dimensions *only*

Function 1: Integrated Class Management & Climate

Function 2: Effective Class/Lesson Planning and Structured Teaching Skills



TEACHERS:
- CHARLIE
- LUCY
- SALLY
- LINUS
- Group Centroid

Canonical discriminant functions of combined set of *both* ISTOF & QoT underlying dimensions

Function1: ***Classroom Management & Climate***, Integrated Class Management & Climate, Structured Teaching Skills, and Meta-Cognitive Skills Teaching

Function 2: *Differentiation & Support*, Clarity & Logic of Presentation, Student Engagement and Effective Class/Lesson Planning

Function 3: Strategies to enhance learning and lesson focus (not shown in the graph on the left)

In Figure 8.2, it is clear that lessons of different teachers are more displaced apart in the first function than in the second function as lessons are more dispersed horizontally than vertically. This actually indicates more variance is found in the first function than in the second function and is reflected in the different squared canonical correlations in the sixth column of Table 8.4, which indicate the amount of variance accounted for by the respective discriminant functions. As the correlation coefficients in the fourth column in Table 8.4 indicate the correlations between the underlying dimensions and their corresponding discriminant function, it becomes clear that *the most important underlying dimension of the first function of each predictor set is related to classroom management and climate*. It is in this aspect that teachers differ most.

As shown in Figure 8.2, the first function distinguishes Charlie and Sally from Lucy and Linus. In contrast, the second function of the QoT predictor set and that of the combined predictor set distinguishes Sally from other teachers. *The key underlying dimension strongly associated with this second discriminant function is the dimension Differentiation and support*. This result is consistent with the result of the *post hoc* contrast on this underlying dimension indicated in Table 8.4. Though Figure 8.2 cannot show the third function of the combined predict set, its distinctive function is indicated in the second function of the ISTOF predictor set because both have *Strategies to enhance learning and lesson focus* as its most important underlying dimension. Thus, *this third function in particular distinguishes Linus from other teachers*.

8.4.5 Characterising the junior and senior form teaching

Both Mr. Kwong and Charlie stated that the junior and senior form teaching differed in *Ming Tak Comprehensive*. Charlie insisted that classroom management is the prerequisite of effective teaching in the junior forms though at times he disagreed with the authoritative approach of some teachers. Mr. Kwong expressed his lower expectation of junior form academic outcomes. He also firmly believed that meta-cognitive skills was crucial in enhancing the academic outcomes as students can have deep learning if they can master these skills. As Charlie correctly noted that most teachers, except himself, taught both junior and senior forms, so the distinction between junior and senior form teaching characteristics should not be understood as a difference between junior and senior form teachers.

Charlie stressed that it was just a coincidence that he had to take up all the Secondary 6 and 7 classes this year. Accordingly, similar statistical procedures using in the last section were performed with the necessary adjustments to explore the contrasts between junior and senior form teaching.

Major contrasts in underlying dimensions and global indicators

In the one-way between-subject ANOVA, the dependent variables were still the mean scores of the various underlying dimensions of teacher behaviours observed using the ISTOF and QoT instruments and the two global indicators of overall teaching effectiveness, but the independent variable was changed to the year group difference: 38 lessons of Form 1 to Form 3 were grouped as the junior year group while 23 lessons of the Forms 4 and 5 as the senior year group, and 15 lessons of Forms 6 and 7 by Charlie as *upper senior* year group. Table 8.5 showed that there were some statistically significant year group differences.

Table 8.5: **Subtle contrasts in the multiple comparisons of year groups on different underlying dimensions and the global indicators of overall teacher effectiveness**

Underlying dimension/Global indicator	F or Brown-Forsythe F	df1	df2	Sig.	Adj. R^2
Meta-cognitive skills teaching*	11.97	2	73	***.001****	.15
Classroom management and climate	12.11	2	69.05	***.000****	.16
Differentiation and support	1.17	2	72.06	.317	.002
Clarity and logic of presentation	4.42	2	71.23	***.016****	.08
Student engagement	1.84	2	70.06	.167	.003
Strategies to enhance learning and lesson focus	1.10	2	70.12	.338	.15
Integrated class management and climate	21.61	2	54.11	***.000****	.22
Structured teaching skills	6.85	2	69.90	***.002****	.08
Effective class/lesson planning	1.59	2	27.04	.223	.08
Overall judgment of quality of teaching	1.07	2	70.10	.350	.02
Good individual involvement by the pupils	8.07	2	66.95	***.001****	.12

Note: Brown-Forsythe F statistic was used instead of the normal F-test as Levene test indicated the variances of five dimensions/indicators shown with an asterisk were heterogeneous and the group sizes were unequal

In order to explore further whether individual teachers might have taught significantly differently in junior and senior form lessons, another two way between-group ANOVA was conducted only on lessons of the junior and senior year groups with *the teacher* was also added as anther independent variable. Except for *Differentiation and support*, many of the year group differences previously indicated in Table 8.5 disappeared when the *upper senior* form results were excluded in the two-way ANOVA, as shown in Table 8.6. *In contrast, the school year effect was found statistically significant in* *many underlying dimensions and the global indicator. The interactions of the school year and teacher effects indicate some teachers' teaching behaviours were found significantly different in junior and senior forms.*

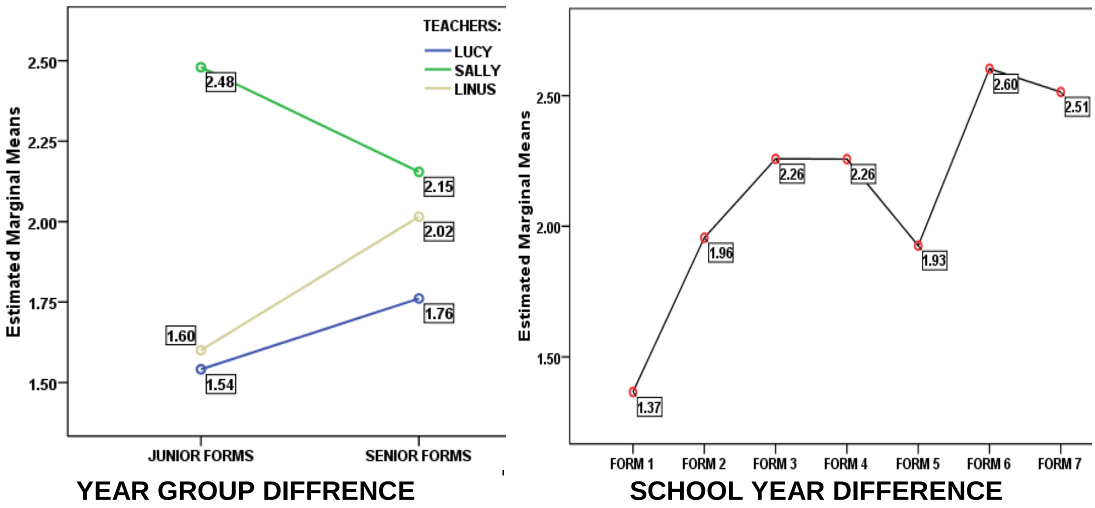
***Table 8.6:* A summary of year group effect and school year effect and their interaction with the teacher effect on different underlying dimensions and the global indicators of overall teacher effectiveness**

Underlying dimension/ Global indicator of overall teaching effectiveness	**Year Group Effect**		**Teacher-Year Group Interaction**		**School Year Effect**		**Teacher-School Year Interaction**	
	Sig.	**Eta R^2**	**Sig.**	**Eta R^2**	**Sig.**	**Eta R^2**	**Sig.**	**Eta R^2**
Meta-cognitive skills teaching*	.50	.008	.13	.072	.048*	.16	.11	.088
Classroom management and climate*	.61	.005	.76	.010	***<.001****	.47	.24	.063
Differentiation and support	***.026****	.087	.47	.027	***.005****	.23	***.045****	.117
Clarity and logic of presentation*	.07	.059	.37	.036	***.010****	.21	***.030****	.130
Student engagement	.10	.049	***.004****	.184	***.001****	.27	***.001****	.236
Strategies to enhance learning and lesson focus	.06	.061	.65	.018	.141	.12	.63	.027
Integrated class management and climate*	.83	.001	.14	.070	***.005****	.28	***.022****	.139
Structured teaching skills*	.32	.018	.07	.092	***.004****	.23	***.048****	.115
Effective class/lesson planning	.44	.011	.32	.041	***.009****	.21	.36	.048
Overall judgment of quality of teaching	.69	.003	.25	.049	***.002****	.25	.15	.080
Good individual involvement by the pupils*	.15	.037	.15	.068	.063	.15	.16	.076

Note: * in the first column indicates statistically significant year group effect when the data of upper senior form were included.

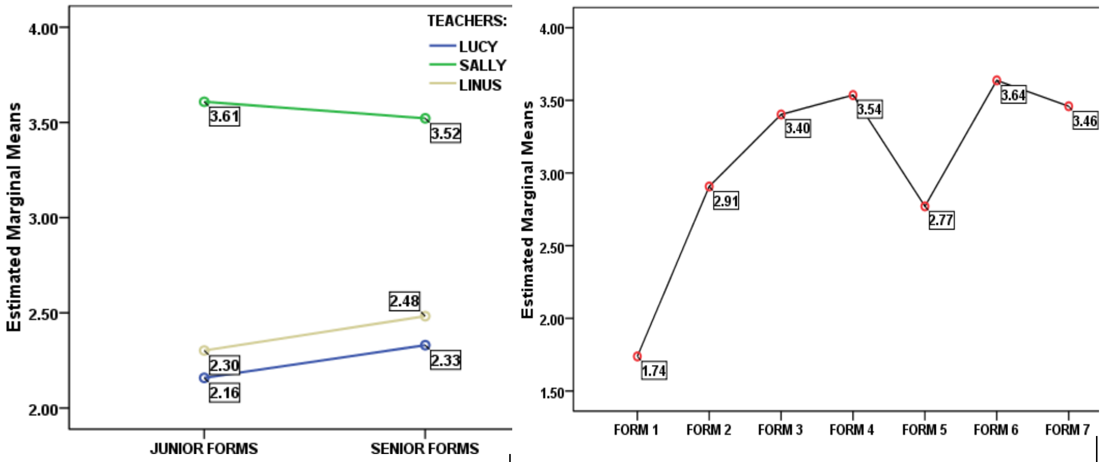
For example, as illustrated in Figure 8.3, while Sally's scores in the dimension *Meta-cognitive skills teaching* declined in the senior forms, those of Lucy and Linus improved. However, a school year by school year comparison indicates meta-cognitive skills teaching improved from the lowest score in the first year of the secondary school, became static in Form 4 and declined in Form 5. These results indicate that deep learning might be only present in the upper senior forms, but still weak in the senior forms.

***Figure 8.3:* Teacher difference, year group difference and school year difference in Factor Meta-cognitive skills teaching**



As for classroom management and climate, both Mr. Kwong and Charlie acknowledged the difficulties that teachers encountered in the junior forms. Their views were supported as teachers' classroom management was significantly better in Form 2 and higher forms. Figure 8.4 shows patterns of this underlying dimension similar to those of Meta-cognitive skills teaching.

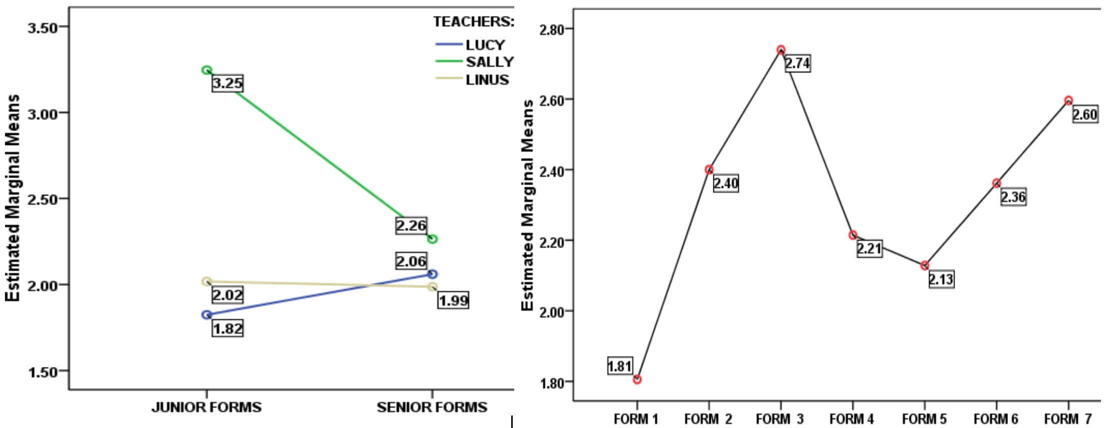
Figure 8.4: **Teacher difference, year group difference and school year difference in Factor Classroom management and climate**



YEAR GROUP DIFFRENCE **SCHOOL YEAR DIFFERENCE**

Senior forms students, particularly those in Form 5, might be better behaved than junior form students, but tended to participate less in the lessons. They might feel passive approaches were better for exam success as this was in line with the private tutorial approaches. Figure 8.5 shows that starting from Form 4 teachers generally could not effectively engage students and Charlie could not engage upper senior form students at levels higher than that in Form 3.

Figure 8.5: **Teacher difference, year group difference and school year difference in Factor Student engagement**

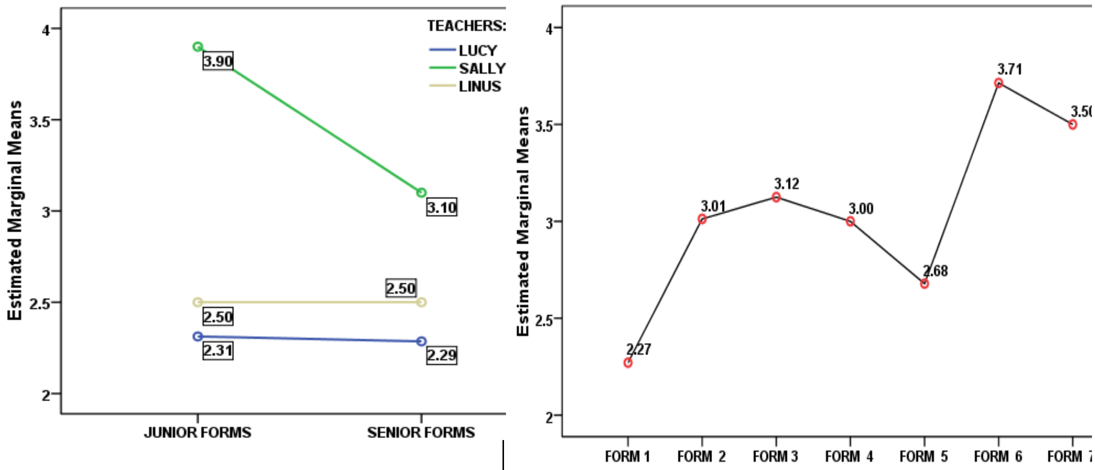


YEAR GROUP DIFFRENCE **SCHOOL YEAR DIFFERENCE**

None of the teachers could engage senior form students more though these students were supposed to be more manageable. As Sally was much more able to engage junior form students than senior form students, both teacher and year group interaction and teacher and school year interaction were found statistically significant, as indicated in Table 8.6.

Interestingly, Figure 8.6 shows that the actual levels of student involvement in junior and senior forms did not vary much. Thus, student participation did not decline as much as the teachers' abilities to engage students and there was not any significant teacher and year group interaction, nor teacher and school year interaction.

Figure 8.6: **Teacher difference, year group difference and school year difference in Indicator Good individual involvement by the pupils**

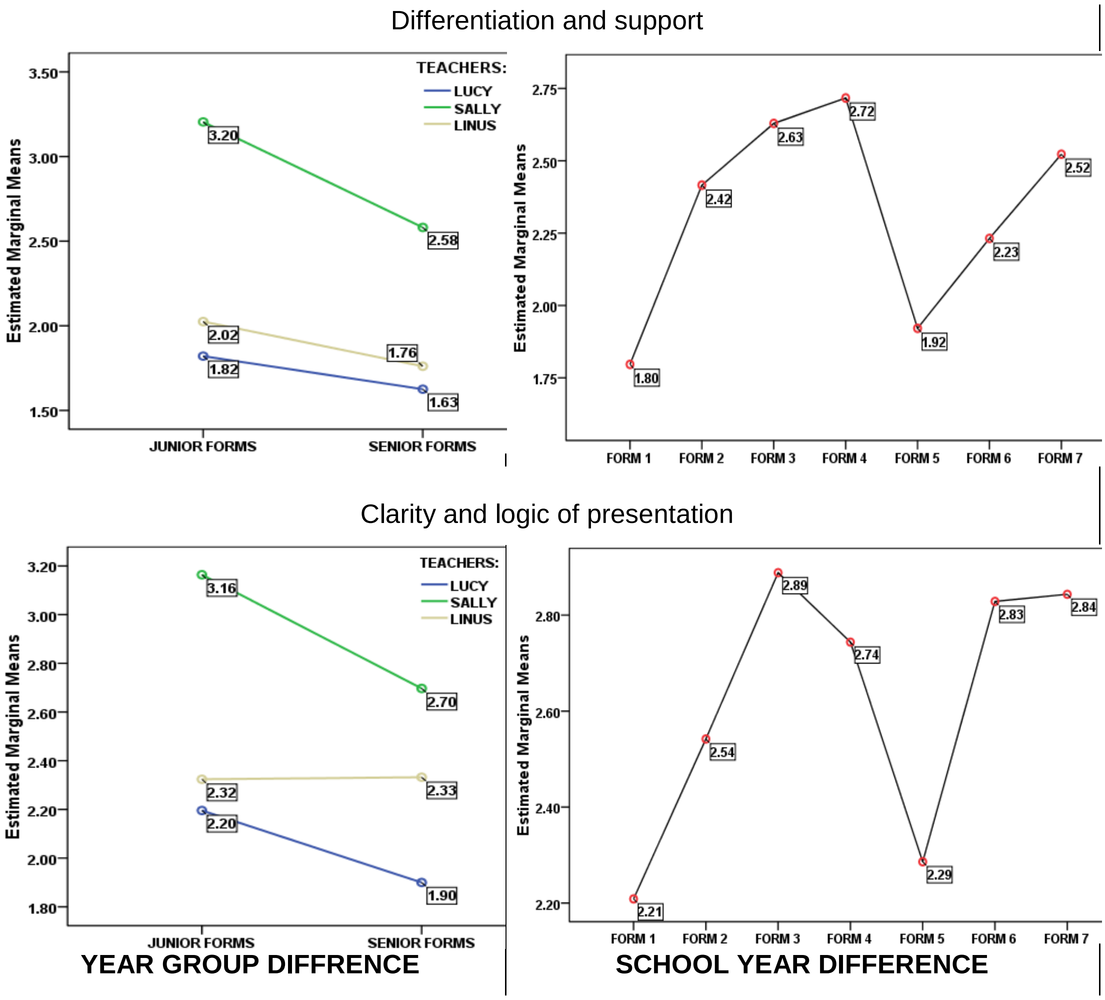


YEAR GROUP DIFFRENCE　　　　SCHOOL YEAR DIFFERENCE

The level of student participation did not decline as much as the teachers' abilities to engage students and there was not any statistically significant teacher and year group interaction, nor teacher and school year interaction. Student participation was particularly high in the upper senior forms. The year group contrast is not as salient as the school year contrast because teachers were in general rated significantly lower for their Form 1 lessons.

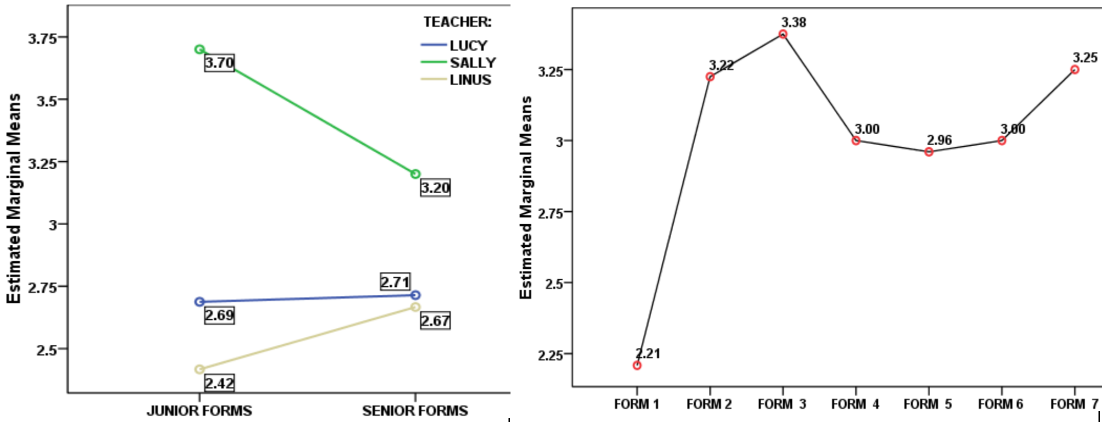
Figure 8.7 shows two of the underlying dimensions which have a statistically significant school year effect and a statistically significant teacher and school year interaction, as indicated in Table 8.6. Both Lucy and Linus scored low for the dimensions *Differentiation and support* and *Clarity and logic of presentation* in their Form 1 and Form 5 lessons. Sally also scored significantly better in both underlying dimensions in her Form 2 classes than Form 4 and Form 5 classes.

Figure 8.7: **Teacher difference, year group difference and school year difference in Factors Differentiation and support and Clarity and logic of presentation**



Although the scores for both underlying dimensions were significantly lower in Form 5 lessons, the judgment of overall teaching quality, as shown in Figure 8.8, was not particularly lower for these lessons. Thus, the global judgment does not seem to reflect some of the less highly rated specific features in these senior form lessons. The decline in these underlying dimensions found in Form 5 lessons seemed to reflect a negative washback effect of the public examination. On the one hand, Linus explained the lack of motivation and participation in these students as a result of their recognition of their predicted results in the public exam or the so called learned helplessness as Sally suggested. On the other hand, the overemphasis on exam skills and drills may well have made the lessons dull and discouraging to those students whose English proficiency levels were not ready for the exam.

Figure 8.8: **Teacher difference, year group difference and school year difference in Indicator Judgment of overall teaching quality**



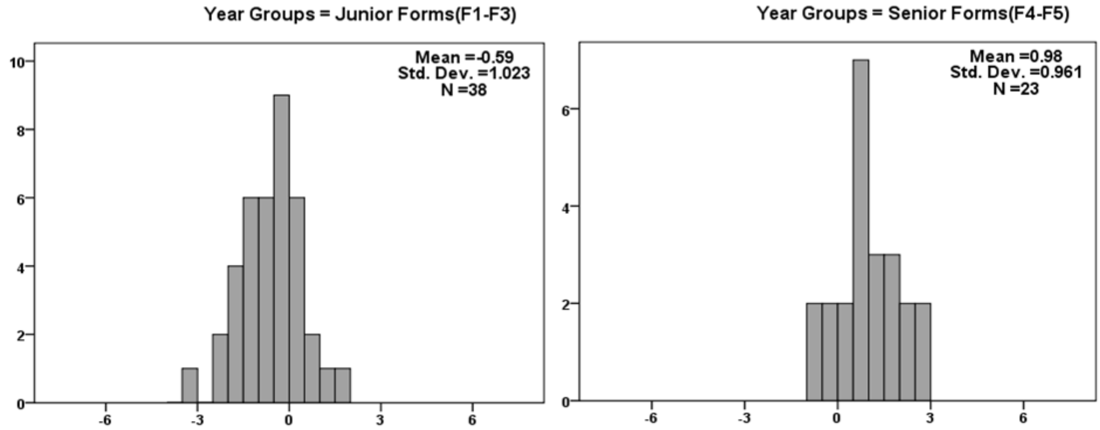
YEAR GROUP DIFFRENCE | **SCHOOL YEAR DIFFERENCE**

Ironically, teaching quality might be hampered in the school year when the students badly need some exceptionally effective teaching to counter the negative school effect and the negative previous teaching effect suggested in Marzano's (2003) model. Contrary to Charlie's perceived lack of positive results in their emphasis of interactive teaching strategies in the junior form teaching, the observed teaching practices were generally favourable in Form 2 and 3 lessons when students became used to the new learning habit and conformed to the expected classroom behaviours. These findings were consistent with the favourable comments made by the Form 2 and Form 3 students in the focus group interview. They also suggest students would show strong motivation to learn whenever they thought that they were taught by effective teachers.

Grouping junior and senior form lessons

In order to characterise junior and senior form lessons through the discriminant function analysis, underlying dimensions of both instruments were used as the predictor set because it was illustrated in the last section that such a combined set of underlying dimensions had the best explanatory power and the highest classification rate. The only one discriminant function identified was significant and, as shown in Figure 8.9, could classify 51 (83.6%) of the 61 junior and senior form lessons correctly into two groups with the respective group centroid at -0.59 and 0.98, though the classification was more successful for junior form lessons (89.5%) than senior form lessons (73.9%).

Figure 8.9: **Junior and senior form lessons grouped by a discriminant function defined by a combined set of underlying dimensions of both instruments in two separate-group scatter plots**



89.5% of original grouped lessons correctly classified; Group centroid= ***-.59***

73.9% of original grouped lessons correctly classified; Group centroid= ***.98***

Table 8.7 indicates that the discriminant function is accountable for about 37% of the variance and the underlying dimension most strongly associated with this discriminant function is *Classroom management and climate*.

Table 8.7: **Relationships of underlying dimensions and their correlations with the discriminant function**

Underlying dimensions defining the discriminant function	Correlation coefficient	Standardised Coefficient	Squared canonical correlation
Classroom management and climate	**0.23**	**0.98**	
Meta-cognitive skills teaching	0.19	0.86	
Strategies to enhance learning and lesson focus	-0.16	-0.26	
Differentiation and support	-0.14	-1.71	
Clarity and logic of presentation	-0.14	0	**.37**
Student engagement	-0.11	-1.52	
Integrated class management and climate	.08	1.90	
Structured teaching skills	-.01	-.06	
Effective class/lesson planning	-.04	-.09	

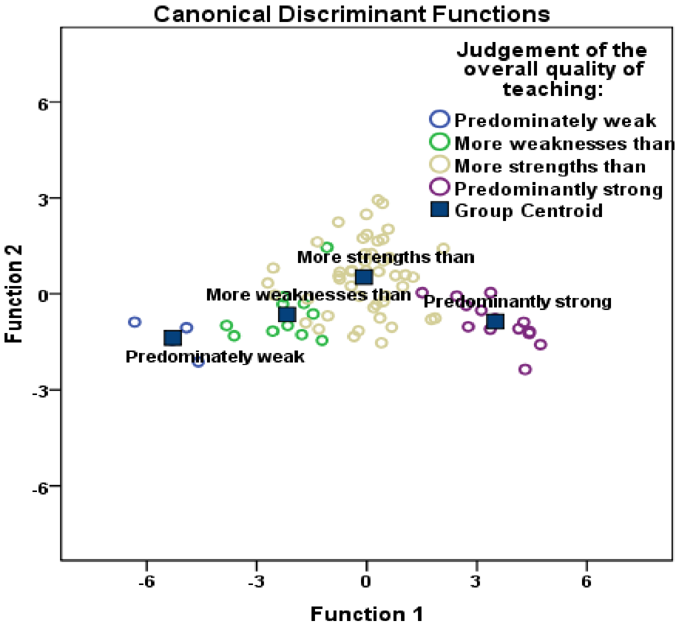
Interestingly, the dimensions *Classroom management and climate* and *Meta-cognitive skills teaching* are *positively* associated with the discriminant function, but many other underlying dimensions such as *Strategies to learning and lesson focus, Differentiation and support*, and *Clarity of Presentation* are *negatively* correlated with this discriminant function. *Classroom management and climate* and *Meta-cognitive skills teaching* are two of the underlying dimensions which did not show a statistically significant year group effect in the one-way ANOVA (see Table 8.6), but were precisely declared by Mr. Kwong as aspects that would distinguish the senior form teaching.

8.4.6 Characterising the more effective and the less effective lessons

The present study could not to distinguish between effective and ineffective lessons, because there was no data available on students' academic outcomes that could be associated with the observed teaching behaviours. Rather, lessons could only be classified based on the ratings of the global indicator of overall judgment of teaching quality. However, this classification is considered as meaningful for the English teachers as they would have a clue of what aspects of teaching behaviours would be crucial in their school.

Accordingly, a simultaneous descriptive discriminant analysis was conducted using the underlying dimensions of both instruments as the predictors and the ratings of the global indicator of overall judgment on teaching quality as the dependent variable. The classification results are illustrated in Figure 8.10.

Figure 8.10: **Effective and ineffective lessons grouped by a discriminant function defined by a combined set of underlying dimensions of both instruments in an all-groups scatter plot and the successful classification rates**



Successful classification rates

Teaching quality of lesson	Percent correctly classified	Predicted/ actual count
Predominately weak	100%	4/4
More weaknesses than strengths	81.8%	9/11
More strengths than weaknesses	89.4%	42/47
Predominantly strong	92.9%	13/14
Total	89.5%	68/76

Figure 8.10 shows the two statistically significant discriminant functions[93] that were found to be able to distinguish the lessons as close to the four categorical ratings of the global indicator of overall judgment of teaching quality. Based on these two discriminant functions, about 90% of the lessons could be correctly classified in the actual categories they were rated in the global indicator. As shown in the lower successful classification rates in Figure 8.10 and the distinctions in group means in Table 8.8, the discriminant functions were less reliable in distinguishing the lessons with either more strengths or weaknesses.

Table 8.8: **Distinctions among lessons as defined by unstandardised canonical discriminant functions evaluated at group means (or group centroids)**

Lesson categories classified by their teaching quality	Discriminant Function	
	1	2
Predominately weak	-5.29	-1.38
More weaknesses than strengths	-2.18	-.64
More strengths than weaknesses	-.08	***.53***
Predominantly strong	***3.51***	-.87

A close examination of the relationship of the underlying dimensions and the two discriminant functions in Table 8.9 reveals that *Structured teaching skills* and *Student engagement* are the two key underlying dimensions of the first discriminant function which distinguishes the more effective lessons from the less effective lessons.

Table 8.9: **The relationship between the two discriminant functions that distinguish teaching quality and their defining underlying dimensions**

Discrim-inant function	Underlying dimensions defining the discriminant function	Correlation coefficient	Standard-ised coefficient	Squared canonical correlation
First	**Structured teaching skills**	**0.77**	.64	**0.82**
	Student engagement	**0.72**	.26	
	Differentiation and support	0.63	.04	
	Integrated class management and climate	0.61	.16	
	Clarity and logic of presentation	0.50	-.08	
	Strategies to enhance learning and lesson focus	0.44	.20	
Second	**Classroom management and climate**	**0.62**	.41	0.33
	Effective class/lesson planning	-0.48	-.52	
	Meta-cognitive skills teaching	0.37	.34	

Note: For limited space here, the correlation coefficients and standardised coefficients listed here are only those coefficients of the discriminant functions with which the underlying dimensions most strongly associated.

93 Wilks'Lambda of the discriminant functions is summarised in the following table:

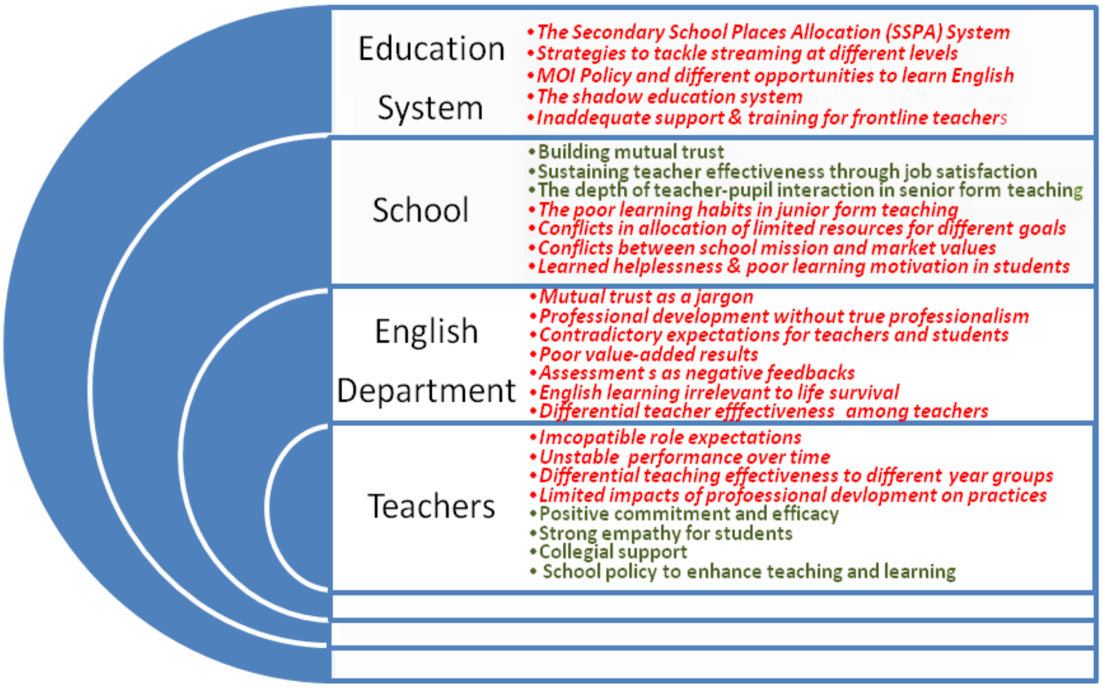
Predictor set	Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig. level
Combined set of both ISTOF & QoT underlying dimensions	1 through 3	.10	156.67	27	.000*
	2 through 3	.58	37.68	16	.002*
	3	.86	10.13	7	.181

In contrast, *Classroom management and climate* is the key dimension of the second discriminant function that distinguishes those lessons with more strengths than weaknesses from other lessons. As the amount of variance that can attribute to the first discriminant function is about 2.5 times of that of the second discriminant function (i.e., 82% vs. 33% as indicated in their squared canonical correlations), teachers and administrators in *Ming Tak Comprehensive* should pay more attention to classroom practices that were found to be associated with the dimensions *Structured teaching skills* and *Student engagement*. These results, however, do not necessarily suggest that classroom management is not important and contradicts the results discussed in Section 8.4.4. Rather, it is more likely that classroom management is a necessary but not a sufficient condition to make lessons effective.

8.5 Synthesis

By integrating the contextual backgrounds in Chapter 2 and the findings from Chapter 6 to the present chapter, the multilevel case model illustrated in Figure 3.3 of Chapter 3 can be updated as Figure 8.11 below (positive factors are shown in green, while negative factors in red).

Figure 8.11: **A multi-level case study model showing the factors affecting teacher practices and teacher effectiveness of EFL teachers at various levels**



Although there were several important negative challenges facing the school, positive factors include the senior managers' willingness to build mutual trust and to sustain teacher effectiveness through job satisfaction and empowerment, the strong commitment and efficacy reported by teachers, collegial support among frontline teachers, and school policy to enhance teaching and learning. The school principal has recognised most of the challenges facing the school, but his strategies addressing to these were not based on evidence of practice in the school. Accordingly, the gap in the learning targets for junior and senior forms, the competitions between departments in allocation of resources, and the tensions between the school mission and the market values might be intensified. That is, without proper classroom management, other aspects of teaching may not be achieved effectively. Both Lucy and Linus taught better in lessons when they could manage their classes.

Factors at the teacher and departmental levels show mixed effects. The English teachers showed their empathy for students' learned helplessness and poor self-concept while recognising that strong determination by the students is required to counter all the negative experiences as this is the reality in the education system. Charlie did recognise the limited impacts of professional training and development on teachers' practices. In particular, Charlie's view that some of their teaching practices to Form 1 students would not work was confirmed in the lower scores found in the underlying dimensions in Form 1 lessons. This finding was consistent with the view that professional development that could not change pedagogies or professional attitudes would contribute little to make a difference. Linus found the functions of their assessments negative to students.

There seemed to be a gap between teachers and students' perceptions about what benefited English learning. Successfully implemented activities that showed high individual participation and enjoyment in Lucy and Sally's lessons indicated that their students learned better, when the tasks were more relevant to students' life contexts as it has been found that students in challenging contexts learned better if their learning is relevant to their life experiences (Muijs et al., 2004). It may be true that there are contradictory expectations for the students in the system, but if they are not replicated in the school through curricula, grouping, and teaching practices, then students can still make progress accordingly.

It is inappropriate to regard Lucy only as an ineffective teacher because the observations showed that she could be effective when the grouping procedures or class composition suited her. She functioned well in higher ability, less behaviourally challenging contexts. The discriminant function analysis showed that Linus was rated as outstanding in *Strategies to enhance learning and lesson focus*, probably because he insisted on implementing the reflective learning policy in his teaching. Thus, school policies on class composition and learning priority can considerably affect the overall and individual dimensions of the teaching effectiveness of individual teachers.

From their self-reports, both Lucy and Linus do not seem to be in crisis in their professional life cycles that may undermine their teaching effectiveness (Day et al., 2006). Rather, Linus and Lucy are teachers who need collegial support. Lucy became vulnerable when students were unruly and unmotivated. She can enhance her teaching effectiveness across contexts by overcoming the negative class composition effect and develop positive class management strategies. As Linus might be too conscious about his role as a SEN teacher to the extent that he became more tolerant to some students' misbehaviors, he can improve his teaching effectiveness by responding proactively and accordingly as an effective teacher who cares about the learning activities of every student in the class. In short, the findings in the individual case study have indicated that the overall teaching effectiveness of the English department can be enhanced if teachers like Lucy can be consistently effective across lessons and teachers like Linus can show less variation in teaching effectiveness in most of the aspects of classroom practices.

A combined set of underlying dimensions was found to be reliable in characterising the differences among the lessons of the four teachers. The major contrasts were found to lie in the dimensions *Classroom management* *and climate*, *Differentiation and support* and *Strategies to enhance learning and lesson focus*. Regarding the junior and senior form teaching, the findings were somewhat mixed. On the one hand, consistent with Mr. Kwong's own analysis as school principal, the junior and senior lessons could be distinctively grouped with the dimensions *Classroom management and climate* and *Meta-cognitive skills teaching*. On the other hand, contrary to what the perceptions of both Mr. Kwong and Charlie, the junior and senior form lessons did not differ significantly except in the score on the factor *Differentiation and support*. Rather, the contrasts should be focused on the relatively lower teaching performance in both Form1 and Form 5 lessons. It seemed that both Form 1 students and their teachers were reciprocally affecting each other and both needed time to adjust their behaviours, while teaching and learning in Form 5 classes were loaded by their exam-orientations.

However, these orientations were not easily adjustable as they were related to a wider context that had made the school and its teachers and students vulnerable. According to Marzano's (2003) probability model of educational effectiveness (Section 2.6.3), English teachers in *Ming Tak Comprehensive* would be more easily discouraged by the general poor student outcomes and the risk of being labelled as ineffective teachers, while students would learn helplessness when they recognised the wide gaps between their actual English proficiencies and the expected requirements in more and more internal and external examinations.

Consistent with the multiple regression results discussed in Chapter 6 (see Table 6.16), the dimension *Structured teaching skills* (see Table 5.9) was found to be the key dimension of the discriminant function that distinguished more effective lessons from less effective lessons. This underlying dimension was more important than the dimension *Classroom management and climate* in making lessons effective, but comparisons between the four teachers in Section 8.4.4 highlighted the importance of classroom management. These results are not contradictory, but together support the view that effective teaching is a *multitask* skill (see Table 6.17) that requires clear instructive and explanations, well-adapted assignment and activities and appropriate teaching strategies that make learning possible, as well as well-structured lesson management and supportive and stimulating lesson climate that make learning easier and pleasurable.

Finally, findings of this chapter contribute much to future classroom observation research and MM research. First, because the combined set of dimensions of the two instruments discriminates the teachers better than sets of dimensions of a single instrument, the two instruments seem to contribute more when they are employed together. This suggests that research and teacher evaluation that rely on single instrument may be subject to unintended biases or limitations. Second, most of the quantitative analyses were performed after the initial phase of analyses as they addressed the additional research questions. This suggests the dynamic MM approach which generates exploratory and explanatory research questions and mixes qualitative and quantitative methods is a fruitful and flexible research strategy (see Tashakkori & Teddlie, 2003a; Sammons, 2010).

The findings of this chapter will be integrated with findings of previous chapters in the next concluding chapter.

CHAPTER 9 : CONCLUSIONS AND IMPLICATIONS

9.1 Introduction

This research contributes to the existing educational effectiveness knowledge base by producing a multilevel empirical case study using a mixed-method (MM) approach to investigate teachers' classroom practice. I employed the same international classroom observation instruments used by Day and his colleagues (2008) in a study of teachers in England, but in a different context. I restricted the scope by looking in much more depth at consistency and variation in observed teaching behaviours of only four EFL teachers of a single school in Hong Kong. All the English lessons of each participating teacher were observed extensively in a five-day observation period, about a month after the commencement of the fall term. This strategy of observing in multiple lessons allowed a sufficient amount of classroom observation data for the subsequent quantitative analyses as well as for the triangulation and integration with qualitative data collected in the form of qualitative field notes of classroom events and semi-structured interviews with teachers intended to explore their own accounts of their intentions and practices in teaching. The qualitative field notes provided information on the classroom processes in details that are necessary to enhance understanding of variation in observed practices and to link with the interview evidence. The interview was intended to elicit the teachers' views of their teaching practices and of the system, school and student factors that they felt influencing these practices. These qualitative data were used to enrich the case studies of individual teachers. Further interviews with the department head and the school principal were conducted to explore the working context and policies of the English department, in the school and in the educational system of Hong Kong.

In the last four chapters, I have presented and discussed the findings of the quantitative and qualitative analyses conducted in order to address the seven key research questions proposed in Chapter 3. In the remaining sections of this concluding chapter, I will start with a brief review (Section 9.2) of these original research questions and four additional research questions that arose in the analysis stage (see Chapter 8). Then, for each of these questions, the main findings are summarised and integrated. Afterwards, the significance of the findings and the implications for future research that seeks to examine the topic of influences on teachers' classroom practices and their effectiveness will be discussed (Section 9.3). These findings are reviewed (Section 9.4) in light of other research findings to generate implications for future direction for school improvement initiatives and teacher development. Section 9.5 will address the limitations of the study. New contributions made by these findings and their implications for future researches are then presented before some concluding remarks on the MM research process.

9.2 A review of research questions, their related chapters and purposes

A set of seven research questions have been put forward and addressed in different relevant chapters:

- What are the characteristics and underlying dimensions found in the observed classroom practices across a large number of lessons? How do they vary with student backgrounds and class compositions? (*Chapters 4-8*)
- To what extent are these characteristics and underlying dimensions comparable to those identified in the English study by Day et al. (2008), despite the sample and contextual differences? (Chapters 4-6)
- To what extent are the characteristics and underlying dimensions identified using different quantitative observations instruments comparable? (Chapter 6)
- To what extent do these characteristics and underlying dimensions contribute to the overall judgment of quality of teaching in the lessons observed? (Chapters 6 and 8)
- To what extent do these characteristics and underlying dimensions contribute to the individual involvement by the students in the lessons observed? (Chapter 6)
- To what extent do these characteristics and underlying dimensions vary among individual teachers and vary across the lessons of each teacher? (Chapters 7 and 8)
- What are the teachers' views and perceptions about their teaching practices, their students' learning and the contextual factors that may affect teaching and learning in the school? In what way are they affected? (Chapters 7 and 8)

The first two research questions are important as their answers would contribute to the theoretical debate between the generic theories of teacher effectiveness (GTE) (e.g., the Dynamic model of educational effectiveness, DEE, Creemers & Kyriakides, 2008) and the differentiated theory of teacher effectiveness (DTE) (e.g., Campbell et al., 2004). According to the GTE, teachers would *consistently show similar strengths and/or areas that might be improved* in different aspects of their classroom practices across different contexts. In contrast, the DTE theory would propose that the *strengths and areas needing improvement* of teachers in different aspects of their classroom practices tend to *vary across different contexts* (e.g., in terms of different age groups or different ability groups of students). An advocate of a generic theory is more likely to hypothesise that different aspects of teaching practices would be similar for individual teachers (i.e., generally effective in most aspects or generally typical or generally less effective) and that effectiveness features would tend to be similar in different cultural contexts.

The third research question is related to the theoretical debate, as a proponent of a generic theory is more likely to believe that different classroom observation instruments could measure similar, rather than different, underlying dimensions of teaching behaviours. Certainly, the finding of this question is very much dependent on the particular instruments selected. Instrument comparison is an important methodological issue that has been rarely addressed in the existing TER literature (see Chapter 2). In the fourth and fifth questions, the distinctive characteristics, or underlying dimensions of observed teacher behaviours identified, are associated with two global indicators of overall teacher effectiveness, namely, *Overall quality of teaching* and *Good individual involvement of pupils*. The multiple regression results show the relative contributions of these characteristics to global ratings of overall teacher effectiveness. The answers to the last two questions are considered to be crucial for providing rich descriptions of a multilevel case study. These descriptions include an account of teachers' perceptions of the challenges facing them as EFL teachers working in a low attaining secondary school in Hong Kong where Chinese is the medium of instruction. In addition, the case study also describes the classroom contexts of these teachers that may affect their teaching practices and the patterns of strengths and areas needing improvements as identified in the various underlying dimensions of their observed teaching behaviours.

Four additional research questions that emerged as important in the later stage of analysis following initial quantitative and qualitative stages of analysis (see Section 3.3.1):

- To what extent are the characteristics identified in different theoretical and empirical models comparable? (*Chapter 6*)
- In what ways does the observed variation in teacher effectiveness across the four case study teachers affect the overall effectiveness of the English department and the school? (*Chapter 8*)
- In what ways does the junior from teaching differ from the senior form one? (*Chapter 8*)
- In what ways do the features of more effective and less effective lessons differ? (*Chapter 8*)

The first of these questions arose when the theoretical and empirical CFA models were found with similar supports in the data. It would be interesting to explore whether they are equally comparable in their relations to global indicators of overall teaching effectiveness. The other three questions were raised during the qualitative analyses after the first phase of quantitative analyses in Chapters 4 to 7 were completed. These questions were considered important, as their answers would help to enhance understanding of why teaching English in *Ming Tak Compreh*ensive was found to be particularly challenging for these four EFL teachers.

9.3 Main findings and the corresponding research questions

9.3.1 Consistency and variation in teaching behaviours *across lessons*

To address the *first* key research question, confirmatory factor analysis (CFA) was chosen to examine whether there were clear and identifiable underlying dimensions of observed teachers' behaviours that could be identified as the distinctive characteristics of their classroom practices. Chapters 4 and 5 have respectively presented the CFA results of using the *International System for Teacher Observation and Feedback* (ISTOF) Scale and the *Lesson Observation Form for Evaluating the Quality* *of Teaching* (QoT) as classroom observation instruments for studying variations in teacher and school effectiveness in classroom behaviours and in a Hong Kong secondary school. The ISTOF model was based on existing review of SER and TER evidence and expert opinion in a large number of countries. In contrast, the QoT model was based on an inspection model of effective classroom practice.

Multidimensionality of teaching identified using the ISTOF instrument

The CFA results presented in Chapter 4 were encouraging as they provided evidence to compare both the GTE and the DTE theories in different ways. First, the six underlying dimensions identified showed high internal consistency and the CFA model they formed was strongly supported by the observation data. These findings lend support to the GTE view that there are distinctive characteristics in the observed teachers' behaviours in the following dimensions of observed teaching behaviours:

- Meta-cognitive skills teaching
- Classroom management and climate
- Differentiation and support
- Clarity and logic of presentation
- Student engagement
- Strategies to enhance learning and lesson focus

Second, as shown in Table 4.14, there are more similarities than discrepancies between these underlying dimensions and the original theoretical components of the two instruments. The first four of these underlying dimensions largely correspond with four of the seven theoretical components included in the original in the scale of ISTOF. In contrast, the last two show a composite structure because they consist of teaching behaviours originally categorised in terms of different theoretical components by the ISTOF instrument. For example, teacher behaviours associated with the dimension *Student engagement* suggest a mixture of strategies to enhance inclusion, to engage students with real life experiences, and to enhance participation and involvement. Similarly, items of *Strategies to enhance learning and lesson focus* dimension include praising students for realising their potentials, explaining the purpose of learning activities, and aligning assignments to the goal of the lesson. Teaching behaviours in the dimensions *Student engagement* and *Strategies to enhance learning and lesson focus* are respectively similar to the constructs *Application* and *Orientation* in the DEE (Creemers & Kyriakides, 2008). This suggests that while some of the original theoretical components of ISTOF may be more important and invariant with context than others, the scale might not have exhausted all the distinctive general aspects of classroom practices. Interestingly, the present data also supported a CFA model formed entirely by its original theoretical components after systematically reducing some items for each component. Thus, the existence of different CFA models that seem to be supported by the data eventually leads to the investigation of the first additional research question.

Finally, Section 4.5 shows that the frequency distributions of these underlying dimensions vary in different extents across lessons. That is, regarding variations in teachers' observed behaviours as described in the ISTOF items, dimensions like *Metacognitive skills teaching* might vary more often than other dimensions like *Classroom management and climate*. This pattern was also found in the ECP sample of effective teachers in England (Ko & Sammons, 2008b; Sammons & Ko, 2008). This study argued that those dimensions showing higher variability might not be the defining characteristics of effective teaching practices. In the present context, the findings seem to support a view in line with the DTE theoretical perspective. That is, teacher effectiveness may vary in different contexts and some dimensions tend to be less stable across contexts than others. Thus, while *effective teaching is most likely a multidimensional construct*, it is also clear that *different dimensions of teaching behaviours can vary considerably across lessons and for different teachers.*

Multidimensionality of teaching identified using the QoT instrument

The CFA results presented in Chapter 5 based on the QoT instrument seem to lend more support to the DTE than the GTE theoretical view, though the kinds of supportive evidence were similar. First, only three main underlying dimensions were identified in a scale originally hypothesised to have nine theoretical factors:

- Integrated class management and climate
- Structured teaching skills
- Effective class/lesson planning

These underlying dimensions showed strong coherence and their CFA model was also strongly supported by the data.

Second, with such a small number of dimensions for QoT and all indicators retained, there were more indicators in a single dimension than in the CFA model of ISTOF. For example, the first dimension has 13 indicators, while the second one has 10. Accordingly, Table 5.12 shows that, except the last dimension, the other two dimensions seem to be associated with more original theoretical factors. Thus, the underlying dimensions of QoT may be viewed as *theoretically less distinctive* than their ISTOF counterparts. The QoT instrument was developed with different theoretical structures at a different time (van de Grift et al., 2004; van de Grift, 2007) and the CFA models based on these theoretical structures seemed to be equally supported in the data. Because teachers' observed behaviours were rated on both instruments for the same set of lessons across a week, direct comparisons can be made between the two instruments in terms of the various underlying dimensions identified.

Third, Section 5.5 shows that the first two underlying dimensions of QoT, *Integrated class management and climate* and *Structured teaching skills*, have a clear bimodal distribution, indicating that in many lessons scores either highly or poorly on these aspects. This finding is in line with the DTE theory of variation in teacher effectiveness, but since the unit of analysis was the lesson, further analysis was required to show that variability did not only exist across lessons of different teachers, but also across lessons of the four individual teachers studied.

Characteristics of effective lessons observed

One of key messages that concerned the participant teachers most is what makes an effective lesson in their school. In order to support the co-

construction of knowledge, the research sought to identify the most important dimensions of the two key discriminant functions found to distinguish the more effective lessons observed. These are reproduced here:

Dimension	Item/ Indicator No.	Description (T=The teacher; S = student; P= pupil)
Structured teaching skills (QoT)	**IND24**	There is good individual involvement by the Ps
	IND41	T gives clear instructions and explanations
	IND52	T makes use of teaching methods that activate the Ps
	IND62	T adapts the assignments and processing to the relevant differences between Ps
	IND71	*T ensures that the teaching materials are orientated towards transfer*
	IND72	*T stimulates the use of control activities*
	IND73	*T provides interactive instruction and activities*
	IND81	*T gives a well structured lesson*
	IND82	*T ensures the orderly progression of the lesson*
	IND83	*T uses learning time efficiently*
Classroom management and climate (ISTOF)	**Item 31**	T demonstrates genuine warmth and empathy toward all Ss in the classroom.
	Item 41	Actions are taken to minimize disruption.
	Item 42	There is clarity about when and how Ss can get help to do their work in class.
	Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.
	Item 45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.

While the dimension *Structured teaching skills* consists of all the indicators of the *Teaching Learning Strategies* and *Effective Classroom Organisation* criteria of the original theoretical QoT scale, the dimension *Classroom management and climate* is dominated by items of the component *Classroom management* of the original theoretical ISTOF scale. The fine-grained elements of behaviours of the dimension *Structured teaching skills* are very similar to those behaviours identified in the profiles of effective teachers mentioned in Section 2.3.1. As the first discriminant function is about 1.5 times more important than the second one, it seems that after all, an effective lesson in *Ming Tak Comprehensive* is not much different from those found in other contexts. Certainly, classroom management still contributes much to an effective lesson in this school, especially in junior form classes. This supports existing TER conclusions.

9.3.2 Consistency and variation *in individual teachers* and *between teachers*

Using the factor scores of the underlying dimensions identified in Chapters 4 and 5, the lessons of each teacher were examined and the *sixth* and *seventh* key research question was addressed. Again, the results can be seen to provide support for both the GTE and DTE perspectives, suggesting that the two camps are compatible, rather than necessarily contradictory. Some teachers show a stronger tendency towards high scores on the underlying dimensions and others a tendency towards mostly lower scores, but all teachers show some variation and some vary more in relation to contextual factors than others.

Results of integrated quantitative and qualitative comparisons

Two teachers, Charlie and Sally, were found to score highly in most of the dimensions across most of the lessons. Sally 's lessons showed the least variability and thus she can be considered as an exemplar to illustrate that *an effective teacher would tend to be effective in all aspects of her teaching at all times*, *as the GTE theory would predict*. *None of Sally's lessons was found weak in Structured teaching skills*. Sally showed her strengths in what Tsui (2003) characterised as "expertise teaching" in *adapting materials and activities to the unexpected needs arose in the immediate classroom context*. The difference between Sally's and Charlie's lessons was subtle. Though generally effective, Charlie could not engage all students in the oral presentation routine and this routine always became a distinguished but unrelated, rather than an integrated, part of many of the observed lessons. This routine *seemed to prevent effective student engagement and blurred its relation with the lesson focus.* Charlie could be another exemplar if he had structured his lessons differently.

Lucy is an interesting case as she showed the greatest variation in classroom practices. In particular, she scored low ratings in terms of most underlying dimensions in most lessons, but she also scored highly on most of the underlying dimensions in a few lessons of some classes. These results indicated that *Lucy's teaching effectiveness was affected by some specific contexts*. *Lucy was more effective when she could effectively manage behaviour of her Form 1 class, when its class composition had changed (with fewer lower ability students and disruptive students)*. *Her scores provide support for the theory of differentiated teacher effectiveness*, which argues that teacher effectiveness is *not only* a generic characteristic of a teacher, but may also *vary in relation to changes in contexts*. That is, *Lucy could be effective in every aspect of teaching when the contexts suited her*. However, class composition is not something that a teacher normally controls because the senior management decides which classes a teacher work with. Mr. Kwong, the school principal, for example, declared that managerial decisions in schools often reflect not only educational, but also political concerns. Mixed ability teaching was not popular among teachers and school administrators despite the government's intention to increase more mixed ability students in the intake (see Section 2.2.2). Instead, in-school setting is perceived and preferred as a more effective strategy to enhance overall student outcomes in a context where enhancing public examination results is always on the top priority.

Linus' teaching behaviours also conformed to the prediction of the GTE theory. According to it, an ineffective teacher would tend to be ineffective in all aspects of teaching. *He was the only teacher who received low scores in most dimensions in a lesson.* In particular, his lower scores for the *Structured teaching skills* dimension appeared to undermine his teaching effectiveness in maintaining an acceptable teaching quality and individual involvement of the students. However, Linus sometimes showed a relatively stronger strength in the dimension *Strategies to enhance learning and lesson focus*, which could also be considered as providing supportive evidence for the alternative DTE theory indicating that teachers may vary in strength in different aspects of teaching. Interestingly, this strength of Linus seemed to be influenced by the current school policy to enhance reflective learning in students. This supports the view that a stronger school focus on teaching and learning may help to enhance teaching effectiveness in teachers.

Results of further quantitative discrimination

The results of the discriminant function analysis showed that the lessons of the four participating teachers can be distinguished on three latent traits:

- **Function 1**: **Classroom management and climate**, Integrated class management & climate, Structured teaching skills, and Meta-cognitive skills teaching
- **Function 2: Differentiation and support**, Clarity and logic of presentation, Student engagement, and Effective class/lesson planning
- Function 3: Strategies to enhance learning and lesson focus

While the first function distinguished Charlie and Sally from Lucy and Linus, the second function further distinguished Charlie and Sally. The third function indicates the unique characteristics of Linus' lessons from other teachers' ones. These results are consistent with the findings discussed above that show considerable differential teacher effectiveness in the English department. Thus, it can be concluded that Marzano's (2003) scenarios of teaching ineffectiveness due to inconsistency in teacher effectiveness in a department or a school may well occur in *Ming Tak Comprehensive*.

9.3.3 The influences of teachers and managers and their interactions with the contexts

To address the reported contextual influences in the *seventh* key research question and the *last three additional* research questions, the individual teacher case studies have been summarised and presented in Chapter 7 against a background of selected characteristics of the education system in Hong Kong and the challenges identified by one senior manager and one middle manager of the particular school where the present research was conducted.

Challenges from wider educational context

Regarding the influence of the wider educational environment in Hong Kong, the following characteristics were identified from a literature review, to be the school's policies on teaching and learning that may have affected individual teachers' teaching practice:

- Selective school places allocation system in Hong Kong
- Different groupings by streaming, setting and mixed ability
- Conflicting goals of Medium of Instruction Policy (Chinese versus English)
- Interdependence between examination-oriented education and private tuition

Challenges within the school

In the interviews with the school principal and the department head, six challenges and three contradictions were identified for *Ming Tak Comprehensive*:

Six Challenges:

- Building mutual trust between the management and the teaching staff

- Sustaining teacher effectiveness through professional development and job satisfaction
- The depth of teach-pupil interaction and student engagement as well as better academic outcomes in senior form teaching
- The poor learning habits and student outcomes in junior form teaching
- Conflicts in the allocation of limited resources to achieve different goals for student outcomes
- Conflicts between the school mission and the competitive market culture

Three Contradictions:

- Mutual trust as an aspiration rather than an achieved goal
- Professional development without true professionalism
- Contradictory expectations

Interestingly, these challenges and contradictions for a Band 3 CMI school (i.e., its student intake is among the bottom one-third of the students from feeder primary schools and its medium of instruction is Chinese) in a deprived area have not been cited by any of teachers to account for their difficulties in classroom practices. Moreover, Challenges 2, 3, 5 and 6 and Contradiction 3 are related to the four challenges of the wider context. "Area context [is] not the only factor driving quality, but it [is] a factor, interacting with market and institutional contexts and with the agency of individual managers and staff" (Lupton, 2004, p.26). Like the successful teacher model proposed by Cheng et al. (2008) and the findings by Gao and Watkins (2001), teacher and teaching effectiveness in Hong Kong is much affected by the system factors that affect the school policies on teaching and learning and that with wider socioeconomic and educational inequalities.

Variation in teaching effectiveness due to the classroom context

It is often argued that classroom contexts affect teaching and teacher effectiveness (see the literature reviewed in Section 2.3), but this issue is rarely addressed specifically in terms of dimensions of teaching practices. In the present analysis, this issue has been explored in relation to both year-group difference and school-year difference. In general, except for the dimension *Differentiation and support*, most of the year group differences disappeared when the *upper senior* form results were excluded. In contrast, statistically significant school year effects were evident in many dimensions and the global indicators. The statistically significantly interactions of the school year and teacher effects were present in junior and senior forms for several dimensions (i.e., *Differentiation and support*, *Clarity and logic of presentation, Student engagement, Integrated class management and climate,* and *Structured teaching skills*).

On the one hand, the school-year variation generally found in these dimensions is not linear, but quadratic like an inverted U-shape, indicating a gradual improvement in teaching quality from the bottom level in Form 1, a leap in Form 2 and 3, but a decline in Form 4 or 5. These results are generally contrary to the perceptions of both the school principal and the head of department, as they showed higher expectation for senior forms. On the other hand, results of the discriminant function analysis indicated that the dimensions *Classroom management and climate* and *Meta-cognitive skills teaching* distinguished the junior form and the senior form lessons. The senior staff's impression that classroom management and climate was crucial to teaching effectiveness in junior form classes was evident in both quantitative and qualitative analyses. Therefore, they might be correct about certain contrasts between the two groups. However, it should be noted that because the participant teachers taught only some of the five classes in a form (i.e. a year group), it would require further data to generalise the teaching of all the classes in the form and the subsequent interpretations have to be dealt with caution.

9.3.4 Consistency and variation in teaching behaviours *between instruments* in association to teaching effectiveness

Comparing the two observation instruments became more complicated when the CFA results in Chapters 4 and 5 indicated both the theoretical and empirical models received similar support in the data. Accordingly, in Chapter 6, the *first additional* research question is addressed together with the *third*, *fourth* and *fifth* key research questions. As in van de Grift (2007), underlying dimensions were also associated with two global indicators of teaching effectiveness, *Judgment of the overall quality of teaching* and *Good individual involvement by the pupils* to indicate the quality of teaching in the observed lessons. Based on Pearson correlations between underlying dimensions, the results extracted from Tables 6.1-6.5 are summarised in Table 9.1

Table 9.1: **Underlying dimensions of observed teachers' behaviours which showed correlations with other dimensions in varied strengths and the strongest correlation with the two global indicators of overall teaching effectiveness (N=76)**

Model	Dimensions strongly correlated with others	Dimensions weakly or moderately correlated with others	Dimensions strongly correlated with Overall teaching quality	Dimensions strongly correlated with Good individual involve-ment by the pupils
ISTOF Empirical – 6 dimensions	Student engagement;	Meta-cognitive skills teaching; Strategies to enhance learning and lesson focus	Student engagement; Differentiation and support	Student engagement; Differentiation and support
ISTOF Theoretical – 7 dimensions	Differentiation and inclusion; Instructional skills; Classroom climate	Classroom management	Differentiation and inclusion; Classroom climate	Instructional skills; Differentiation & inclusion; Classroom climate
QoT Empirical – 3 dimensions	Structured teaching skills; Effective class/lesson planning	Effective class/lesson planning	Integrated class management and climate; Structured teaching skills	Integrated class management & climate; Structured teaching skills
QoT Original Theoretical – 8 dimensions	Stimulating learning climate; Effective classroom organisation	Clear objectives	Effective classroom organisation; Stimulating learning climate; Teaching learning strategies	Stimulating learning climate; Teaching learning strategies
QoT New Theoretical – 5 dimensions	Clear Instruction; Teaching learning strategies	Adaptation of teaching	Efficient classroom management; Teaching learning strategies	Teaching learning strategies; Clear instruction

Nearly all correlations between the underlying dimensions with other dimensions in the model were found positive and statistically significant at above 0.01. In the cross-model comparisons in Pearson correlation analyses, the underlying dimensions which showed more statistically significant and strong positive correlations (*r* above 0.80), as shown in Table 9.2, are similar to those listed in the first two columns in the last table.

Table 9.2: **Underlying dimensions of observed teachers' behaviours which showed stronger correlations with dimensions in models of other instrument (N=76)**

Model	Dimensions strongly correlated with dimensions in models of other instrument	Dimensions weakly or moderately correlated with dimensions in models of other instrument
ISTOF Empirical – six dimensions	Student engagement; ***Classroom management and climate***	Strategies to enhance learning and lesson focus
QoT Empirical – three dimensions	***Teaching learning strategies;*** Effective class/lesson Planning	Effective class/lesson planning
QoT Original Theoretical – eight dimensions	Stimulating learning climate; Effective classroom organisation	Clear objectives
QoT New Theoretical – five dimensions	Clear Instruction; Teaching learning strategies	Adaptation of teaching

The only two exceptions are *Classroom management and climate* of the ISTOF empirical model and *Teaching learning strategies* of the QoT original theoretical model. These results have indicated that *underlying dimensions of* *the two instruments are largely correlated as they are correlated with other underlying dimensions in the same instrument*. This provides some evidence of construct validity.

Multiple regression was also employed to establish the relative predictability of the various models and the results are summarised in Table 9.3 (an extracted summary of Table 6.23 and Table 6.26).

Table 9.3: **Relative strengths of different models in terms of their impacts on the judgement on overall quality of teaching as shown in multiple regression (N=76)**

Model and its underlying dimensions with statistically significant impacts	***Unique variance explained in the model***	***Explanatory power of the model in terms of adj. R^2***
In predicting *Overall quality of teaching*		
ISTOF Empirical vs QoT Original Theoretical		**.82**
Effective classroom organisation (QoT)	**8.6%**	
Strategies to enhance learning and lesson focus (ISTOF)	**2.5%**	
Clear Objectives (QoT)	**2.3%**	
Safe & orderly school climate (QoT)	1.2%	
In predicting *Good individual involvement of the pupils*		
ISTOF Empirical vs QoT Empirical		**.85**
Structured teaching skills (QoT)	84.8%	
ISTOF Empirical vs QoT Original Theoretical		**.87**
Stimulating learning climate (QoT)	**8.3%**	
Teaching learning strategies (QoT)	3.6%	

Table 9.3 shows a comparison between the empirical model for ISTOF and the original theoretical model of QoT in predicting the judgment on overall quality of teaching and two comparisons in predicting the global indicator *Good individual involvement by the pupils*, between the two empirical models of the two instruments and between the empirical model for ISTOF and the original theoretical model of QoT. *The results show a dominance of the QoT models in the prediction*, as *Strategies to enhance learning and lesson focus is the only underlying dimensions of ISTOF that showed a significant unique impact*. However, this may also reflect the nature of the two instruments as the QoT sought to make higher inference judgments of quality rather than identifying the frequencies of specific behaviours.

In predicting the judgement on the overall teaching quality of the lessons observed, the crucial teachers' behaviours are related to the dimension *Effective classroom organisation* (i.e., *The teacher gives a well* *structured lesson, ensures the orderly progressions of the lesson, uses learning time efficiently and ensures effective classroom management*). In predicting greater individual involvement of the pupils, the crucial teachers' behaviours are related to the dimension *Structured teaching skills* of the empirical model of QoT or to the dimension *Stimulating learning climate* of the original theoretical model of QoT. As the indicators in the dimension *Structured teaching skills* are mainly indicators of the dimensions *Teaching learning strategies* and *Effective classroom organisation* of the original theoretical model (see Table 6.11 in Section 6.3.1), it can be concluded that *Effective classroom organisation is the single most important underlying dimension in the associations with the two global indicators of overall teaching effectiveness*. Interestingly, the teacher behaviours described in this dimension are very similar to the concept of opportunity to learn in Creemers' (1994) model of effective teaching.

9.3.5 Consistency and variation in teaching behaviours *across samples*

Given the obvious sample and contextual differences between the current study and the earlier English study by Day et al. (2008), it was expected that the empirical models of each study would not be strongly supported in the other study. In particular, the mean scores in the ECP sample are generally higher and most item variances are lower, reflecting the attempt to select effective teachers in the English research. Cross-validation results indicated that the empirical model of ISTOF was less well supported than the empirical model in the ECP data. The ECP sample was found to contribute to only about 30% to the chi-square, when the empirical model was tested, but about 36% when the theoretical model was tested. In either case, the samples still shared a lot in common. Among the various empirical and theoretical models of QoT, the pooled data showed the strongest support for the original theoretical model, though the ECP sample still only contributed only 32% to the chi-square. The goodness-of-fit indices also indicated that the QoT models were better supported in the data than the ISTOF models. These results suggested that the original theoretical models might have stronger external validity than the empirical models and the QoT might be seen as a relatively better instrument than ISTOF for classroom observations in both Hong Kong and England. Interestingly, both these systems have shown inspection base to school accountability.

9.4 Six conclusions from the findings related to previous researches

The present quantitative and qualitative findings on teachers' behaviors in lessons, the wider educational environments and the immediate working ecology in the school and other forms of contexts should be examined and compared in light of the several factors identified in the literature by Muijs and his colleagues (Muijs, Harris, Chapman, Stoll, & Russ, 2004). "These include focus on teaching and learning, effective distributed leadership, creating an information-rich environment, creating a positive school culture, creating a learning environment and a strong emphasis on continuous professional development" (Muijs et al., 2004, p.168). These characteristics do not seem to be uniquely applicable for effective schools in challenging contexts, but also found to be common in other contexts (e.g., see Section 2.3.4 and Table 2.5). This justifies a further review of the present findings in light of the factors identified by Muijs and his colleagues and this led to the following six conclusions.

School's responses to challenges

First, *there are more external and internal challenges than supports for Ming Tak Comprehensive and the school's responses to these challenges may not always be appropriate or effective.* Due to the meritocratic SSPA system and the post-colonial MOI policy, many schools serving socioeconomically disadvantaged areas like *Ming Tak Comprehensive* are in a vulnerable situation because most of their students have lower prior attainment or ability to learn other subjects in English, a strong motive to learn English, and the familial support to develop English language skills. Given the examination-orientation in the senior forms and the competitive market culture driven by parental choice emphasised in the SSPA, school managers tend to favour more in-school setting and assign more teachers who are effective to teach senior form classes. The present study also found that mixed ability was difficult for some teachers e.g., Lucy and Linus. Thus, it is not surprising that the Principal, Mr. Kwong, might prefer to invest the time of his better teachers on those more able students and senior form classes in hoping that his school can become more competitive in the education "market" if his students can do better in the public exams. It is not clear how much the senior management of the school needs, and how much it is preparing, "to compensate for the lack of resources in the pupils' homes" that schools in disadvantaged areas may need to do (Muijs et al., 2004, p.152). Nobody interviewed had seriously considered other alternative strategies to enhance their effectiveness such as establishing transformational leadership or stronger instructional leadership, involving parents in students' learning, building a learning community, creating an information-rich environment, or creating a more positive school culture (see Muijs et al., 2004, for a list of strategies used by improving schools in difficult areas and Gu, Sammons, & Mehta, 2008, for the leadership strategies that enhanced the capacities to enable significant improvement in those schools identified as academically effective and improved schools).

Teachers' responses to challenges

Second, *teachers' responses to the external challenges were generally found to be largely passive and reactive.* For example, teachers like Lucy and Linus preferred streaming as they were less confident in teaching students with mixed abilities. The observations revealed that Lucy indeed taught more effectively once the class composition of her class changed to a streamed group with fewer 'difficult' students. The shadow education system reinforced by the high demand to achieve good public exam results for access to higher education seems to encourage senior form teachers to adopt a teacher-centred approach that emphasises examination skills in their teaching, relies more on L1 in the lessons supposed to be learning in English, and reflects pedagogical approaches by private tutors*. Because teachers are held accountable for the student outcomes,* even more effective teachers such as Charlie may feel lack confidence in using a more student-centred approach for senior forms, because many students expect them to teach more like the private tutors. *Teachers seem to allow outside influences such as the student* *and parent expectations or the markets to influence their pedagogy.* None of the teachers who taught Form 5 seemed to be able to achieve high level of student engagement, even though these students were supposed to be more manageable. However, it might help them pass if that is what private tutors also emphasise.

Relative strength of teaching and learning focus

Third, *the current focus on teaching and learning did not seem to be strong enough in Ming Tak Comprehensive to turn things around.* It might be an unavoidable disadvantage for the school to be a Band 3 CMI school. There is a contradictory system-wide expectation for its English teaching and learning, which exists in the SSPA system, the MOI policy and the public examination requirements. That is, the education authority and parents implicitly have set a lower aim in English learning for the junior form students in a CMI school like *Ming Tak Comprehensive*, but their aim for these students in the public examinations is as high as that for their counterparts in schools of higher bands. English teachers of a Band 3 school are somehow expected to reduce the wide gap existed between the low expectation for junior form student outcomes and the high expectation for senior form student outcomes. Teaching in a Band 3 school means that teachers have to counter a stronger impact of low attaining intake school attempting to exceed the "normal efforts" (Maden, 2001; Maden & Hillman, 1993). In order to graduate at a higher percentile after five years, most of the students in *Ming Tak Comprehensive* whose academic abilities were among the bottom one-third in the primary schools are likely to need some exceptionally effective teaching. Unfortunately, according to Charlie, the school has not created a positive school culture and learning community. Even compared with similar schools, *Ming Tak Comprehensive'*s value-added results are below average.

Due to generally low expectations in *Ming Tak Comprehensive*, the focus on students' academic achievement in the junior forms was not strong compared expectations for the senior forms. The wide gap in the learning targets between the junior forms and senior forms indicate that the English teachers have not integrated the curriculum across forms (or grades). Lucy expected her new Form 1 class to have stronger English proficiency than her

Form 2 class given her higher usage of L2 in instruction. Sally explicitly admitted her Form 2 and Form 3 class had better English than her Form 5 class. Linus claimed that his Form 5 students might only have primary school proficiency in English. The inconsistent and low scores in the dimensions of observed teaching practices for Lucy and Linus may undermine the overall teaching effectiveness of the English department. The results suggest they had not structured their lessons appropriately in order that students could progress with a succession of higher and higher targets. Besides, Mr. Kwong's emphasis on deep learning was well supported by research (e.g., Biggs, 1988; Biggs & Collis,1989; Bowden, 1988; Marton & Saljo, 1976), but Hattie (2009) argued that learners need both surface and deep learning and the key is to balance them in a context or set of domain knowledge. It may be critical for the EFL teachers to teach in accordance with the learning characteristics of the Chinese learner, even though the subject is a foreign language.

Relative strength of leadership

Fourth, *instructional leadership was not strong enough in the school management*. It was found that the quality of pedagogy and the achievement of students would improve substantially only when both transformational and shared instructional leadership coexist in an integrated form of leadership in its influence on school performance (Marks & Printy, 2003). However, neither the school principal nor the head of department, Charlie, perceived themselves as transformational leaders. Rather, they tended to believe that they had qualities of instructional leadership and distributed leadership as they are willing to keep their minds open regarding instruction and to share leadership with junior colleagues. For example, Charlie thought that he is able to communicate with the English teachers openly and equally. However, Muijs et al. (2004, p.170) noted that what type of leadership is needed is dependent upon the existing phase of the school and its strengths and weaknesses:

> ….. while strongly distributed leadership sees to characterise effective schools, it may be that those that are in an early phase of improvement may need more forceful top-down methods to set the basics in place, as suggested by both

> contingency theory (fit to circumstances) and the compensatory model….. Schools that are in an early phase of improvement, or who appear to be failing, may need a lot of external support, strong leadership, and a focus on the basics, and may not be in a position to get parents to be strongly involved in the school.

Thus, whether Charlie can successfully enhance the overall teaching effectiveness of the English department in future depends on whether he can lead his colleagues like Lucy to manage diverge class compositions and help teachers like Linus to minimise the variation in his teaching effectiveness. This may require qualities of transformational and instructional leadership more than those of distributed leadership to initiate improvement programs in teaching and learning in the school. Similar findings have been elaborated by Day et al., 2009 in relation to school context, organisational structure and the kinds of leadership needed to bring about improvement.

Employment of data

Fifth*, lack of data/information richness seems to be a more serious problem than inadequate leadership in Ming Tak Comprehensive.* Both Mr. Kwong and Charlie did not mention using data to inform their work. Their practices contrasted the emphasis of using a broad variety of data in school improvement in the quality assurance framework in Hong Kong (EDB, 2009c). Recent research also indicated that schools in challenging contexts had to rely more on data to enhance teaching effectiveness (Muijs et al., 2004; Gu et al., 2008) and that strong leadership in schools showed utilising data to devise strategies for action (Day, 2004; Day et al., 2009; Mulford, Silins, & Leithwood, 2004). Since patriarchal leadership was still common in school management (e.g., the last school of Lucy and Linus, according to Linus), no interviewees mentioned about the possibility of collective leadership shared among key stakeholders though its effects on student achievement is generally evident in the western countries (Day et al., 2009; Leithwood & Mascall, 2008; Robinson, Lloyd, & Rowe, 2008).

What Charlie lacks seems to be an inquiry mind on what works and what does not work in his school. For example, Charlie reported that they had done many things to improve the junior form teaching, but he thought they had achieved very little. On the contrary, from the classroom events in

Sally's junior form lessons and Lucy's Lessons 14 and 15, students were fully engaged in the learning activities. Sally's students also reported that they were learning English more and better than they had been in the primary schools. Thus, Charlie seems to underestimate what the teachers and students have achieved in the junior forms when compared with the classroom observation results and students' reports. By contrast, Linus and Sally seemed to understand their students the junior form students better. Linus realised the gap between the formative and summative assessments might have discouraged students from learning as they could not see their efforts would succeed. He was not pleased with the limited function of summative assessments in the school as they served to give only negative feedbacks to students. Linus also chose to assign homework regularly as a means to promote reflective learning and monitor students' progress. Charlie probably may not have realised that Linus's strength in the dimension *Strategies to enhance learning and lesson focus* was so much contingent on the school's new policy. Sally commented that their students gradually learned helplessness again in *Ming Tak Comprehensive* as they had done in their primary schools because of low attainments in English. Therefore, she rewarded her students in different ways as she recognised her students needed her recognition as positive reinforcement for their learning. Lucy also positively reinforced her students with approving smiles, changes in MOI, and the use of more demanding tasks. If more frontline teachers are as perceptive as Linus and Sally and were consistent in implementing the school policy on homework as Linus, *Ming Tak Comprehensive* might improve its quality of student experiences and its effectiveness.

Impacts of professional development

Sixth and finally*, as yet continuous professional development in Ming Tak Comprehensive has not demonstrated sufficient positive impacts in teaching and learning.* Certainly, enhanced professionalism in English teaching has been evident in the government's higher demands for quality of teaching (e.g., required relevant qualification, specific language benchmark tests, and ongoing professional training requirements). Nonetheless, department heads like Charlie acknowledge that some teachers are still weak in basic classroom management to the extent that the quality of teaching in the junior forms is poor because some teachers are spending valuable teaching and learning time on trying to control behaviour and using ineffective strategies such as excessive punishment creating a negative climate (see Section 2.4.4 for similar problems found in the effective school in Chen et al., 2004). Without purposeful teaching, teachers cannot expect students would take up their responsibilities in learning if their rights are not respected on the same regard. In his explanation for some teachers' authoritative approach in junior form lessons, Charlie declared that professional development and training is a necessary, but not the sufficient condition for enhancing teacher effectiveness, because the influence of professional development seems to weaker than those deeply rooted in teachers' upbringing. All teachers reported that they would reflect on their teaching in lessons where their own standards were not met and did attempt to make adjustments in the upcoming lessons. However, none of them reported tackling the problem systematically through action research or using strategies such as peer observation and joint planning, though they did share teaching materials for a common curriculum and teaching schedule. The worksheets were found unsuitable in Sally's Lesson 7 (see Section 7.3.4) and the co-teaching in various Linus' lessons showed little collaboration. Moreover, lack of consistent expectations and support by the school and the educational authority for junior and senior form teaching, lack of awareness and policy to tackle variation in teacher effectiveness among teachers, and lack of the intention to apply knowledge into practices and reflect practices based on evidence might serve to have reduced the overall teaching effectiveness of the English department at *Ming Tak Comprehensive*.

9.5 Limitations of the study

This study is subject to many limitations including sample selection, methodology, data collection, analysis and interpretation. These limitations and some recommendations for future work can be summarised as follow.

9.5.1 Sample selection

The present sample consists of only four EFL teachers in one department in one school in Hong Kong, whose teaching behaviours might not represent the teaching practices found in other teachers, in other departments, and/or in other schools. However, focusing on teachers teaching the same subject in one school also has its own advantage. It was expected that interpretations of variation found between such teachers would be more easily to identify influences related to the unique contexts of the department and of the school.

The decision to study the English department in an underperforming school reflected a personal as well as academic interest. It was personal as I had been an EFL teacher for fourteen years in an underperforming school, the findings could be important for my ex-colleagues and me to understand their teaching and the challenges faced better. A strong academic interest in schools in challenging contexts has emerged in the U.K. (e.g., Chapman & Allen, 2006; Chapman et al., 2006; Harris et al., 2003, 2006; Muijs et al., 2004), but this topic is rarely studied in Hong Kong. The multilevel case study in the present research can thus contribute to our knowledge of the challenges of those teachers who work within in the less effective schools in the socio-economically disadvantaged areas in Hong Kong. In these schools, teaching English has been particularly challenging since the new MOI policy introduced in 1998 because the total English inputs to which students could exposed in schools sharply declined and students in these schools generally lack the ability to command the language and the familial support to develop the language skills in English.

The accounts on the English department, *Ming Tak Comprehensive*, and the educational system in Hong Kong were brief and selective. They were used to provide background information and raise awareness of context. There was always a dilemma for a researcher to include more data that may enrich the descriptions of a case. School documents, internal examination results, departmental minutes, and the government policy documents were available, but analysing all these additional data would be too much to an extent that it would shift the focus away from that chosen for the research study, namely, consistency and variation in teachers' classroom practices. To pursue a scrutiny for this issue, I considered that it was practical and reasonable to limit the scope of the triangulation of the data to quantitative observation scores, field notes and interview transcripts. Howeever, any future work to expand and knowledge base of teachers' classroom practices can be benefited if more teachers in different contexts are included in the case spectrum.

9.5.2 Sample size

Most of the quantitative analyses were based on the lesson as the unit of analysis. A sample size of only 76 lessons remained rather small for doing factor analysis, multiple regression or discriminant function analysis. The problems related to small size are well-documented and discussed (e.g., Marsh & Hau, 1998). As the sample size for each teacher was unequal and underlying dimensions were highly correlated, there were problems like violation of equal variances and multicollinearity. However, these problems were inherent in the data that could not be easily resolved simply by increasing the sample size. Moreover, these problems do not seem to have affected the statistical significance of findings much, as both the liberal and the more restrictive measures used (e.g., normal F test vs. Brown-Forsythe F statistic for used when group variances were unequal in ANVOA) in the same analyses often produced similar results.

There was an attempt to keep the lesson sample size close to 100, but the planned 5–day observation period did not work out due to an interruption of a missing school day for the arrival of a typhoon. The original plan to observe all the teachers teaching Form 5 was also revised, when one teacher withdrew her initial consent and Charlie, the department head, was therefore observed instead. There were also several lessons planned for observation, but eventually unobserved because of the interruptions of some school functions like a field trip, school mass, and form test. This reflected that one could not easily estimate opportunity to learn in a school just by calculating the official school days and the official school time-tables. Certainly, one could consider extending the observation period, but the total number of lessons observed for an individual teacher may have reached the limit that s/he would feel comfortable. As it was not easy to recruit teachers to participate in a doctorial research in Hong Kong due to their heavy workloads, I could not increase the sample size as much as I wished.

9.5.3 Insufficent discriminant validity for the underlying dimensions

Although the present study succeeded in identifying several underlying dimensions of teaching, they generally lacked adequate discriminant validity. Further work is needed to deal with this problem. In Section 4.4, it has been suggested that the model could be further developed by adding in some second–order variables to account for the high intercorrelations found between some factors. Another possibility is suggested in a recent study by Malmberg, Hagger, Burn, Mutton & Colls (personal communication), who tried to measure classroom quality in terms of four aspects: emotional support, instructional support, classroom organisation, and student engagement. However, these researchers used a different classroom observation instrument. This suggests that for some instruments, discriminant validity might be more difficult to establish. Thus, we may have to refine these observational instruments such that more distinctive dimensions can be generated.

It should be also noted that underlying dimensions of effective teaching may be by nature less distinctive as constructs in psychometric tests if certain observed behaviours are perceived to be related to different aspects of teaching. It may also be the case that for highly effective teachers or for highly effective lessons, different dimensions of teaching may show strong correlations as it was found in Sally or Charlie and their lessons. This means that when the sample selection is more restrictive, it would be more difficult to establish discriminant validity. Again, this implies the sample spectrum of future research should be as broad as possible when resources permit.

9.5.4 Case study and insider research

While the selection of the sample reflected the difficulty to get access to school, the case study methodology was also limited. Compared to Tsui's (2003) study of four ESL teachers in Hong Kong, the present case study was not data-rich, because she collected her data in three months, video-taped and transcribed all the lessons[94]. The present interviews were also not extensive enough for an in-depth address on issues such as teacher knowledge, pedagogical decisions, professional development, enactment of the curriculum, teacher-student relationship, and teacher collaboration. In other words, the potential of case study methodology might have not been fully utilised. However, a case study approach was considered compatible with the MM approach and the researcher's previous background as an insider of the school.

As a piece of insider research, this study was subject to the limitations addressed in Section 3.6. I attempted to minimise researcher bias by applying a professional critique to the context within which the teachers and other staff were situating, rather than focusing on the teachers as the source of 'inconsistent teaching behaviours' or 'ineffective teaching'. Observing several teachers in at least 15 lessons each seems to provide a more objective lens to study individual teachers than studies relying on one or two lessons without referencing to their immediate or greater contexts. While effectiveness identified in classroom observation cannot be equated with valued-added effectiveness, it is more easily attributed to what the teachers had done and more readily understood by the practitioners. Although value-added statistics have been available to *Ming Tak Comprehensive* for several years, teachers generally expressed that they could not inform them much other than their relative ineffectiveness in comparison with other schools.

9.5.5 Scope and nature of data

A major limitation of this study was certainly the lack of objective student outcome data. Actually, this concerned the relevancy of the data and the objectivity of the data. Without the student level data, I could not evaluate the value-added teaching effectiveness of the teachers by relating their observed teaching behaviours with the academic outcomes of their students, the most widely accepted type of student outcome data in the SER and TER. Some of the academic outcomes of the school in the public examinations

94 It should be noted that as a professor and the teacher of her participants, Tsui (2003) received more resources and support from her students and their schools.

were actually available in its school profile, but it only reflected the general performance of the school and the English department. However, these results were less relevant for evaluating the performances of individual teachers as they were not the only teachers in the department. Like most schools, *Ming Tak Comprehensive* did not have value-added data at the teacher level. Instead, the overall teaching effectiveness was assessed by the global judgments of the teaching quality of the lesson observed and the overall individual involvement by the students observed in the lesson.

Based on the QoT instrument, my judgment as a researcher might not be biased and unreliable, because I was trained to use the instruments and had received postgraduate level professional training in classroom observation and school inspection. However, objectivity remains an issue I could not explore as ratings could not be compared and assessed by judgments of other raters in this single researcher study. In other words, both the numerical item scores and the global indicator ratings should be viewed as high-inference and subjective by nature.

It has been argued earlier in Section 3.5.3 the advantages of using instruments developed for application in international contexts. However, these instruments are still under development and piloting in various countries. The results presented in Chapters 4 and 5 indicated the lack of discriminant validity for the factors, suggesting that further research is required to address and establish multidimensionality of teaching (see Section 2.3.5). Multiple regressions results in Chapter 6 showed the relative strengths of the instruments in terms of the associations between the factors identified and the indicators of overall teaching quality. These instruments also differ in the inclusion/exclusion of certain aspect(s) of teaching. For example, ISTOF does not categorise items in terms of lesson objectives, while the current results indicated that clear objectives were strongly associated with overall quality of teaching (see Table 6.23). Accordingly, it is doubtful whether these instruments are sensitive enough to capture some unique aspects of teaching in Hong Kong contexts. For example, it may not be able to measure those teaching behaviours relevant to the traditional concept of teacher and teaching in Confucian philosophy as discussed in

Section 2.5.4. This also suggests that research relying on using one single instrument may be subject to biases and further research is required to compare different instruments, like the present study, and to compare them in different contexts.

9.5.6 Quantitative and qualitative synthesis

Given the scope and nature of the data, the analyses conducted were predominantly quantitative. Qualitative data were mainly used to confirm the quantitative findings (e.g., to describe the classroom processes of lessons of some teachers tended to be consistently scored highly in most underlying dimensions) and also to explore and help to explain the quantitative findings. The qualitative data were also used to identify themes that required further quantitative analyse (e.g., to explore whether junior form teaching and senior form teaching are different). Some quantitative data were also qualitised to characterise teacher effectiveness and contexts. Accordingly, the research design of the present study did not quite fit the exemplars of MM designs proposed by Creswell and Plano Clark (2007). This indicated the use of quantitative and qualitative data can be more dynamic and changed during the research process.

Merriam (1998) argued that all qualitative analyses are filtered through the investigator’s world views, values and perspectives. It is unlikely that an MM study involving qualitative analyses like the current research could be exceptional. Given my past working experience, I might be more empathetic to the difficulties in teaching English in *Ming Tak Comprehensive* or schools in similar settings. Some researchers from non-Confucian backgrounds may be more ready to be amazed by the outstanding student outcomes of Hong Kong, Singapore, Taiwan, or Japan in international investigations like PISA and TIMSS. They probably would not agree with me that a common curriculum, a meritocratic school places allocation system, a competitive educational system, an examination-oriented culture, a post-colonial MOI policy, and an one-size-fits-all type of public examinations are necessarily always negative challenges to a disadvantaged school. Challenges and contradictions are not value-free terms as they imply that some contexts or circumstances are more desirable.

Despite the above limitations, this piece of research can be considered a unique effort to investigate the relationship between teaching effectiveness and classroom practices with methodological and theoretical significances.

9.6 Significance of findings and their implications of for future researches

9.6.1 Implications for Teacher Effectiveness and Professional Development Research

There are two major interrelated contributions of the present research. 1) Providing new evidence about the multi-dimensionality of teaching effectiveness and intensifying the relative significance of different dimensions of teacher practices as prediction of two forms of observed teaching effectiveness, namely, a global indicator of teaching quality and an indicator of student engagement; 2) Investigating the evidence for alternative theories of teacher effectiveness, generic and differentiated, and showing how these may be viewed as comparable rather than opposed. The GTE perspective has been well supported as the distinctive dimensions of teachers' behaviours identified across lessons largely corresponded with the theoretical ones. Different cases lent different degrees of support to the GTE concept.

Consistently high scores across different dimensions and consistency across lessons by Sally provides an example of what may be termed as an 'expert' teacher who excels in all aspects of teaching at all times. By contrast, Lucy showed the same consistent observed practices across different dimensions but generally obtained low scores, but not always across lessons or contexts. The DTE theory would suggest that Lucy should not be labeled only as an ineffective teacher as her observed practices seen to be much affected by the contexts of the classes she taught. Charlie is another case to illustrate how a teacher may not fit the GTE's standard of being effective in all observed dimensions and across different lessons.

Linus' patterns of results are in accord with the GTE's classification of an ineffective teacher. However, like Lucy, Linus' teacher effectiveness could be enhanced as the data indicate that he was responsive to the demands of the school. His role as a SEN teacher appeared to have affected his classroom management approach negatively. It was possible that Linus' multiple and expanding roles might have had a negative impact on his overall teaching effectiveness. As the DTE theory suggests, teaching ineffectiveness may not be best viewed as a generic characteristic of ineffective teachers, but rather as an outcome that is affected by the teaching contexts and expectations, including the changing roles of teachers. Certainly, evidence based on a single case of Linus' teaching patterns may not be sufficient to make a strong claim on this. However, Sally is a counter-example to indicate that some teachers can still be effective despite changes in contexts and expectations.

Thus, it would be fruitful for future research to explore further the impacts of contexts and roles on different dimensions of teachers' behaviours and their relations to traditional measures of student outcomes. There is a need to employ additional measures such as student progress and their self-reported social and affective outcomes. The focus group interviews with students in the later phase of the study could be expanded and structured and other student level data should be collected in future. The current findings need to be confirmed using different samples like teachers of different subjects or schools of different bandings. That is, different areas of the GTE and DTE have to be explored in future (Muijs et al., 2005). This study has provided some comparisons between results for the Hong Kong and the English samples, but cross-cultural and comparative issues are certainly worth much deeper further explorations.

9.6.2 Implications for Educational Effectiveness and School Improvement Research

Different forms of teacher feedback and professional development, including action research may assist teachers to develop strategies to minimise their variability across different dimensions and sustain their teaching effectiveness across contexts. Peer observations and feedback and lesson study are also feasible systematic strategies for teachers to enhance teachers' understanding their strengths and identify areas needed improvements. In other words, the link between TER and professional development research should be addressed in future studies.

The current research has contributed to EER by studying cross-level interactions (between school, department, and class) in a multilevel case study. This is an alternative strategy to study multilevel educational effects when the researcher is confined by limited resources and the need for accessible data. Creemers and Kyriakides' (2008) DEE model is inspiring but it also has posed a heavy burden for researchers to verify it with adequate empirical evidence. Cross-level effects are particularly difficult to measure quantitatively and sometimes their effects are only explicit on the occasions when there are some new governmental policies such as changes in public examination requirements (e.g., Cheng, 1998, 2004, 2005; Cheng & Falvey, 2000), curriculum innovation (Chen, 2006), and changes in pedagogy (like MOI, Tsang, 2004; Yip, Tsang & Cheung, 2003). This may explain why qualitative descriptions and inferences proved valuable in enriching this multilevel case study by examining cross-level interactions in terms of teachers' perceptions and accounts in interviews.

Studying teachers teaching parallel classes of the same form in the present study has allowed me to illustrate differences in observed teacher practices and draw conclusions about teaching and teacher effectiveness. Some researchers like Luyten and de Jong (1998) have argued that differences between teacher effects on student achievement would be small and limited to classroom instruction when the content and goals of instruction were controlled through coordination efforts. The current results, however, suggest that the differential teacher effects could be large. Marzano's (2003) probability model educational effectiveness was used to illustrate the relative impacts of differential teacher effectiveness. Recently, Sammons and Luyten (2009) have examined different methodological approaches to explore *schooling effects* and *school effects.* The current multilevel case study design has provided qualitative evidence that attributes departmental and schooling effects to the teachers by examining variation in their individual teaching behaviours and indicators of their teaching effectiveness in different forms (or grades) as well as by examining their collective teaching behaviours and teaching effectiveness in different year groups and different forms. Triangulating these results with themes in the interview has led to a range of conclusions in Section 9.3.3. These findings have strong implications for school improvement strategies in the school observed and these may be applicable to other similar schools in challenging contexts.

Given the specific characteristics of the educational system in Hong Kong, future research on EER should employ multilevel latent curve growth models to explore differential school effects among schools of different bandings (i.e., different ability-compositions in the intakes), as Palardy (2008) has recently done in his study of the differential school effects among schools with different social class compositions. In order to conduct this type of research, stronger collaboration between the academics and the school personnel would be needed. However, the latter generally lack the expertise to handle the abundant amount of data they can collect through formative and summative assessments. These are the longitudinal data that may be used to explore the schooling and teacher effects clearly. Hong Kong is particularly more advanced on this regard as every school is provided with tools to collect student level emotional and social outcomes as well as platforms to collect stakeholder views and contextualised value-added data. Although schools are data rich, in most cases they do not know how to use such data for school improvement. The present research results were considered valuable by the school personnel because they have provided them with rich descriptions of practices and analysis of observations that inform their practice. Moreover, at present it is very difficult for researchers to gain access to school level value-added data, in contrast to other systems such as England, where such data have frequently made available to the research community in the later decade.

9.6.3 Implications for Mixed Methods Research

As an example of MM research, the present study has sought to demonstrate four essential qualities: synergistic by findings, non-linear by research process, richness-driven by choice of evidence, and unlimited by domain. It has been argued that MM designs can offer a *richer evidence* *base* that would thus be more fruitful in promoting new understandings and contributing to knowledge that would inform policy and practice than studies relying on approaches from either the quantitative or qualitative research paradigm alone and that MM research approach allows a study to address new research questions more suited to the study of complex social institutions such as schools that could not be explored by reliance on a single paradigm alone (Sammons, 2010). The present research has achieved this aim by combining statistical prediction and explanation of variation in teachers' behaviours observed and rated with two instruments, with other case study data including extensive qualitative field notes and subsequent in-depth semi-structured interviews that provided rich descriptions of practice and explored teachers' own perceptions and understandings to supplement the development of both empirical and theoretical inferences and enhanced understandings.

The current multilevel case study did not completely conform to existing classifications proposed by key theorists in the MM research area (see Chapter 3 for details). This was not a shortcoming of the present research for not falling in a particular existing category, but an inevitable consequence if the MM research is going to enable the development *of new synergistic understandings* (Day et al., 2007; Day, Sammons & Qu, 2008; Sammons, 2010). This was considered a flexibility that made the research process dynamic and dialectic. For example, the four emerging research questions were raised at different times during the qualitative and quantitative analyses. The process was both exploratory and explanatory. On the one hand, although the original analyses showed a stronger emphasis on the quantitative data, these quantitative data were later qualitised in the case studies to characterise the profiles of individual teachers. On the other hand, while the qualitative field notes help to validate the quantitative CFA findings, the challenges and paradoxes identified in the analyses of the interview transcripts led to the later originally unanticipated explorations using multiple regression analyses and discriminant function analyses. These later analyses also contributed to profiling individual teachers and cross-case analyses.

Throughout the analysis process, the researcher had to triangulate data across different levels without attributing a fix weight to data just by their nature. The weight of the data should be evaluated on the basis of their relevance to the research questions, rather than its nature or amount. Only when data, and thus evidence, are evaluated by their relevance, can the researcher *reveal their richness*. The new emerging research questions also indicated that *the research process was not linear, but a dialectic exercise* in which the researcher had to constantly search for evidence for emerging themes during the integration of analyses and interpretations. The researcher cannot ignore the emerging questions as they "*inform and support closer links with applied research* and evaluations that can promote effective school improvement initiatives and teacher development programmes" (Sammons, 2010; Teddlie & Sammons, 2010). By creating individual teacher case studies and examining cross case studies the researcher sought to produce trustworthy accounts and enhanced understanding of variation in teachers' observed practices in a single EFL department in a school in a challenging context. This is seen to enhance the contribution to theory testing and generating and provide evidence of relevance and value to practitioners.

9.6.4 Implications for Classroom Observation Research

This research has contributed to classroom observation research by using systematic classroom observation instruments focusing on studying teachers' behaviours in the classroom and by comparing two instruments. Studying teachers' behaviours in classroom observation has "the advantage of possibly being more objective due to the outsider's perspective" (Muijs, 2006, p.58). Employing two instruments has enabled exploration of the different underlying dimensions implied in the instruments. Variation in classroom observation instrument is an important issue that has not received little attention in the existing TER and SER literature. It has been found and argued that generialisability across occasions using eight classroom observation instruments was poor (Calkins, Borich, Pascone, Kugle, & Marston, 1997), this problem was somewhat reduced in this study by observing a teacher at least 15 times concurrently with two instruments and using field notes and interviews to supplement and enrich the quantitative data and to create individual teacher case studies.

Certainly, lessons can be observed without using systematic classroom observation instruments (e.g., see some examples in Hargreaves & Woods, 1998). The few texts on classroom observation instruments (Croll, 1986; Muijs & Reynolds, 2005; Wragg, 1999) have only introduced a very few of the range of instruments available in the literature (see Simon & Boyer, 1967-1970; Borich & Madden, 1977; Galton, 1978) and treated the instruments mainly as tools for professional development. In contrast, the development of the ISTOF and QoT instruments has respectively represented the academic and inspectorates' interests in creating instruments for classroom observation as well as the different theoretical orientations (e.g., frequency vs. effectiveness of particular teacher behaviour) implied in their instruments and models.

The current research is significant as it has compared the instruments used by international researchers differed by orientations. The findings tend to suggest that QoT may be more useful as it highlighted the importance of *Effective classroom organisation* and its empirical and expanded variant, *Structured teaching skills*. The results also indicated that its new and more parsimonious theoretical structure might not be necessarily superior to the original one and more research is needed to explore the various factors of the instrument. The current findings show that QoT is more value-oriented than ISTOF and thus may be more compatible with, and a better predictor of, the global indicator of overall teaching effectiveness. The strength of ISTOF may lie in its attempt to distinguish teaching behaviours into theoretically important domains, but the number of items for each domain has not been balanced for adequate testing.

Recently, Creemers and Kyriakides (2008) developed new scales to test their Dynamic model of educational effectiveness. Rather than comparing different aspects of teaching simultaneously in a single instrument, they tested each in terms of five theoretical factors (i.e., frequency, stage, focus, quality and differentiation). Their work and their instruments have represented a completely different conceptualisation about how different aspects of teaching could be measured. This would be another major contribution if future research can compare their instruments with those such as ISTOF and QoT and examine their applications in different cultural contexts and for different phases of education.

9.7 Concluding remarks

I have intended to tell a story with numbers in this piece of research. Numbers are meaningless by itself without appropriate interpretation, but a pure story without numbers would also involve limitations. I have tried to use MM to address a range of research questions about quantifiable variation in teachers' observed classroom practices via a multilevel case study of four teachers of a single department of a school and tell a story about the challenges of teaching English as a foreign language in a particular context, namely, a less effective department and low attaining school in Hong Kong.

REFERENCES

Adams, G., & Engelmann, S. (1996). *Research on Direct Instruction: 25 Years beyond DISTAR.* Seattle,. WA: Educational Achievement Systems.

Anderson, R., Greene, M., & Loewen, P. (1988). Relationships among teachers' and students thinking skills, sense of efficacy and student achievement. *Alberta Journal of Educational Research, 17*, 86-95.

Antoniou, P. (2009, Sept). *Using the dynamic model of educational effectiveness to improve teaching practice: Building an evaluation model to test the impact of teacher professional development programs*. Cyprus: Unpublished PhD Thesis.

Ashton, P. T. (1985). Motivation and the teacher's sense of efficacy. In C. Ames, & R. Ames (Eds.), *Research on motivation in education: Vol. 2. The classroom milieu* (pp. 141-174). Orlando, FL: Academic Press.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York, NY: Longman.

Askew, M., Rhodes, V., Brown, M., William, D., & Johnson, D. (1997). *Effective teachers of numeracy: Report of a study carried out for the teacher training.* London: School of Education, King's College, London .

Aubrey, C. (1997). *Mathematics teaching in the early years: An investigation into teachers' subject knowledge.* London: Falmer Press.

Auerbach, E. (1993). Reexamining English only in the ESL classroom. *TESOL Quarterly , 27*, 9–32.

Bailey, C. (2005). “The UK lecturers don’t teach me anything”: Chinese students’ expectations of their teachers and implications for UK HE providers. *Conference Proceedings: 2nd Annual Conference: The Chinese and South East Asian Learner: The Transition toUK Higher Education: September 2005.* (pp. 1-12). Southampton: Southampton Solent University.

Beard, R. (2000). *National Literacy Strategy. Review of Research and other Related Evidence.* Suffolk: DfEE .

Bennett, N. (1976). *Teaching styles and pupil progress.* London: Open Books.

Bennett, N. (1988). The effective primary school teacher: the search for a theory of pedagogy. *Teaching and Teacher Education, 4*(1), 19-30.

Bickel, E., & Bickel, D. (1986). Effective schools, classrooms and instruction: implications for special children. *Exceptional Children , 52*(6), 489-500.

Biggs, J. B. (1988). The role of the metacognition in enhancing learning. *Australian Journal of Education, 32*, 127-138.

Biggs, J. (1998). Teaching across cultures. *International Conference on the Application of Psychology to the Quality of Teacher and Learning.* Hong Kong: The University of Hong Kong.

Biggs, J., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy.* New York, NY: Academic Press.

Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment: using the SOLO Taxonomy. *Australian Journal of Education, 33*, 143-161.

Bloom, B. (1976). *Human characteristics and school learning.* New York, NY: McGraw Hill.

Borich, G. (1996). *Effective teaching methods* (3rd ed.). New York, NY: Macmillan.

Borich, G., & Madden, S. (1977). *Evaluating classroom instruction: Sourcebook of instruments. Final Report.* Austin, TX: Research and Development Centre, University of Texas.

Borman, G., Hewes, G., Overman, L., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*, 125-230.

Bosker, R. J., & Guldemond, H. (2009). School effects on students' progress: a dynamic perspective. *School Effectiveness and School Improvement , 20* (2), 255-268.

Bosker, R. J., & Scheerens, J. (1989). Issues in the interpretation of the results of school effectiveness research. *International Journal of Educational Research, 3*(7), 741-751.

Bosker, R. J., & Scheerens, J. (1994). Alternative models of school effectiveness put to the test. *International Journal of Educational Research , 21* (2), 159-180.

Bourdieu, P., & Passeron, J.-C. (1990). *Reproduction in education, society and culture.* (2nd ed.). London: Sage.

Bowden, J. (1988). Achieving change in teaching practices. In P. Ramsden (Ed.), *Improving learning: new perspectives.* (pp. 255-267). London: Kogan Page.

Brannen, J. (1992). Combining qualitative and quantitative approaches: an overview. In J. Brannen (Ed.), *Mixing methods: Qualitative and quantitative research.* (pp. 3-38). Aldershot: Avebury.

Brannen, J. (2005). Mixing methods: The entry of qualitative and quantitative approaches into the research. *International Journal of Social Research Methodology, 8*(3), 173-184.

Bray, M. (1999). *The shadow system: private tutoring and its implication for planners.* Paris: UNESCO.

Bray, M. (2006). Private supplementary tutoring: comparative perspectives on patterns and implications. *Compare: A Journal of Comparative and International Education, 36*(4), 515-530.

Bray, M., & Kwok, P. (2003). Demand for private supplementary tutoring: Conceptual considerations, and socio-economic patterns in Hong Kong. *Economics of Education Review, 22* , 611-620.

Brewer, J., & Hunter, A. (1998). *Multimethod research: A synthesis of styles.* Newbury Park, CA: Sage.

Brimer, A., Johnson, K., Leung, J., Sweeting, T., Ip, B., Tong, S. M., et al. (1985). *Effects of the medium of instruction on the achievement of Form 2 students in Hong Kong secondary schools.* Hong Kong : University of Hong Kong and Education Department.

Brookhart, S. (1997). Effects of the classroom assessment environment on mathematics and science achievement. *Journal of Educational Research , 90*(6), 323-330.

Brophy, J. (1992). Probing the subtleties of subject matter teaching. *Educational Leadership, 19*, 4-8.

Brophy, J., & Good, T. (1986). Teacher behaviour and student achievement. In M. Wittrock, *Handbook of research on teaching* (3rd ed., pp. 328-375). New York, NY: Macmillan.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York, NY: Guilford.

Bryman, A. (1988). *Quantity and quality in social research.* London, UK: Unwin Hyman.

Bryman, A. (2006a). Integrating quantitative and qualitative research: How is it done? *Qualitative Research* , 97-113.

Bryman, A. (2006b). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology, 9*, 111-126.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research, 1*(1), 8-22.

Bunton, D., & Tsui, A. B. (2002). Setting language benchmarks. *Journal of Asian Pacific Communication ,12* (1), 63–76.

Calkins, D., Borich, G. D., Pascone, M., Kugle, C. L., & Marston, P. T. (1997). Generalizability of teacher behaviors across classroom observation systems. *Journal of Classroom Interaction, 13*(1), 9-22.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin , 56*, 81-105.

Campbell, J., Kyriakides, L., Muijs, D., & Robinson, W. (2004). *Assessing teacher effectiveness: Developing a differentiated model.* Abingdon, UK: RoutledgeFalmer.

Carroll, J. (1963). A model of school learning. *Teachers College Record, 64*, 723-733.

Carroll, J. (1989). The Carroll Model: A 25-year retrospective and prospective view. *Educational Researcher, 18*(1), 26-31.

Carroll, J. B., & Spearritt, D. (1967). *A study of a model of school learning (Monograph No. 4).* Cambridge, MA: Center for Research and Development in Educational Differences, Harvard University.

Chan, C.-C., Chan, K.-Y., Cheung, W.-M., Ngan, M., & Yeung, V.-M. (1992). Primary school teacher self concept: Its relationship with teacher behaviors and students' educational outcomes. *Primary Education, 3*(1), 9-28.

Chapman, C., & Allen, T. (2006). Collaborative reform for schools. *Improving Schools , 3*, 291–301.

Chen, L.-M. (2006, Summer). *Washback effects on curriculum innovation.* Retrieved on Feb 22, 2009, from Academic Exchange Quarterly, 6: http://findarticles.com/p/articles/mi_hb3325/is_2_10/ai_n29284295/

Cheng, L. (2005). *Changing language teaching through language testing: A washback study.* Cambridge: Cambridge University Press.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing.* (pp. 147-170). Mahwah, NJ: Erlbaum.

Cheng, L. (1998). The washback effect of public examination change on classroom teaching: An impact study of the 1996 Hong Kong Certificate of Education in English on the classroom teaching of English in Hong Kong secondary schools. *HKU Theses Online, Curriculum Studies: Theses* . Hong Kong: University of Hong Kong.

Cheng, L., & Falvey, P. (2000). What works? The washback effect of a new public examination on teachers' perspectives and behaviors in classroom teaching. *Curriculum Forum , 9*(2), 1-33.

Cheng, Y. C. (1995). *Function and effectiveness of eductaion* (3rd ed.). Hong Kong: Wide Angle Press.

Cheng, Y. C. (1996). *School effectiveness and school based management.* London: Falmer.

Cheng, Y. C. (2000). Cultural factors in educational effectiveness: a framework for comparative research. *School Leadership and Management , 20* (2), 207–225.

Cheng, Y. C., Cheung, W. M., & Tam, W. M. (2002). The Pacific Rim and Australia- Hong Kong. In D. Reynolds, B. Creemers, S. Stringfield, C. Teddlie, & E. Schaffer, *World class schools: Internationa lperspectives on school effectiveness* (pp. 138-155). London: RoutledgeFalmer.

Cheng, Y. C., & Tsui, K. T. (1996). Total teacher effectiveness: new conception and improvement. *International Journal of Educational Management, 10*(6), 7-17.

Cheng, Y. C., & Tsui, K. T. (1999). Multimodels of teacher effectiveness: Implications for research. *The Journal of Educational Research* , 141-150.

Cheung, L. M., Cheng, M. H., & Pang, K. C. (2008). Building a model to define the concept of teacher success in Hong Kong. *Teaching and Teacher Education, 24*, 623–634.

Choi, P. K. (2003). The best students will learn English: ultra-utilitarianism and linguistic imperialism in education in post-1997 Hong Kong. *Journal of Education Policy, 67*(3), 673 — 694.

Clark, C., & Peterson, P. (1986). Teachers thought processes. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 255-296). New York, NY: Macmillan.

Coleman, J. S. (1987). Families and schools. *Educational Research, 16*, 32-38.

Cook, V. (2001). Using the first language in the classroom. *Canadian Modern Language Review, 57* (3), 402-423.

Cook, V. (2008). *Second language learning and language teaching* (4th ed.). London: Hodder Education.

Costin, F., & Grush, J. (1973). Personality correlates of teacher-student behavior in the college classroom . *Journal of Educational Psychology, 65*, 410-44.

Cotton, K. (1995). *Effective schooling practices: A research synthesis 1995 Update.* Portland, OR: Northwest Regional Educational Laboratory.

Creemers, B. (1994). *The effective classroom.* London: Cassell.

Creemers, B. (2001). The effective teacher: What changes and remains. In Y. Cheng, M. Mok, & K. T. Tsui (Eds.), *Teaching effectiveness and teacher development: towards a new knowledge base* (pp. 85-109). Hong Kong: Hong Kong Institute of Education/Kluwer Academic Publisher.

Creemers, B., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools.* Abingdon: Routledge.

Creemers, B., & Reezigt, G. (1999). The role of school and classroom climate in elementary school learning environments. In H. Freiberg (Ed.), *School climate: measuring improving and sustaining healthy learning environments* (pp. 30-47). London: Falmer.

Creemers, B., Stringfield, S., & Guldemond, H. (2002). The quantitative data. In D. Reynolds, B. Creemers, S. Stringfield, C. Teddlie, & E. Schaffer, *World class schools: International perspectives on school effectiveness* (pp. 33-55). London: RoutledgeFalmer.

Creswell, J. W. (2007). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Boston: Allyn & Bacon.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209-240). Thousand Oaks, CA: Sage.

Croll, P. (1986). *Systematic classroom observation.* Lodon: Falmer Press.

Croll, P. (1996). Teacher-pupil interaction in the classroom. In P. Croll, & N. Hastings (Eds.), *Effective primary teaching* (pp. 14-28). London: David Fulton .

Cuttance, P. (1987). *Modelling variation in the effectiveness of schooling.* Edinburgh: Center for Educational Sociology, University of Edinburg.

Dang, D., & Rogers, F. (2008). The growing phenomenon of private tutoring: Does it deepen human capital, widen inequalities, or waste resources? *The World Bank Research Observer, 23* (2), 161-200.

Darling-Hammond, L. (1999). *Teacher quality and student achievement: A review of state policy evidence.* Seatle, WA: Center for the study of teahing and policy, University of Washington.

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Educational Policy Analysis Archives , 8*, 1.

Darling-Hammond, L. (2007). Recognizing and enhancing teacher effectiveness: A policymaker's guide. In L. Darling-Hammond, & C. D. Prince, *Strengthening teacher quality in high-need schools—policy and practice.* Washington, DC: The Council of Chief State School Officers.

Darling-Hammond, L. (2008). *Measuring and improving the effectiveness of high school teachers.* Retrieved September 10, 2009, from Alliance for Excellent Education: http://www.all4ed.org/files/TeacherEffectiveness.pdf

Day, C. (1998). Working with the different selves of teachers: beyond comfortable. *Educational Action Research, 6*(2), 255-275.

Day, C. (1999). Professional development and reflective practice: purposes, processes and partnerships. *Pedagogy, Culture & Society, 7* (2), 221-233.

Day, C. (2004). Change Agendas: The roles of teacher educators. *Teaching Education , 15*(2), 145-158.

Day, C. (2005). Principals who sustain success: Making a difference in schools in challenging circumstances. *International Journal of Leadership in Education , 8*(4), 273–290.

Day, C., Sammons, P., & Gu, Q. (2008). Combining qualitative and quantitative methodologies in research on teachers' lives, work, and effectiveness: From integration to synergy. *Educational Resaercher , 37*(6), 330-342.

Day, C., Sammons, P., Kingston, E., & Regan, E. K. (2008). *Effective classroom practice (ECP): A mixed-method study of influences and outcomes.* Swindon: ESRC.

Day, C., Sammons, P., Stobart, G., Kington, A., & Gu, Q. (2007). *Teachers matter: Connecting work, lives and effectiveness.* Maidenhead: Open University Press.

Day, C., Stobart, G., Sammons, P., & Kington, A. (2006). Variations in the work and lives of teachers: relative and relational effectiveness. *Teachers and Teaching: Theory and Practice, 12* (1), 169-192.

De Jong, R., Westerhof, K. J., & Kruiter, J. H. (2004). Empirical evidence of a comprehensive model of school effectiveness: A multilevel study in mathematics in the first year of junior general education in the Netherlands. *School Effectiveness and School Improvement , 15*(1), 3-31.

Doyle, W. (1987). Research on teaching effects as a resource for improving instruction. In M. Wideen, & I. Andrews (Eds.), *Staff development for school improvement* (pp. 91-102). Lewes: Falmer Press.

Driessen, G., & Sleegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement , 11* (1), 57-79.

Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching.* New York, NY: Holt, Rinehart and Winston.

Education Bureau (EDB). (2010a). *Education Reform Highlights*. Retrieved on Jan 31, 2010, from Education Bureau: http://www.edb.gov.hk/index.aspx?nodeID=88&langno=1

Education Bureau. (2010b, Jan 10). *Secondary School Places Allocation (SSPA) System*. Retrieved on Jan 10, 2010, from Education Bureau: http://www.edb.gov.hk/index.aspx?nodeID=1565&langno=1

Education Bureau. (2010c). *Meduium of Instruction*. Retrieved on Jan 9, 2010, from Education Bureau: http://www.edb.gov.hk/index.aspx?nodeID=1900&langno=1

Education Bureau. (2010d, Jan 20). *Leadership Training Program for Principals Consultation Paper.* Retrieved on Jan 30, 2010, from Eductaion Bureau: http://www.edb.gov.hk/index.aspx?nodeID=420&langno=1

Education Commission. (1990). *Report No. 4.* Hong Kong: Government Printer.

Education Comission. (2005). *Report on review of medium of instruction for Secondary Schools and Secondary School Places Allocation.* Hong Kong: Printing Department.

Education Department. (1997a). Arrangements for firm guidance on secondary schools' medium of instruction. Consultation document. Hong Kong: Government Printer.

Education Department. (1997b). Medium of instruction guidance for secondary schools. Hong Kong: Government Printer.

Elliott, J. Methodology and ethics. In C. Adelman (Ed.), *The politics And ethics of evaluation* (pp. 19-25). Beckenham: Croom Helm.

Elliot, J., & Lukeš, D. (2008). Epistemology as ethics in research and policy: The use of case studies. *Journal of Philosophy of Education, 42*(1), 87-119.

Emmer, E., & Peck, R. (1973). Dimensions of classroom behavior. *Journal of Educational Psychology , 64*(2), 223-240.

Erickson, F. (1985). *Qualitative methods in research on teaching.* East Lansing, MI: Institute for Research on Teaching, Michigan State University.

Evans, S. (2000). Hong Kong's new English language policy in education. *World Englishes, 19*(2), 185-204.

Evans, S. (2002). The medium of instruction in Hong Kong: Policy and practice in the new English and Chinese streams. *Research Papers in Education, 17*(1), 97–120.

Evans, S. (2009). The medium of instruction in Hong Kong revisited: policy and practice in the reformed Chinese and English streams. *Research Papers in Education , 24*(3), 287-309.

Evans, S., & Green, C. (2003). The use of English by Chinese professionals in post-1997 Hong Kong. *Journal of Multilingual and Multicultural Development , 24*(5), 386–412.

Farrell, A. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research, 63*(3), 324-327.

Festinger, L., Riecken, H. W., & Schachter, S. (1956). *When Prophecy Fails: A Social and Psychological Study of A Modern Group that Predicted the Destruction of the World.* Minneapolis: University of Minnesota Press.

Fitz-Gibbon, C. T. (1991). *Evaluation of school performance in public examinations: A report for the Scottish Office Education Department.* Durham: Center for Evaluation and Monitoring.

Fitz-Gibbon, C. T. (1996). *Monitoring education: Indicators, quality and effectiveness.* London: Continuum.

Fok, S. (2004). Values orientations of Hong Kong's reform proposals. *School Effectiveness and School Improvement , 15*(2), 201-214.

Fontana, A., & Frey, J. (2005). The interview: From neutral stance to political involvement. In N. Denzin, & Y. Lincoln (Eds.), *The SAGE handbook of qualitative research* (pp. 695-727). London: Sage.

Frey, J., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal, 28* (2), 175-197.

Gadamer, H. (1975). *Truth and method.* London: Sheed and Ward.

Gage, M. (1978). *The scientific basis of the art of teaching.* New York, NY: Columbia University Teacher College Press.

Gage, N. (1989). The paradigm wars and their aftermath: A "Historical" sketch of research on teaching since 1989. *Educational Researcher* , 4-10.

Galton, M. (1978). *British mirrors: a collection of classroom observation systems.* Leicester: University of Leicester, School of Education.

Galton, M. (1979). Systematic classroom observation: British research. *Educational Research , 21*(2), 109-115.

Galton, M. (1995). *Crisis in the primary classroom.* London: David Fulton.

Galton, M., & Croll, P. (1980). *Inside the primary classroom.* London: Routledge and Kegan Paul.

Galton, M., & Hargreaves, L. (1996). Today I felt I was actually teaching: The effects of class size on teachers' classroom behaviour. *Education Review , 10*(2), 26-33.

Galton, M., Hargreaves, L., Comber, C., Pell, T., & Wall, D. (1999). *Inside the primary classroom: 20 Years On.* London: Routledge.

Galton, M., & Simon, B. (Eds.). (1980). *Progress and performance in the primary classroom.* London: Routledge and Kegan Paul.

Gamoran, A. (2000). High standards: A strategy for equalizing opportunities for learning? In R. D. Kahlenberg, *A notion at risk: Preserving public education as an engine for social mobility* (pp. 93-126). New York: The Century Foundation.

Gamoran, A. (2002). *Standards, inequality and ability grouping in schools.* University of Edinburg. Edinburg: Center for Educational Sociology.

Gao, L., & Watkins, D. (2001). Identifying and assessing the conceptions of teaching of secondary school physics teachers in China. *British Journal of Educational Psychology, 71*, 443-469.

Getzels, J., & Jackson, P. (1963). The teacher' s personality and characteristics. In M. Gage (Ed.), *Handbook of research on teaching* (pp. 506-582). Chicago: Rand McNally.

Glass, G. V. (1977). Integrating findings: the meta-analysis of research. *Review of Research in Education, 47*(5), 351-379.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis.* Washington, D.C.: National Cmephensive Centre for Teacher Quality.

Goe, L., & Croft, A. (2009, March). *Methods of evaluating teacher effectiveness.* Retrieved Dec 9, 2009, from National Comprehensive Centre for Teacher Quality: http://www.tqsource.org/publications/RestoPractice_EvaluatingTeacherEffectiveness.pdf

Goldhaber, D. (2007). Everyone's doing it, but what does teacher testing tell us about teacher effectiveness. *Journal of Human Resources, XLII* (4), 765-794.

Goldhaber, D., & Anthony, E. (2007). Can teacher quality be effectively assessed: National Board Certification as a signal of effective teaching. *The Review of Economics and Statistics, 89*(1), 134-150.

Good, T. L., Biddle, B. J., & Brophy, J. E. (1983). *Teacher effectiveness: Research findings and policy implications.* Washington, D.C.: Dingle Asssociates, Inc.

Good, T., & Brophy, J. (1999). *Looking in Classrooms* (8th ed.). Boston: Allyn & Bacon.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational research .* London: Open University Press.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. In J. C. Greene, & V. J. Caracelli (Eds.), *Advances in mixed-method evaluation: The challenges and benefits of integrating diverse paradigms (New Directions for Evaluation, No. 74).* (pp. 5-17). San Francisco, CA: Jossey-Bass.

Greene, J., & Caracelli, V. (2003). Making paradigmatic sense of mixed methods practice. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 91-110). Thousand Oaks, CA: Sage.

Gu, Q., Sammons, P., & Mehta, P. (2008). Leadership characteristics and practices in schools with different effectiveness and improvement profiles. *School Leadership and Management, 38* (1), 43-63.

Hair, J. F. Jr., Black, W., Babin, B., Anderson, R. E., & Tatham, R. (2006). *Multivariate data analysis* (6th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Halvorsen, A.-L., Lee, V., & Andrade, F. H. (2009). A mixed-method study of teachers' attitudes about teaching in urban and low-income schools. *Urban Education, 44* (2), 181-224.

Hammersley, M. (1996). The relationship between qualitative and quantitative research: Paradigm loyalty versus methodological eclecticism. In J. Richardson, *Handbook of qualitative research methods for psychology and the social sciences* (pp. 159-174). Leicester UK: BPS Books.

Hargreaves, A., & Woods, P. (Eds.). (1984). *Classrooms and staffrooms: Sociology of teachers and teaching.* Buckingham: Open University Press.

Harlen, W., & Malcom, H. (1997). *Setting and streaming: A research review.* Scottish Council for Research in Education: Edinburgh.

Harris, A. (1998). Effective teaching: a review of the literature. *School Leadership and Management, 18*(2), 169-183.

Harris, A., & Chapman, C. (2005). Towards differentiated improvement for schools in challenging circumstances. *British Journal of Educational Studies , 52*(4), 45 – 67.

Harris, A., Chapman, C., Muijs, D., Russ, J., & Stoll, L. (2006). Improving schools in challenging contexts: Exploring the possible. *School Effectiveness and School Improvement, 17*(4), 409-424.

Harris, A., Day, C., Hopkins, D., Hatfield, M., Hargreaves, A., & Chapman, C. (2003). *Effective leadership for school improvement.* London: RoutledgeFalmer.

Harris, A., James, S., Gunraj, J., Clarke, P., & Harris, B. (2006). *Improving schools in exceptionally challenging circumstances: Tales from the frontline.* London: Continuum Press.

Harris, A., Muijs, D., Chapman, C., Stoll, L., & Russ, J. (2003). *Raising attainment in the former coalfield areas.* Moorfoot, UK: Department for Education and Skills.

Hattie, J. (2009). *Visible Learning: A synthesis of over 800 meta-analyses relating to achievement.* Abington: Routlege.

Hill, P., & Rowe, K. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement , 7*(1), 1-34.

Hill, P., & Rowe, K. (1998). Modelling Student Progress in Studies of Educational Effectiveness. *School Effectiveness and School Improvement , 9* (3), 310-333.

Ho, B., & Ho, K. (2004). The developmental trend of the medium of instruction in secondary schools of Hong Kong: Prospect and retrospect. *Lamguage and Education, 18*(5), 400-412.

Hockey, J. (1993). Research methods -- researching peers and familiar settings. *Research Papers in Education, 8*(2), 199 - 225.

Hopkins, D. (2001). *School improvement for real.* London: Routledge.

Hopkins, D., Youngman, M., Harris, A., & Wordsworth, J. (1999). Evaluation of the initial effects and implementation of Success for All in England. *Journal of Research in Reading , 22*(3), 257-270.

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis, or dogmas. *Educational Researcher, 17*(8), 10-16.

Husen, T., & & Postlethwaite, N. (Eds.). (1994). *The international encyclopedia of education* (2nd ed.). Oxford: Pergamon Press.

Ip, C. H., & Chan, G. H. (1985). *Studies on the modes of language of instruction at junior secondary levels in Anglo-Chinese secondary schools.* Hong Kong: Education Department and the University of Hong Kong.

James, M. (2006). Balancing rigour and responsiveness in a shifting context: meeting the challenges of educational research. *Research Papers in Education, 21*(4), 365-380.

James, M., Black, P., McCormick, R., Pedder, D., & Wiliam, D. (2006). Learning how to learn, in classrooms, schools and networks: aims, design and analysis. *Research Papers in Education , 21*(2), 101-118.

Jin, L., & Cortazzi, M. (1998). Dimensions of dialogue: Large classes in China. *International Journal of Educational Research , 29*, 739-761.

Johnson, B., & Christensen, L. (2004). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd ed.). Boston, MA: Pearson .

Johnson, R. (1983). Bilingual switching strategies: A study of the modes of teacher-talk in bilingual secondary school classrooms in Hong Kong. *Language Learning and Communication, 2*(3), 267–85.

Johnson, R. K., Chan, R. M., Lee, L. M., & Ho, J. C. (1985). *An investigation of the effectiveness of various language modes of presentation, spoken and written, in Form III in Hong Kong Anglo-Chinese secondary schools.* Hong Kong: Education Department and the University of Hong Kong.

Johnson, R., & Lee., P. (1987). Modes of instruction: Teaching strategies and student responses. In R. Lord, & H. Cheng, *Language education in Hong Kong* (pp. 99–121). Hong Kong: Chinese University Press.

Jordan, H., Mendro, R. L., & Weerasinghe, D. (1997, July). *Teacher Effects on Longitudinal Student Achievement. A preliminary report on research on teacher effectiveness.* Retrieved on Feb 22, 2009, from http://www.dallasisd.org/eval/research/articles/Jordan-Teacher-Effects-on-Longitudinal-Student-Achievement-1997.pdf

Jöreskog, K., & Sörbom, D. (1996). *LISREL 8 : User's Reference Guide* (2 nd ed.). Lincolnwood: Scientific. Software Internationa.

Joyce, B., Calhoun, E., & Hopkins, D. (2008). *Models of learning, tools for teaching* (3rd ed.). Buckingham: Open University Press.

Joyce, B., Weil, M., & Calhoun, E. (2005). *Models of teaching* (7th ed.). Boston: Allyn & Bacon.

Kaya, S., & Rice, D. C. (2009). Multilevel effects of student and classroom factors on elementary sciencen achievement in five countries. *International Journal of Science Education*, 1-27.

Ko, J., & Sammons, P. (2008a). *Using systematic observation schedules to explore variations in effective classroom practices.* Nottingham: School of Education, University of Nottingham.

Ko, J., & Sammons, P. (2008b). *Variation in effective classroom practices: Confirmatory factor analysis.* School of Education, University of Nottingham.

Konstantopoulos, S. (2007). *How long do teacher effects persist.* Bonn: Institute for the Study of Labor (IZA).

Korthagen, F. A. (2004). In search of the essence of a good teacher: towards a more holistic approach in teacher education. *Teaching and Teacher Education , 20*, 77-97.

Krashen, S. D. (1981). *Second language acquisition and second language learning.* Oxford: Pergamon.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Hemel, UK: Prentice-Hall.

Krashen, S., D. & Terrell, T. D. (1983). *The natural approach: Language acquisition in the classroom.* Oxford: Pergamon.

Kyriacou, C. (2007). *Effective teaching in schools-Theory and practice* (3rd Edition ed.). Cheltenham: Nelson Thornes.

Kyriakides, L. (2005). Extending the comprehensive model of educational effectiveness by an empirical investigation. *School Effectiveness and School Improvement, 16*(2), 103-152.

Kyriakides, L. (2008). Testing the validity of the comprehensive model of educational effectiveness: a step towards the development of a dynamic model of effectiveness. *School Effectiveness and School Improvement, 19* (4), 429-446.

Kyriakides, L., & Campbell, R. J. (2003). Teacher evaluation in Cyprus: Some conceptual and methodological issues arising from teacher and school effectiveness research. *Journal of Personnel Evaluation in Education, 17*(1), 21-40.

Kyriakides, L., & Creemers, B. P. (2008a). Using a multidimensional approach to measure the impact of classroom-level factors upon student achievement: A study testing the validity of the dynamic model. *School Effectiveness and School Improvement, 19*(2), 183-205.

Kyriakides, L., & Creemers, B. P. (2008b). A longitudinal study on the stability over time of school and teacher effects on student outcomes. *Oxford Review of Education, 34*(5), 521-545.

Lam, R., Wong, K. C., & Ho, L. M. (2002). School effectiveness of a streamed-school system a multilevel modelling of the Hong Kong secondary schools. *Australian Journal of Education, 46*(3), 287-304.

Le Gallais, T. (2008). Wherever I go there I am: reflections on reflexivity and the research stance. *Reflective Practice, 2*, 145-155.

Lee, J. C., Lam, W. P., & Li, Y. Y. (2003). Teacher evaluation and effectiveness in Hong Kong: Issues and challenges. *Journal of Personnel Evaluation in Education, 17*(1), 41-65.

Leithwood, K., & Mascall. (2008). Collective leadership effects on student achievement. *Educational Administration Quarterly, 44*(4), 529-561.

Leithwood, K., & Steinbach, R. (2002). Successful leadership for especially challenging schools. *Journal of Leadership in Education, 79* (2), 73-82.

LeTendre, G. K. (2002). Advancements in conceptualizing and analyzing cultural effects in cross-national studies of educational achievement. In A. Porter, & A. Gamoran (Eds.), *Methodological advances in cross-national surveys of educational achievement* (pp. 198-230). Washington, DC: National Academies Press.

Levine, D. (1991). Creating effective schools: Findings and implications from research and practice. *Phi Delta Kappan, 72*(5), 389-393.

Levine, D., & Lezotte, L. (1990). *Unusually effective schools: A review and analysis of research and practice.* Madison, WI: National Centre for Effective Schools Research and Development.

Levy, J., Wubbels, T., Brekelmans, M., & Morganfield, B. (1997). Language and cultural factors in students' perceptions of teacher communication style. *International Journal of Intercultural Relationships, 21*, 29-56.

Libman, Z. (2009). Teacher licensing examinations: True progress or an illusion? *Studies in Educational Evaluation , 35*(1), 7-15.

Light, R., & Smith, P. (1971). Accumulating evidence: procedures for resolving contradictions among different studies. *Harvard Educational Review , 41*(4), 429-471.

Lo, M., Chik, P., & Pang, M. (2006). Patterns of variation in teaching the colour of light to Primary 3 students. *Instructional Science, 34*(1), 1-19.

Lo, M., Pong, W., & Chik., P. M. (2005). *For each and everyone - Catering for individual differences through learning study.* Hong Kong: Hong Kong University Press.

Lupton, R. (2004). *Schools in disadvantaged areas: Recognising context and raising quality.* London: Centre for Analysis of Social Exclusion, London School of Economics.

Luyten, H. (1994). Stability of school effects in dutch secondary education: The impact of variance across subjects and years. *International Journal of Educational Research*, 197-216.

Luyten, H. (1995). Teacher change and instability across grades variation of student scores across grades in Dutch elementary schools. *Educational Research and Evaluation, 1*(1), 67 - 89.

Luyten, H., & De Jong, R. (1998). Parallel classes: Differences and similarities, teacher effects and school effects in secondary schools. *School Effectiveness and School Improvement, 9*(4), 437-473.

Luyten, H., Cremers-van Wees, L., & Bosker, R. J. (2003). The Matthew Effect in Dutch primary education: Differences between schools,cohorts and pupils. *Research Papers in Education, 18*(2), 167-195.

Ma, X. (2001). Stability of school academic performance across subject areas. *Journal of Educational Measurement, 38*(1), 1-18.

Macaro, E. (1997). *Target language, collaborative learning and autonomy.* Clevedon, UK: Multilingual Matters.

Macaro, E. (2001a). Analysing student teachers' codeswitching in foreign language classrooms: Theories and decision making. *The Modern Language Journal, 85*(4), 531-548.

Macaro, E. (2001b). *Learning strategies in second and foreign language classrooms.* London: Continuum.

MacBeath, J., Gray, J., Cullen, J., Frost, D., Steward, S., & Swaffield, S. (2007). *Schools on the Edge: Responding to Challenging Circumstances.* London: Paul Chapman.

MacCallum, R. C., Browne, M. W., & & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130-149.

Maden, M. (2001). *Success against the odds: Five years on.* London: Routledge.

Maden, M., & Hillman, J. (Eds.). (1993). *Success against the odds: Effective schools in disadvantaged areas.* London: Routledge.

Magno, C., & Sembrano, J. (2007). The role of teacher efficacy and characteristics on teaching effectiveness, performance, and use of learner-centered practices. *The Asia Pacific-Education Researcher , 16*(1), 73-90.

Mandeville, G. K., & Liu, Q. (1997). The effect of teacher certification and task level on mathematics achievement. *Teaching and Teacher Education , 13* (4), 397-407.

Marks, H., & Printy, S. (2003). Principal leadership and school performance: An integration of transformational and instructional leadership. *Educational Administration Quarterly, 39*(3), 370-397.

Marsh, H. W., & Hau, K. T. (1999). Confirmatory factor analysis: Strategies for small sample sizes. In R. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 251-284). Thousand Oaks , CA : Sage.

Marsh, H. W., & Hau, K. T. (2003). Big fish little pond effect on academic self-concept: A crosscultural (26-country) test of the negative effects of academically selective schools. *American Psychologist*, 364-376.

Marsh, H. W., Kong, C., & Hau, K. (2000). Longitudinal multilevel modeling of the Big Fish Little Pond Effect on academic self-concept: Counterbalancing social comparison and reflected glory effects in Hong Kong high schools. *Journal of Personality and Social Psychology*, 337-349.

Martin, N. K., Baldwin, B., & Yin, Z. (1995, Apr). *Beliefs regarding classroom management style: Relationships to particular teacher personality characteristics.* Retrieved on Jan 23, 2009, from ERIC: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/2f/97.pdf

Marton, F. (1986a). Educational research: Then and now. Reflections on practices and impact. *Australian Educational Researcher, 13*, 5-31.

Marton, F. (1986b). Phenomenography: A research approach to investigating different understandings of reality. *Journal of Thought, 21*, 28- 49.

Marton, F., & Booth, S. (1997). *Learning and Awareness.* Mahwah, NJ: Erlbaum.

Marton, F., & Saljo, R. (1976). On qualitative differences in learning - I: Outcome and process. *British Journal of Educational Psychology, 46*, 4-11.

Marton, F., & Tsui, A. B. (2004). *Classroom discourse and the space of learning.* Mahwah, NJ: LEA.

Marton, F., Watkins, D., & Tang, C. (1997). Discontinuities and continuities in the experience of learning: An interview study of high-school students in Hong Kong. *Learning and Instruction, 7*, 21–48.

Marzano, R. (2003). *What works in schools: Translating research into action.* Alexandria, VA: ASCD.

Maxcy, S. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 51-89). Thousand Oaks, CA: Sage.

Maxwell, J. A., & Loomis, D. M. (2003). Mixed methods design: An alternative approach. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 241-271). Thousand Oaks, CA: Sage.

Mcber, H. (2000). *Research into teacher tffectiveness: a model of teacher effectiveness.* London: DfEE.

McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability.* Santa Monica: RAND.

McIlrath, D., & Huitt, W. (1995, Dec). *The teaching-learning process: A discussion of models.* Retrieved on Feb 20, 2009, from Educational Psychology Interactive: http://chiron.valdosta.edu/whuitt/papers/modeltch.html

Md-Basir, H., Mahzan, S., Jaafar, N_H., Mokhtar, R., Abdl-Aziz, N., Rahman, M.A., & Daud, D. (2010) The online construct validity calculator (CVC) Version 2.0. Malaysia: Universiti Teknologi MARA. Retrieved on April 28, 2010 from http://www.hishammb.net/cvc2/

Medley, D. M. (1982). Teacher effectiveness. In H. Mitzel, *Encyclopedia of Educational Research* (5th ed., pp. 1841-1851). New York, NY: Free Press.

Meehan, M., Cowley, K., Fich, N., Chadwick, K., Ermolov, L., & Riffle, M. (2004). *Special Strategies Observation System-Revised:A Useful Tool for Educational Research and Evaluation.* Charleston, WV: AEL.

Mercer, J. (2007). The challenges of insider research in educational institutions: wielding a double-edged sword and resolving delicate dilemmas. *Oxford Review of Education, 33* (1), 1-17.

Merriam, S. B. (1998). *Qualitative research and case study applications in education* (Revised ed.). San Francisco, CA: Jossey-Bass.

Meyers, L.S., Gamst, G., & Guarino, A.J. (2006). *Applied multivariate research*. Thousand Oaks, CA: Sage.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York, NY: Harper & Row.

Monk, D. H. (1994). Subject matter preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review, 13*(2), 125-145.

Moon, T. C. (1971). A study of verbal behavior patterns in primary grade classrooms during. *Journal ofResearch in Science Teaching, 8*, 171-177.

Moore, W., & Esselman, M. (1992). Teacher efficacy, power, school climate and achievement: A desegregating district's experience. *Annual Meeting of the American Educational Research Association, San Francisco.*

Morrison, K., & Lui, I. (2000). Ideology, linguistic capital and the medium of instruction in Hong Kong. *Journal of Multilingual and Multicultural Development , 21*(6), 471-486.

Morse, J. M. (2003). Principles of mixed and multi-method research design. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioural research* (pp. 189-208). Thousand Oaks, CA: Sage.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *School matters: The junior years.* Somerset : Open Books.

Muijs, D., & Reynolds. (2000). School effectiveness and teacher effectiveness in mathematics: Some preliminary findings from the

Evaluation of the Mathematics Enhancement Programme (Primary). *School Effectiveness and School Improvement, 11*(3), 273-303.

Muijs, D., & Reynolds, D. (2002). Teachers' beliefs and behaviors: What Really Matters? *Journal of Classroom Interaction, 37*(2), 3-15.

Muijs, D., & Reynolds, D. (2005). *Effective teaching: Evidence and practice* (2nd ed.). London: Sage.

Muijs, D., Campbell, J., Kyriakides, L., & Robinson, W. (2005). Making the case for differentiated teacher effectiveness: An overview of research in four key areas. *School Effectiveness and School Improvement , 16*(1), 51-70.

Muijs, D., Harris, A., Chapman, C., Stoll, L., & Russ, J. (2004). Improving schools in socioeconomically disadvantaged areas - A review of research evidence. *School Effectiveness and School Improvement , 15*(2), 149-175.

Mulford, B., Silins, H., & Leithwood, K. (2004). *Leadership for organisational learning and student outcomes:outcomes: A problem-based learning approach.* Dordrecht: Kluwer.

Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing, 5* (2), 159-168.

Muñoz, M., & Chang, F. (2007). The elusive relationship between teacher characteristics and student academic growth: A longitudinal multilevel model for change. *Journal of Persnnel Evaluation in Education, 20*, 147–164.

Newman, I., Ridenour, C., Newman, C., & DeMarco, G. J. (2003). A typology of research purposes and its relationship to mixed methods research. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 167-188). Thousand Oaks, CA: Sage.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis , 26*(3), 237-257.

Ober, R., Wood, S., & Cunningham, G. (1970). *Simultaneous use of four different observational systems to assess student teacher classroom behavior.* ERIC ED No.041830.

Opdenakker, M., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School mprovement, 11*(2), 165-196.

Opdenakke, M.-C., & Van Damme, J. (2006). Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and Teacher Educatio , 22*(1), 1-21.

Pennycook, A. (1995). English in the world/the world in English. In J. Tollefson (Ed.), *Power and inequality in language education* (pp. 34-58). Cambridge: Cambridge University Press.

Perrott, E. (1982). *Effective Teaching.* London: Longman.

Philippou, G., & Christou, C. (1999). Teachers' conceptions of mathematics and students' achievement: A cross-cultural study based on results from TIMSS. *Studies in educational evaluation, 25*, 379-398.

Philpott, J. (2009). *Captivating your class: Effective teaching skills.* London: Continuum.

Pollard, A., Broadfoot, P., Croll, P., Osborn, M., & Abbott, D. (1994). *Changing english primary schools?: The impact of the Education Reform Act at Key Stage 1.* London: Continuum.

Poon, A. (1999). Chinese medium instruction and its impact on English learning in post-1997 Hong Kong. *International Journal of Bilingual Education and Bilingualism, 2*(2), 131-146.

Porter, A., & Brophy, J. (1988). Synthesis of research on good teaching: insights from the work of the Institute for Research on Teaching. *Educational Leadership , 46*, 74-85.

Porter, A., & Gamoran, A. (Eds.). (2002). *Methodological advances in cross-national surveys of educational achievement.* Washington, DC: National Academies Press .

Pratt, D., Kelly, M., & Wong, W. (1999). Chinese conceptions of effective teaching' in Hong Kong: towards culturally sensitive evaluation of teaching. *International journal of lifelong education, 18*(4), 241-258.

Preacher, K. J. (2002). Exploratory factor analysis in behavior genetics research: Factor recovery with small sample sizes. *Behavior Genetics , 32*, 153-161.

Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal, 83*(4), 426-452.

Rabinowitz, W., & Travers, R. (1953). Problems of defining and assessing teacher efeectiveness. *Educational Theory, 3*(3), 212 - 219.

Relich, J. (1996). Gender, self-concept and teachers of mathematics: Effects on attitudes to teaching and learning. *Educational Studies in Mathematics , 30*, 179-195.

Reynolds, D., Bollen, R., Creemers, B., Hopkins, D., Stoll, L., & Lagerweij, N. (Eds.). (2005). *Making good schools: Linking school effectiveness and school improvement.* London: Routledge.

Reynolds, D., Creemers, B., Stringfield, S., Teddlie, C., & Schaffer, G. (Eds.). (2002). *World class schools: International perspectives on school effectiveness.* London: RoutledgeFalmer.

Robinson, V. M., Lloyd, C., & Rowe, K. (2008). The impact of leadership on student outcomes: An analysis of the differential effects of leadership types. *Educational Administration Quarterly, 44* (5), 635-674.

Rocco, T. S., Bliss, L. A., Gallagher, S., Pérez-Prado, A., Alacaci, C., Dwyer, E. S., et al. The pragmatic and dialectical lenses: Two views of mixed methods use in education. In A. Tashakkori, & C. Teddlie (Eds.), *The handbook of mixed methods in the social and behavioral sciences* (pp. 595-615). Thousand Oaks, CA: Sage.

Rosenshine, B. (1970). The stability of teacher effects upon student achievement. *Review Review of Educational Research , 40*, 647-662.

Rosenshine, B. (1971). *Teaching behaviors and student achievement.* London: National Foundation for Educational Research.

Rosenshine, B. (1973). The smallest meaning sample of classroom transactions. *Journal of Research in Science Teaching, 10*(3), 221-226.

Rosenshine, B. (1983). Teaching functions in instructional programs. *Elementary School Journal, 83*(4), 335- 351.

Rosenshine, B., & Furst, N. (1973). The use of direct observation to study teaching. In R. Travers (Ed.), *Second handbook of research on teaching* (pp. 122-183). Chicago: Rand McNally.

Rosenthal, R., & Fode, K. L. (1963). Psychology of the scientist: V. Three experiments in experimenter bias. *Psychological Reports, 12*, 491-511.

Rowe, K. (2006). Effective teaching practices for students with and without learning difficulties Issues and implications. *Australian Journal of Learning Disabilities , 11*(3), 99-115.

Rowe, K. (2007). *The imperative of evidence-based instructional leadership--Building capacity within professional learning communities via a focus on effective teaching practice.* Retrieved September 20, 2009, from Australian Council for Educational Research: http://www.acer.edu.au/documents/Rowe_InstructionalLeadershipFeb2007.pdf

Sammons, P. (1999). *School effectiveness: Coming of age in the twenty-first century.* Lisse, The Netherlands: Swets & Zeitlinger.

Sammons, P., (2010). The contribution of mixed methods to recent research on educational effectiveness, Chapter 27. In A. Tashakkori & C. Teddlie (Eds.) (2nd ed.), *Handbook of mixed methods research* (pp 697-723). Thousand Oaks, CA: Sage.

Sammons, P., Day, C., Kington, A., Gu, Q., Stobart, G., & Smees, R. (2007). , Exploring variations in teachers' work, lives and their effects on pupils: key findings and implications from a longitudinal mixed methods study. *British Educational Research Jounal, 33*(5), 681-701.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools: A review of school effectiveness research.* London: OFSTED.

Sammons, P., & Ko, J. (2008). *Using Systematic Classroom Observation Schedules to Investigate Effective Teaching: Overview of Quantitative Findings. An Effective Classroom Practice (ECP) Project Report.* Swindon: ESRC.

Sammons, P., & Luyten, H. (2009). Editorial article for special issue on alternative methods for assessing alternative methods for assessing. *School Effectiveness and School Improvement, 20*(2), 133-143.

Sammons, P., Mortimore, P., & Thomas, S. (1996). Do schools perform consistently across? In J. Gray, D. Reynolds, C. Fitz-Gibbon, & D. Jesson, *Merging traditions: The future of research on school effectiveness and school improvement* (pp. 1-21). London: Cassells.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a re-analysis of the Inner London Education Authority's Junior School Project. *British Educational Research Journal, 19*, 381-405.

Sammons, P., Siraj-Blatchford, I., Sylva, K., Melhuish, E., Taggart, B., & Elliot, K. (2005). Investigating the Effects of Pre-school Provision: Using mixed methods in the EPPE research. *International Journal of Social Research Methodology, Special issue on Mixed Methods in Educational Research, 8* (3), 207-224.

Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: Effective schools and effective departments.* London: Sage.

Sammons, P., West, A., & Hind, A. (1997). Accounting for variations in pupil attainment at the end of. Key Stage I. *British Educational Research Journal* , 489-511.

Sanders, W. (1998). Value-added assessment. *School Administrator , 11*(3), 24-27.

Sanders, W. (2000). Value-added assessment from student achievement data:opportunities and hurdles. *Journal of Personnel Evaluation in Education , 14*(4), 329-39.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement (Research progress report).* Retrieved on Dec 20, 2009, from http://mdk12.org/practices/ensure/tva/tva_2.html

Schaffer, E., Nesselrodt, P., & Stringfield, S. (1991). *The groundings of an observational instrument: the teacher behaviour - student learning research base of the special strategies observation system.* Kaohsiung, Taiwan: International school effects research workshop, September 26-27, 1991 at Kaohsiung Normal University.

Schaffer, E., Nesselrodt, P., & Stringfield, S. (1994). The contribution of classroom observation to school effectiveness research. In B. Creemers, P. Nesselrodt, E. Schaffer, S. Stringfield, & C. Teddlie (Eds.), *Advances in school effectiveness research and practice* (pp. 133-150). Oxford: Pergamon Press.

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement, 1*(1), 61- 80.

Scheerens, J. (1992). *Effective schooling. Research, theory and practice.* London: Cassell.

Scheerens, J. (2004). *Review of school and instructional effectiveness research.* Retrieved on Feb 10, 2008, from http://unesdoc.unesco.org/images/0014/001466/146695e.pdf

Scheerens, J. (2008). *Review of research on school and instructional effectiveness.* Enschede: University of Twente.

Scheerens, J., & Bosker, R. (1997). *The foundation of educational effectiveness.* Oxford: Pergamon.

Schereens, J., & Creemers, B. (1989). Conceptualising school effectiveness. *International Journal of Educational Research, 13* (Special), 691-706.

Scheerens, J., Vermeulen, C. J., & Pelgrum, W. J. (1989). Generalizibility of instructional and school effectiveness indicators across nations. *International Journal of Educational Research, 13*(7), 789-800.

Schmidt, W., McKnight, C., & Raizen, S. (1996). *A splintered vision: An investigation of US science and mathematics education.* Lansing, MI: US National Centre for TIMSS, Michigan State University.

Scholz, R., & Tietje, O. (2003). *Embedded case study methods: Integrating quantitative and qualitative knowledge.* London: Sage.

Schunk, D., & Rice, J. (1993). Strategy fading and progress feedback: Effects on self-efficacy and comprehension among students receiving remedial reading services. *Journal of Special Education, 27*, 257–276.

Shek, C. C., Johnson, R. K., & Law, E. H. (1991). Survey of the language policy and practice in 193 Hong Kong secondary schools. *New Horizons in Education, 32*, 1–10.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Resercher, 57*(1), 4-14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1-17.

Simon, A., & Boyer, E. G. (Eds.). (1967-1970). *Mirrors for behavior: An anthology of classroom observation instruments .* Philadelphia: Research for Better Schools.

Slavin, R. (2006). *Educational psychology* (8th ed.). Boston: Allyn & Bacon.

Slavin, R., & Madden, N. (Eds.). (2001). *Success for All: Research and reform in elementary education.* Mahwah, NJ: Erlbaum.

Sober, R. (1991). The ethics of the rhetoric of teacher professionalisation. *Teaching and Teacher Education, 7* (3), 295-302.

Spradley, J. (1979). *The ethnographic interview.* New York: Holt, Rinehart and Winston.

Stanovich, K. E. (1986). Matthew Effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360–407.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Stoll, L., & Sammons, P. (2007). Growing together: School effectiveness and school improvement in the UK. In T. Townsend (Ed.), *International handbook of school effectiveness and improvement* (pp. 207-222). London: Springer.

Stringfield, S., & Herman, R. (1996). Assessment of the state of school effectiveness research in the United States of America. *School Effectiveness and School Improvement, 7*(2), 159-180.

Stringfield, S., & Slavin, R. (1992). A hierarchical longitudinal model for elementary school effects. In B. Creemers, & G. Reezigt (Eds.), *Evaluation of educational effectiveness* (pp. 35-69). Groningen: ICO.

Stronge, J. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: ASCD.

Stufflebeam, D., & Shinkfield, A. (1995). *Teacher evaluation: Guide to effective practice.* Norwell: Kluwer.

Suknanadan, L., & Lee, B. (1998). *Streaming, setting and grouping by ability: a review of the literature.* Slough: NFER.

Sun, H., Creemers, B., & de Jong, R. (2006). Contextual factors and effective school improvement. *School Effectiveness and School Improvement, 18*(1), 93-122.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2004). *Effective provision of pre-school education (EPPE) project: Final Report.* London: DfES.

Tabachnick, B., & Fidell, F. (2006). *Using multivariate statistics.* Boston, MA: Allyn & Bacon.

Tam, W. (2009). Sufficient conditions for sustainable instructional changes in the classroom: The case of Hong Kong. *Journal of Educational Change, 10* (4), 315-336.

Tan, J. (1998). The marketization of education in Singapore: Policies and implications. *International Review of Education, 44*(1), 47-63.

Tansel, A., & Bircan, F. (2007, July). *Private supplementary tutoring in Turkey recent evidence on its various aspects.* Retrieved on Sept 20, 2009, from http://ssrn.com/abstract=1122804

Tashakkori, A., & Creswell, J. (2007). The new era of mixed methods. *Journal of Mixed Methods Research, 1* (1), 3-7.

Tashakkori, A., & C. Teddlie (2003a). (Eds.), *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage

Tashakkori, A., & Teddlie, C. (2003b). The past and future of mixed methods research: From data triangulation to mixed model designs. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 671-701). Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches.* Thousand Oaks, CA: Sage.

Teddlie, C. (1991). The integration of classroom and school process data in school effects research. *International school effects research workshop, Kaoshiung, Taiwan* (pp. 71-100). Kaoshiung: Kaoshiung Normal University.

Teddlie, C., Creemers, B., Kyriakides, L., Muijs, D., & Yu, F. (2006). The international system for teacher observation and feedback: evolution of an international study of. Educational Research and Evaluation, 12(6), 561-582.

Teddlie, C. & Sammons, P. (2010) Applications of mixed methods to the field of educational effectiveness research, chapter 7. In B.P.M. Creemers, L. Kyriakides, & P. Sammons (Eds*.) Methodological advances in educational effectiveness research* (pp 115-152). London: Routledge.

Teddlie, C., & Stringfield, C. (1993). *Schools make a difference: Lessons learned from a 10 year study of school effects.* New York, NY: Teachers College Press.

Teddlie, C., Stringfield, S., & Burdett, J. (2003). International comparisons of the relationships among educational effectiveness, evaluation and improvement variables: An overview. *Journal of Personal Evaluation in Education, 17*(1), 5-20.

Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp. 3-50). Thousand Oaks, CA: Sage.

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in Schools , 13*(1), 12-28.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods resqarch: Integrating quantitative and qualitative approaches in the social and behavioral sciences.* Thousand Oaks, CA: Sage.

Teddlie, C., Virgilio, I., & Oescher, J. (1990). Development and validation of the Virgilio Teacher Behavior Inventory. *Educational and Psychological Measurement , 50*(2), 421-430.

Thomas, S., Pan, H., & Goldstein, H. (1994). *Report on analysis of 1992 examination.* London: Association of Metropolitan Authorities.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997a). Differential secondary school effectiveness: comparing the performance of different pupil groups. *British Educational Research Journal , 23*(4), 451-469.

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997b). Stability and consistency in secondary schools' effects onstudents' GCSE outcomes over three years. *School Effectiveness and School Improvement , 8*(2), 169–197.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-147). New York, NY: MacMillan.

Thompson, M. (1995). *Professional ethics and the teacher: an educated talent in the service of society.* London: General Teacher Council for England .

Tollefson, J. W. (1991). *Planning language, planning inequality: Language policy in the community.* London: Longman.

Trautwein, U., Koller, O., Schmitz, B., & Baumert, J. (2002). Do homework assignments enhance achievement? A multilevel analysis in 7th-grade mathematics. *Contemporary Educational Psychology , 27*(1), 26-50.

Trochim, W. (2006, Oct 20). *The multitrait-multimethod matrix*. Retrieved on July 20, 2009, from The Research Methods Knowledge Base: http://www.socialresearchmethods.net/kb/mtmmmat.php

Tsang, W. (2004, Aug 14). *Evaluating the medium of instruction policy in a post-colonial society: The case of Hong Kong Special Administration Region.* A paper presented at the annual meeting of the American Sociological Association, Hilton San Francisco & Renaissance Parc 55 Hotel, San Francisco. Retrieved on Mar 16, 2009, from http://www.allacademic.com/meta/p110895_index.html

Tsang, W. K. (2009, July). *The effect of medium-of-instruction policy on education advancement.* Retrieved on Sept 20, 2009, from The Public Policy Digest: http://www.ugc.edu.hk/rgc/ppd1/eng/05.htm

Tsui, A.B. (2003) Understanding expertise in teaching: Case studies of ESL teachers. Cambridge: Cambridge University Press.

Tsui, A. B. (2007). Language policy and the construction of identity: The case of Hong Kong. In A. B. Tsui, & J. W. Tollefson, *Language Policy, Culture and Identity in Asian Contexts* (pp. 121-142). Mahwah, NJ: Erlbaum.

Tsui, A. B., Shum, M. S., Wong, C. K., & Tse, S. K. (1999). Which agenda? Medium of instruction policy in post-1997 Hong Kong. *Lanuage, Culture and Currciulum, 12*(3), 196-214.

Tymms, P. B. (1993). Accountability - Can it be fair? *Oxford Review of Education, 19*(3), 291-299.

van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research, 49*(2), 127-152.

van de Grift, W., Matthews, P., Tabak, l., & de Rijcke, F. (2004). *Comparative research into the inspection of teaching in England and the Netherlands.* London: Ofsted.

Virgilio, I., Teddlie, C., & Oescher, J. (1991). Variance and context differences in teaching at differentially effective schools. *School Effectiveness and School Improvement, 2*(2), 152-168.

Walberg, H. J. (1982). What makes schooling effective? A synthesis and a critique of three national studies. *Contemporary Education Review ,1*, 102-120.

Walberg, H. J. (1986). Syntheses of research on teaching. In M. Wittrock, *Handbook of research on teaching* (3rd ed., pp. 570-602). New York, NY: Macmillan.

Walberg, H. J. (1990). Productive teaching and instruction: assessing the knowledge base. *Phi Delta Kappa, 71*(6), 470-478.

Walberg, H. J., & Tsai, S. L. (1983). Matthew Effects in Education. *American Educational Research Journal, 20*, 359–73.

Wang, M., & Walberg, H. (1991). Teaching and educational effectiveness: Research synthesis and consensus from the field. In H. Waxman, & H. Walberg (Eds.), *Effective teaching: Current research* (pp. 81–104). Berkeley, CA: McCutchan Publishing Corporation.

Watkins, D., & Biggs, J. (Eds.). (1996). *The Chinese learner: cultural psychological and contextual influences.* Hong Kong: Comparative Education Research Centre, The University of Hong Kong.

Webb, R., & Vulliamy, G. (1999). Changing times, changing demands: a comparative analysis of classroom. *Research Papers in Education, 14*(3), 229-255.

Widhiarso, W. (2007). *Estimating reliability for multidimensional measure.* Unpublished Research Summary. Faculty of Psychology, Gadjah Mada University. Retrieved on April 28, 2010 from http://ugm.academia.edu/documents/0044/2514/Widhiarso_-_Estimating_Reliability_For_Multidimensional_Measure.pdf

Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating. *Journal of Educational Measurement , 26*, 209-232.

Wittrock, M. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York, NY: Macmillan.

Wood, S., Brown, B., Ober, R., & Soar, R. (1969). *A factor analysis of three sets of simultaneously collected observational data--Theory and implications.* ERIC No. ED028993.

Wragg, E. (1984). *Classroom teaching skills.* London: Croom Helm.

Wragg, E. (1999). *An introduction to classroom observation* (2nd ed.). London: Routledge.

Yin, R. (1994). *Case study research: Design and methods* (2nd ed.). London: Sage.

Yin, R. (2003). *Case study research: Design and methods* (3rd ed.). London: Sage.

Yip, D., Tsang, W., & Cheung, S. (2003). Evaluation of the effects of medium of instruction on the scinece learning of Hong Kong secondary students. *Bilingual Research Journal , 27*(2), 295-331.

Zhou, Y. (2008). *Success for All: A comprehensive educational reform for improving at-risk students in an urban school in China.* Charlotte, NC: Information Age Publishing.

APPENDICES

Appendix I: International System for Teacher Observation and Feedback (ISTOF) Scale

ISTOF Teacher Observation Protocol
(2006 version)

COMPONENT 1: ASSESSMENT AND EVALUATION

(1) Indicator #1.1: The teacher gives explicit, detailed and constructive feedback

Item #1: *The teacher makes explicitly clear why an answer is correct or not.*

Item #2: *The teacher provides appropriate feedback to the answers given by the students.*

(2) Indicator #1.2: Assessment is aligned with goals and objectives

Item #3: *Assignments given by the teacher are clearly related to what students learned.*

Item #4: *The teacher explains how assignments are aligned to the learning goals of the lesson.*

COMPONENT 2: DIFFERENTIATION AND INCLUSION

(3) Indicator #2.1: The teacher creates an environment in which all students are involved

Item #5: *Students communicate frequently with one another on task-oriented issues.*

Item #6: *All students are actively engaged in learning.*

(4) Indicator #2.2: The teacher takes full account of student differences

Item #7: *The teacher makes a distinction in the scope of the assignments for different groups of students.*

Item #8: *The teacher gives additional opportunities for practice to students who need them.*

COMPONENT 3: CLARITY OF INSTRUCTION

(5) Indicator #3.1: The teacher shows good communication skills

Item #9: *The teacher regularly checks for understanding.*

Item #10: *The teacher communicates in a clear and understandable manner.*

(6) Indicator #3.2: Clear explanation of purpose

Item #11: *The teacher clarifies the lesson objectives at the start of the lesson.*

Item #12: *The teacher asks students to identify the reasons why specific activities take place in the lesson.*

(7) Indicator #3.3: Lessons are well structured

Item #13: *The teacher presents the lesson with a logical flow that moves from simple to more complex concepts.*

Item #14: *The teacher implements the lesson smoothly moving from one stage to another with well-managed transition points.*

COMPONENT 4: INSTRUCTIONAL SKILLS

(8) Indicator #4.1: The teacher is able to engage students

Item #15: *The teacher provides sufficient wait time and response strategies to involve all types of learners.*

Item #16: *The teacher gives assignments that stimulate all students to active involvement.*

(9) Indicator #4.2: The teacher possesses good questioning skills

Item #17: *The teacher poses questions which encourage thinking and elicit feedback.*

Item #18: *The length of the pause following questions varies according to the difficulty level of questions (e.g., a question calling for application of abstract principles requires a longer pause than a factual question).*

(10) Indicator #4.3: The teacher uses various teaching methods and strategies

Item #19: *The teacher uses a variety of instructional strategies during the class period.*

Item #20: *The teacher uses different, appropriate instructional strategies for different groups of students.*

COMPONENT 5: PROMOTING ACTIVE LEARNING AND DEVELOPING METACOGNITIVE SKILLS

(11) Indicator #5.1: The teacher helps pupils develop problem-solving and meta-cognitive strategies

Item #21: *The teacher invites students to use strategies which can help them solve different types of problems.*

Item #22: *The teacher invites students to explain the different steps of the problem solving strategy which they are using.*

Item #23: *The teacher explicitly provides instruction in problem-solving strategies.*

(12) Indicator #5.2: The teacher gives students opportunities to be active learners

Item #24: *The teacher encourages students to ask one another questions and to explain their understanding of topics to one other.*

Item #25: *The teacher gives students the opportunity to correct their own work.*

(13) Indicator #5.3: The teacher fosters critical thinking in students.

Item #26: *The teacher motivates the students to think about the advantages and disadvantages of certain approaches.*

Item #27: *The teacher asks the students to reflect on the solutions/answers they gave to problems or questions.*

Item #28: *The teacher invites the students to give their personal opinion on certain issues.*

(14) Indicator #5.4: The teacher connects material to students' real world experiences

Item #29: *The teacher systematically uses material and examples from the students' daily life to illustrate the course content.*

Item #30: *Students are invited to give their own examples.*

COMPONENT 6: CLASSROOM CLIMATE

(15) Indicator #6.1: All students are valued.

Item #31: *The teacher demonstrates genuine warmth and empathy toward all students in the classroom.*

Item #32: *The teacher shows respect for the students in both in his/her behaviour and use of language.*

(16) Indicator #6.2: The teacher initiates active interaction and participation.

Item #33: *The teacher creates purposeful activities that engage every student in productive work.*

Item #34: *The teacher's instruction is interactive (lots of questions and answers).*

(17) Indicator #6.3: The teacher interacts with all students

Item #35: *The teacher gives turns to and/or involves those students who do not voluntarily participate in classroom activities.*

Item #36: *The teacher seeks to engage all students in classroom activities.*

(18) Indicator #6.4: The teacher communicates high expectations

Item #37: *The teacher praises children for effort towards realizing their potential.*

Item #38: *The teacher makes clear that all students know that he/she expects their best efforts in the classroom.*

COMPONENT 7: CLASSROOM MANAGEMENT

(19) Indicator #7.1: Learning time is maximized

Item #39: *Teacher starts lesson on time.*

Item #40: *Teacher makes sure that students are involved in learning activities until the end of the lesson.*

Item #41: *Actions are taken to minimize disruption.*

(20) Indicator #7.2: Clear rules are evident

Item #42: *There is clarity about when and how students can get help to do their work in class.*

Item #43: *There is clarity about what options are available when the students finish their assignments.*

(21) Indicator #7.3: Misbehaviours and disruptions are effectively dealt with

Item #44: *The teacher corrects misbehaviour with measures that fit the seriousness of the misconduct (e.g., she does not overreact).*

Item #45: *The teacher deals with misbehaviour and disruptions by referring to the established rules of the classroom.*

Appendix II: The Lesson Observation Form for Evaluating the Quality of Teaching (QoT)

LESSON OBSERVATION FORM FOR EVALUATING THE QUALITY OF TEACHING

(PLEASE FILL IN THE ANSWER, OR CIRCLE THE CORRECT ANSWER)

School name /number:	________	teacher:	m / f	
group:	________	age of teacher:	... years	
Activity/subject matter:	________	mixed age group:	y / n	
name inspector:	________	# of pupils in school:		____
% pupils in school from socially deprived families	____	# of pupils in classroom:		____
% pupils in school for which the language of instruction is not their native language	____	# of residents in community:		____

Indicator: The teacher ...			Grade[1]	Good practice examples: *The teacher ...*	Obs[2]
Safe and orderly school climate	11	...ensures a relaxed atmosphere	1 2 3 4	...addresses the children in a positive manner	0 1
				...reacts with humour and stimulates humour	0 1
				...allows children to make mistakes	0 1
	12	...promotes mutual respect	1 2 3 4	...encourages children to listen to each other	0 1
				...intervenes when children are being laughed at	0 1
				...takes (cultural) differences and idiosyncrasies into account	0 1
	13	...supports the self-confidence of pupils	1 2 3 4	...feeds back on questions and answers from pupils in a positive way	0 1
				...expresses positive expectations to pupils about what they are able to take on	0 1
	14	...shows respect for the pupils in behaviour and language use	1 2 3 4	...allows pupils to finish speaking	0 1
				...listens to what pupils have to say	0 1
				...makes no role-confirming remarks	0 1
Stimulating learning climate	21	...ensures cohesion	1 2 3 4	...honours the contributions made by children	0 1
				...ensures solidarity between pupils	0 1
				...ensures that events are experienced as group events	0 1
	22	...stimulates the independence of pupils	1 2 3 4	...allows pupils to work independently on another assignment or to take up an individually selected task after completing an assignment	0 1
				...allows pupils to work with self-correcting materials	0 1
				...has pupils working on daily and weekly tasks	0 1

1 Please circle the correct answer: 1 = predominantly weak; 2 = more weaknesses than strengths 3 = more strengths than weaknesses; 4 = predominantly strong. Score only 3 when all good practice examples (if applicable) are really observed.

2 Please circle the correct answer: 0= no, I didn't observe this; 1= yes, I have observed this.

1

	23	...promotes cooperation between pupils	1 2 3 4	...provides opportunities for pupils to help each other	0 1
				...gives assignments that incite cooperation	0 1
				...gives pupils the opportunity to play together or to carry out assignments together	0 1
	24	There is good individual involvement by the pupils	1 2 3 4	Pupils listen to the instructions actively	0 1
				Pupils take part in learning/group discussions	0 1
				Pupils work on the assignments in a concentrated and task focused way	0 1
Clear objectives	31	...clarifies the lesson objectives at the start of the lesson	1 2 3 4	... informs pupils at the start of the lesson about the aims of the lesson	0 1
				...clarifies the aim of the assignment and what the pupils will learn from it	0 1
	32	...evaluates whether the objectives have been achieved at the end of the lesson	1 2 3 4	...verifies and/or evaluates whether the aims of the lesson have been achieved	0 1
				...checks the pupils' achievements	0 1
Clear instruction	41	...gives clear instructions and explanations	1 2 3 4	...activates the children's prior knowledge	0 1
				...explains in sequential stages	0 1
				...asks questions that are understood by the pupils	0 1
				...summarises the lesson materials from time to time	0 1
	42	...gives clear explanations of the learning materials and the assignments	1 2 3 4	...ensures that every child knows what he has to do	0 1
				...clearly indicates the materials that can be used as learning aids	0 1
	43	...gives feedback to pupils	1 2 3 4	...checks whether pupils have understood the lesson materials when instructing the class	0 1
				...checks whether pupils are completing the assignments correctly	0 1
				...gives feedback on the way pupils arrive at their answers	0 1
				...gives feedback on the social functioning involved in the completion of the tasks (group work)	0 1
Activating pupils	51	...involves all pupils in the lesson	1 2 3 4	...gives assignments that stimulate pupils into active involvement	0 1
				...poses questions that initiate reflection	0 1
				...ensures that pupils listen carefully and keep on working	0 1
				...waits sufficiently long to allow children to reflect after posing a question	0 1

2

				...gives the opportunity to respond to pupils who don't put their hands up	0 1
	52	...makes use of teaching methods that activate the pupils	1 2 3 4	...makes use of conversational forms and discussion forms	0 1
				...provides graduated exercises	0 1
				...permits working in corners/groups	0 1
				...makes use of ICT	0 1
Adaptation of teaching	61	...adapts the instruction to the relevant differences between pupils	1 2 3 4	... allows pupils who need less instruction to commence with the work	0 1
				...gives extra instruction to small groups or individual pupils	0 1
				...does not direct himself exclusively to the middle bracket	0 1
	62	...adapts the assignments and processing to the relevant differences between pupils	1 2 3 4	...makes a distinction in the scope of the assignments between individual children	0 1
				...does not give all children the same time to complete the assignment	0 1
				...allows some children to make use of auxiliary materials	0 1
Teaching learning strategies	71	...ensures that the teaching materials are orientated towards transfer	1 2 3 4	...teaches pupils solution strategies or search and reference strategies	0 1
				...teaches pupils the use of organisation resources	0 1
				...promotes the conscious use of what has been learned in other (different) areas of learning	0 1
	72	...stimulates the use of control activities	1 2 3 4	...gives attention to estimatory calculation/anticipatory reading	0 1
				...has solutions relate to the context	0 1
				...stimulates the use of alternative solutions	0 1
	73	...provides interactive instruction and activities	1 2 3 4	...facilitates mutual interaction between pupils	0 1
				...ensures interaction between pupils and the teacher	0 1
Effective classroom organisation	81	...gives a well structured lesson	1 2 3 4	...ensures clearly recognisable components in the lessons (lesson structure)	0 1
	82	...ensures the orderly progression of the lesson	1 2 3 4	Entering and leaving the classroom takes place in an orderly manner	0 1
				... intervenes in a timely and appropriate way to any order disruptions	0 1

3

				...acts as a 'watchdog' for agreed codes of behaviour and rules	0 1
	83	...uses learning time efficiently	1 2 3 4	There is no loss of time at the start, during, or at the end of the lesson	0 1
				There are no dead moments	0 1
				The children are not left waiting	0 1
	84	...ensures effective classroom management	1 2 3 4	... makes clear which lesson materials should be used The lesson materials are ready to use	0 1
				The lesson materials are adapted to the level and the experience of the pupils	0 1
				The lesson materials are adapted to the level and the experience of the pupils	0 1
Effective classroom layout	91	...ensures that classroom layout supports the pupil activities	1 2 3 4	The furniture is easy to re-arrange	0 1
				The materials are easily accessible	0 1
	92	The teaching environment is educational and contemporary	1 2 3 4	The classroom décor supports the lesson activities	0 1
				The classroom is a linguistically challenging learning environment	0 1
Final judgement	100	I judge the overall quality of teaching as:	1 2 3 4		

4

Appendix III: Pre-Observation Teacher Survey

The University of Nottingham

Teacher Questionnaire

Effective Classroom Practice: A differentiated model

Thank you for helping us with this extension of an important research project by taking the time to fill in this questionnaire, which is a two-year study of teacher's' effective practice. Funded by the Economic and Social Research Council (ESRC), the research is being carried out by an independent team of researchers from the School of Education, University of Nottingham.

Your answers will be treated in the strictest confidence and all findings will be anonymised so that responses cannot be traced back to individual people or schools. The questionnaire should take about 20 minutes to complete the questionnaire. If you have any queries about this questionnaire, please feel free to discuss them with the researcher conducting the visit. Alternatively, contact James Ko on 66443595 (ttxypk@nottingham.ac.uk).

Please return your completed questionnaire to: James Ko, School of Education, Rm C8 Dearing Building, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB.

Background Information

1. Gender — Male ☐1 Female ☐2
2. Level — a) Primary teacher ☐1 Secondary teacher ☐2
   b) Maths teacher ☐1 English teacher ☐2
3. What year group(s) are you currently teaching? (Please tick all that apply)
   Year 1 ☐1 Year 2 ☐2 Year 3 ☐3 Year 4 ☐4 Year 5 ☐5 Year 6 ☐6 RECP ☐7
   Year 7 ☐8 Year 8 ☐9 Year 9 ☐10 Year 10 ☐11 Year 11 ☐12 Year 12 ☐13 A level ☐14
4. What is your age?
   30 or under ☐1 31-35 ☐2 36-40 ☐3 41-45 ☐4 45-50 ☐5
   51-55 ☐6 56-60 ☐7 61-65 ☐8 over 65 ☐9
5. What is your education background? (Please tick all that apply)
   Non-graduate Certificate/Diploma ☐1 Degree ☐2 Postgraduate Certificate/Diploma ☐3
   Masters ☐4 Doctorate ☐5 Other ☐6 Please specify__________
6. How long have you been a teacher?
   a) .... in total?
   0-3 years ☐1 4-7 years ☐2 8-15 years ☐3 16-23 years ☐4 24 years or more ☐5
   b) .... in this school?
   0-3 years ☐1 4-7 years ☐2 8-15 years ☐3 16-23 years ☐4 24 years or more ☐5
7. How many schools have you taught in for a year or more?
   1 ☐1 2-5 ☐2 6-9 ☐3 10+ ☐4
8. What is your current post and responsibility? (Please tick all that apply)
   Class teacher ☐1 Advanced Skills Teacher KS Manager
   Subject coordinator ☐4 Deputy Head of Year ☐5 Head of Year ☐6
   Deputy Head of Department ☐7 Head of Department ☐8 Assistant Head ☐9
   Deputy Head ☐10 Head ☐11
   Other ☐12 Please specify ______________________

1

The University of Nottingham

9. How many hours per week do you usually work (during term time)?

34-40 ☐1 41-45 ☐2 46-51 ☐3 52-60 ☐4 61-65 ☐5 65+ ☐6

10. For the following statements please circle the number closest to how feel:

	Very high	High	Moderate	Low	Very low
a) My current motivation as a teacher is	⑤	④	③	②	①
b) My current commitment to teaching is	⑤	④	③	②	①
c) My current feeling of job satisfaction is	⑤	④	③	②	①
d) My ability to make a difference to pupils learning is	⑤	④	③	②	①
e) The likelihood that my teaching role will change in the next 2 years is	⑤	④	③	②	①
f) The likelihood that my teaching job will change in the next 2 years is	⑤	④	③	②	①
g) The likelihood that my career will change in the next 2 years is	⑤	④	③	②	①
h) The current level of stress I experience in my work as a teacher is	⑤	④	③	②	①

Classroom Teaching

11. Please indicate how often you do each of the following in your classroom teaching:

	All or majority	Often	Sometimes	Rarely	Never
a) Introduce content through formal presentations	⑤	④	③	②	①
b) Pose open-ended questions	⑤	④	③	②	①
c) Engage the whole class in discussions	⑤	④	③	②	①
d) Require students to explain their reasoning when giving an answer	⑤	④	③	②	①
e) Ask students to explain concepts to one another	⑤	④	③	②	①
f) Ask students to consider alternative methods for solutions	⑤	④	③	②	①
g) Allow students to work at their own pace	⑤	④	③	②	①
h) Read and comment on the reflections students have written, e.g., in their journals	⑤	④	③	②	①
i) Take students' prior understanding into account when planning the curriculum	⑤	④	③	②	①
j) Develop students' conceptual understanding of the subjects I teach them	⑤	④	③	②	①
k) Lead a class of students using investigative strategies	⑤	④	③	②	①
l) Manage a class of students engaged in hands-on/project-based work	⑤	④	③	②	①
m) Have students work in cooperative learning groups	⑤	④	③	②	①
n) Listen/ask questions as students work in order to gauge their understanding	⑤	④	③	②	①
o) Use the textbook as a resource rather than the primary instructional tool	⑤	④	③	②	①
p) Teach groups that are heterogeneous in ability	⑤	④	③	②	①
q) Teach students who have limited English proficiency	⑤	④	③	②	①
r) Recognise and respond to student cultural diversity	⑤	④	③	②	①

Relationships

The University of Nottingham

12. Please indicate how satisfied you are with your ability to do the following:

	Very satisfied	Satisfied	Dissatisfied	Very dissatisfied	Not applicable
a) Get your students to follow classroom rules	⑤	④	③	②	①
b) Control disruptive behaviour in the classroom	⑤	④	③	②	①
c) Get students to believe they can do well in schoolwork	⑤	④	③	②	①
d) Get parents to become involved in school activities	⑤	④	③	②	①
e) Assist parents in helping their children do well in school	⑤	④	③	②	①
f) Make parents feel comfortable coming to school	⑤	④	③	②	①
g) Make the school a safe place	⑤	④	③	②	①
h) Make students enjoy coming to school	⑤	④	③	②	①
i) Get students to trust teachers	⑤	④	③	②	①
j) Help other teachers with their teaching skills	⑤	④	③	②	①
k) Motivate students who show little interest in schoolwork	⑤	④	③	②	①
l) Help your students' value learning	⑤	④	③	②	①
m) Craft high quality questions for your students	⑤	④	③	②	①
n) Calm a student who is disruptive or noisy	⑤	④	③	②	①
o) Establish a classroom management system with each group of students	⑤	④	③	②	①
p) Use a variety of assessment strategies	⑤	④	③	②	①
q) Provide an alternative explanation or example when students are confused	⑤	④	③	②	①
r) Implement alternative strategies in your classroom	⑤	④	③	②	①

Q13. Are there are other elements of our classroom practice that influence your relationships with:

Students:______________________________

Parents:______________________________

Other colleagues:______________________________

Appendix IV: Post-observation Interview Schedule

The University of Nottingham

Post-Observation Interview

PhD Project: **Effective Classroom Practice: A mixed methods comparative study of an extended dynamic model of educational effectiveness**

Preamble checklist:

- Thank the teacher for his/her time and continued support of the project and for allowing the observation of the class.
- Outline the aims of the visit: discuss the lesson that was just taught.
- Explain that the post-observation interview should last approximately 50-60 minutes and check that the timing is not a problem for her/him.
- Remind her/him that everything said remains confidential and that the names of all teachers and schools will be changed in publications and reports.
- Ask permission for the interview to be recorded.
- Finally, ask if s/he has any questions before the interview begins and check that the teacher is happy to continue.

<u>A. Teacher's views on teaching</u>

1. How often do you reflect on your teaching? And what do you usually think of?
2. What do you usually do after that?
3. What is/are your goal(s) in teaching? What you value most in teaching?
4. What areas do you think effective teaching should cover? And which matters most to you? In which areas do you show more strength? And in which areas do you find more challenges/difficulties?
   a) Classroom climate
   b) Classroom management
   c) Clarity of instruction
   d) Questioning and presentation skills
   e) Catering individual differences
   f) Diversity of teaching strategies
   g) Assessment and evaluation
   h) Others?
5. Which language skill(s) you can teach better than others?
6. Is there any difference in your teaching practices in your teaching of your major subject and minor subject, if any?

1

The University of Nottingham

7. How would you rate your performance in the past week of teaching? Which lesson do you like most? Why?

8. Which do you like least? Why?

B: Views on teacher effectiveness

1. In your opinion, what is it in or about this school that most influences teaching practices?
   Prompts
   What support systems, policy or influences of leadership influence teacher effectiveness, like beliefs, values, colleagues, Panel Head or the Principal?

2. What would you say were the key changes you have made that have been fundamental to the improvement of pupil outcomes and the effectiveness of your teaching?

3. What factors do you think might have affected your teaching practices in various lessons which I had observed?
   Subject differences- major vs minor
   Pupil characteristics- age, form, ability, motivation
   Teaching characteristics- preparation, teacher-student relationship, T-P interaction, commitment, etc.
   Organizational characteristics- departmental leadership, collegial collaboration, support from senior management

4. Do external policy agendas, such as *Chinese as Medium of Instruction* and *The New Academic Structure for Senior Secondary Education and Higher Education*, affect your teaching practices? If so, how? *How do you manage these?*

C. Teacher's views on pupils

1. How would you like to describe your students in terms of the following?
   a) Achievements/aptitudes
   b) Motivation;
   c) Self-concept;
   d) Engagement in school

2. What do think how they look at themselves?
   a) Achievements/aptitudes
   b) Motivation;
   c) Self-concept;
   d) Engagement in school

3. Would your views about your students differ in terms of their abilities in different subjects? Can you explain why?

4. How would you like to attribute the achievements and under-achievements of your students?

The University of Nottingham

D. Teacher Efficacy

1. In your opinion, have you achieved your goals in teaching?
   How successful are you in doing so?
   What hindrances do you have?
   What are you going to do to further any changes?

2. In your opinion, what factors influence your teaching most and how they may also affect your students?

3. How would you like to describe your role in making your students achieve?

4. Do you feel able to help all your pupils to learn what is expected of them?
   If not, why?

5. How confident you can accomplish the expectations that have been set for your classes?

6. In what ways does the principal develop motivation amongst staff?
   Prompt –
   What are the main factors influencing the motivation of staff in the school?

7. What, if anything, would you change or do you think should be changed in order to improve your teaching practices?

E: Impact of leadership

1. What support, if any, do you receive from what leaders to help you improve your teaching practices?

   Prompt - Is this done on a formal or informal basis?
   Prompt - Have they had a positive impact on the ways you promote and manage pupil outcomes?

2. Could you briefly describe your current developmental needs?
   Probes
   What types of support (formal and informal) do you receive as a teacher?
   What CPD (formal learning) experiences, if any, are promoted amongst staff?

3. In what ways does the principal develop motivation amongst staff?
   Prompt –
   What are the main factors influencing the motivation of staff in the school?

F: Impact of support and teacher identity

1. What does the word "professionalism" mean to you?

2. Please describe the different ways in which teachers can share knowledge and expertise?
   Probes: Are they formal or informal?

3

The University of Nottingham

3. Thinking of three aspects of teachers' lives:
   a) personal dimension
   b) professional role
   c) current teaching situation

   Do you feel these are currently in harmony?

   If yes, how does this harmony impact on your teaching?

   If no, is there one of these aspects that is more dominant for you at the moment?

   If yes, which one?

   Is it having a positive or negative impact on your work?

3. Please could you comment on:
   - Your current level of (*motivation*, *commitment*, *and job satisfaction*)

     Probes

     In your opinion, has your motivation increased, decreased or stayed the same over the last three years?

     What factors have caused this?

     Do you think there have been any consequences or implications of this in terms of your effectiveness as a teacher? If so, what?

G: Well-being

1. Briefly, what are the aspects of your life outside of school, if any, that help or hinder you in your effectiveness as a teacher?
   - age
   - family
   - additional roles at home
   - other

2. Briefly, looking at it the other way round, is there anything about school and your role as a teacher that helps and/or hinders you in your life outside of school?
   - age
   - school context
   - career phase
   - colleagues
   - additional roles at school
   - part time/job share
   - other

3. Is there anything, in your opinion, that the principal/Subject Panel Head/EDB could do to improve your work-life balance?

Do you have any questions or things you would like clarified while I'm here?

This interview schedule is an adaptation of one produced originally for the Effective Classroom Practice Project, an ESRC funded study in England (Day, Sammons & Kington, 2006).

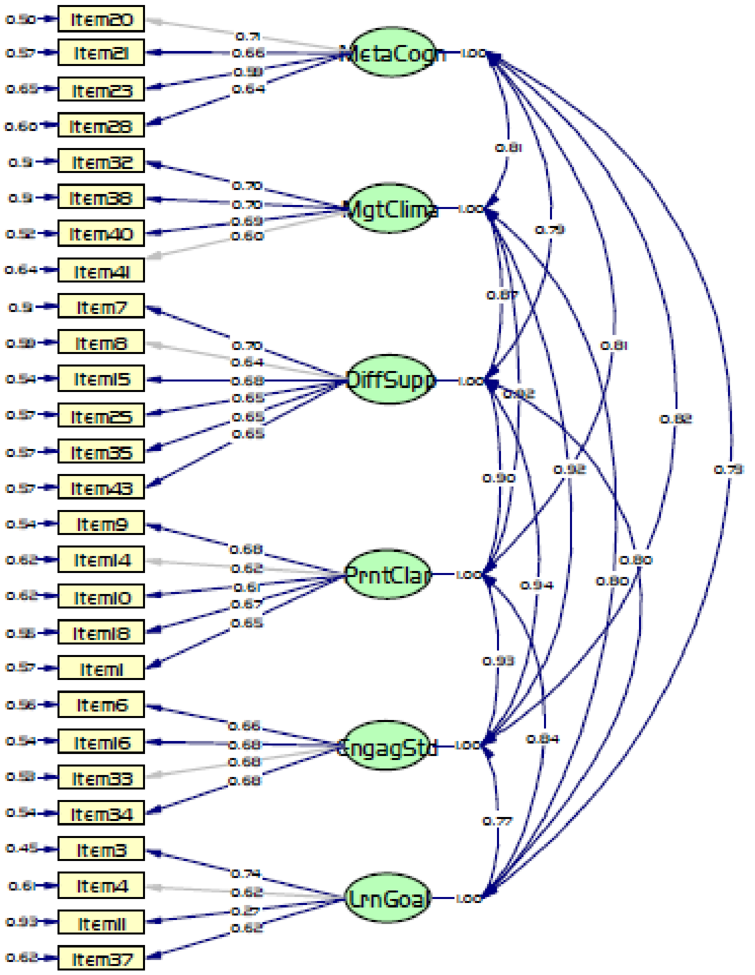
Appendix V: Descriptive Statistics for ISTOF Items

Item No	Item Description (T = The teacher; S= student)	Mean	Std. Deviation	Skew-ness	z-skewness	Kurtosis	z-kurtosis
Item 10	T communicates in a clear and understandable manner.	*4.33*	.89	-1.06	3.85*	.04	0.08
Item 1	T makes explicitly clear why an answer is correct or not.	*4.24*	.81	-.92	3.34*	.41	0.74
Item 40	Teacher makes sure that Ss are involved in learning activities until the end of the lesson.	*4.24*	1.00	-1.23	4.45*	.73	1.35
Item 3	Assignments given by T are clearly related to what Ss learned.	*4.21*	.84	-.98	3.55*	.55	1.01
Item 2	T provides appropriate feedback to the answers given by the Ss.	*4.13*	.84	-.53	1.94	-.65	1.19
Item 9	T regularly checks for understanding.	*4.07*	.90	-1.04	3.76*	1.16	2.12*
Item 42	There is clarity about when and how Ss can get help to do their work in class.	*4.03*	1.23	-.93	3.37*	-.38	0.69
Item 39	Teacher starts lesson on time.	*4.00*	1.06	-.97	3.52*	.61	1.11
Item 41	Actions are taken to minimize disruption.	3.96	1.24	-.92	3.34*	-.40	0.74
Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	3.91	1.31	-1.00	3.63*	-.21	0.39
Item 4	T explains how assignments are aligned to the learning goals of the lesson.	3.82	.93	-.22	0.82	-.91	1.67
Item 31	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	3.80	1.13	-.68	2.45*	-.31	0.56
Item 11	T clarifies the lesson objectives at the start of the lesson.	3.79	.79	-.44	1.61	.02	0.03
Item 45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	3.78	1.41	-.73	2.66*	-.90	1.65
Item 32	T shows respect for the Ss in both in his/her behaviour and use of language.	3.72	1.13	-.41	1.48	-.78	1.42
Item 38	T makes clear that all Ss know that he/she expects their best efforts in the classroom.	3.70	1.02	-.67	2.44*	-.05	0.09
Item 37	T praises children for effort towards realizing their potential.	3.59	1.10	-.24	0.87	-.77	1.41
Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts.	3.54	.96	-.72	2.62*	.45	0.82
Item 6	All Ss are actively engaged in learning.	3.50	1.15	-.27	0.98	-.83	1.52
Item 33	T creates purposeful activities that engage every S in productive work.	3.50	1.19	-.48	1.75	-.51	0.93
Item 14	T implements the lesson smoothly moving from one stage to another with well-managed transition points.	3.46	1.06	-.34	1.23	-.46	0.84
Item 15	T provides sufficient wait time and response strategies to involve all types of learners.	3.38	1.01	-.19	0.69	-.52	0.96
Item 17	T poses questions which encourage thinking and elicit feedback.	3.38	1.07	-.29	1.04	-.62	1.14
Item 36	T seeks to engage all Ss in classroom activities.	3.37	1.06	-.03	0.10	-.99	1.82
Item 16	T gives assignments that stimulate all Ss to active involvement.	3.34	1.05	.04	0.16	-.68	1.25
Item 43	There is clarity about what options are available when the Ss finish their assignments.	3.33	1.40	-.01	0.05	-1.35	2.48*
Item 18	The length of the pause following questions varies according to the difficulty level of questions	3.26	.93	-.55	2.01*	-.24	0.43
Item 25	T gives Ss the opportunity to correct their own work.	3.24	1.23	-.12	0.42	-.87	1.60
Item 34	T's instruction is interactive (lots of questions and answers).	3.20	1.07	-.07	0.26	-.60	1.10
Item 8	T gives additional opportunities for practice to Ss who need them.	3.18	1.16	-.11	0.40	-.97	1.78
Item 19	T uses a variety of instructional strategies during the class period	3.07	1.04	.01	0.05	-.34	0.63
Item 29	T systematically uses material and examples from the Ss' daily life to illustrate the course content.	3.07	1.11	.05	0.17	-.80	1.47
Item 5	Ss communicate frequently with one another on task-oriented issues.	3.05	1.16	.16	0.56	-.96	1.75
Item 12	T asks Ss to identify the reasons why specific activities take place in the lesson.	3.01	.96	.44	1.59	-.13	0.24
Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	2.97	1.18	.15	0.55	-.89	1.63
Item 24	T encourages Ss to ask one another questions and to explain their understanding of topics to one other.	2.96	1.32	.11	0.40	-1.19	2.18*
Item 23	T explicitly provides instruction in problem-solving strategies.	2.95	.86	-.15	0.56	-.47	0.87
Item 30	Ss are invited to give their own examples.	2.91	1.19	.13	0.48	-.89	1.63
Item 7	T makes a distinction in the scope of the assignments for different groups of Ss.	2.84	1.07	.19	0.69	-.34	0.63
Item 21	T invites Ss to use strategies which can help them solve different types of problems.	2.83	1.04	-.38	1.39	-.74	1.36
Item 20	T uses different, appropriate instructional strategies for different groups of Ss.	2.82	.96	.11	0.38	-.24	0.45
Item 27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	2.80	1.14	.01	0.05	-1.14	2.08*
Item 26	T motivates the Ss to think about the advantages and disadvantages of certain approaches.	2.79	1.10	-.06	0.23	-1.06	1.95
Item 22	T invites Ss to explain the different steps of the problem solving strategy which they are using.	2.63	.91	-.18	0.67	-.70	1.28
Item 28	T invites the Ss to give their personal opinion on certain issues.	2.62	1.13	.29	1.07	-.71	1.31

Appendix VI: Frequency table for ISTOF Items

Item No	Item Description (T = The teacher; S= student)	Strongly Disagree	Moderately Disagree	Neutral	Moderately Agree	Strongly Agree
Item 10	T communicates in a clear and understandable manner.	0.00%	3.90%	15.80%	23.70%	56.60%
Item1	T makes explicitly clear why an answer is correct or not.	0.00%	3.90%	11.80%	40.80%	43.40%
Item 40	Teacher makes sure that Ss are involved in learning activities until the end of the lesson.	1.30%	6.60%	13.20%	25.00%	53.90%
Item 3	Assignments given by T are clearly related to what Ss learned.	0.00%	5.30%	10.50%	42.10%	42.10%
Item2	T provides appropriate feedback to the answers given by the Ss.	0.00%	2.60%	21.10%	36.80%	39.50%
Item 9	T regularly checks for understanding.	1.30%	5.30%	13.20%	46.10%	34.20%
Item 42	There is clarity about when and how Ss can get help to do their work in class.	3.90%	10.50%	18.40%	13.20%	53.90%
Item 39	Teacher starts lesson on time.	3.90%	2.60%	23.70%	28.90%	40.80%
Item 41	Actions are taken to minimize disruption.	3.90%	14.50%	10.50%	23.70%	47.40%
Item 44	T corrects misbehaviour with measures that fit the seriousness of the misconduct.	7.90%	10.50%	10.50%	25.00%	46.10%
Item 4	T explains how assignments are aligned to the learning goals of the lesson.	0.00%	7.90%	30.30%	34.20%	27.60%
Item 31	T demonstrates genuine warmth and empathy toward all Ss in the classroom.	3.90%	9.20%	23.70%	28.90%	34.20%
Item 11	T clarifies the lesson objectives at the start of the lesson.	0.00%	6.60%	23.70%	53.90%	15.80%
Item 45	T deals with misbehaviour and disruptions by referring to the established rules of the classroom.	9.20%	14.50%	13.20%	15.80%	47.40%
Item 32	T shows respect for the Ss in both in his/her behaviour and use of language.	2.60%	11.80%	28.90%	23.70%	32.90%
Item 38	T makes clear that all Ss know that he/she expects their best efforts in the classroom.	2.60%	11.80%	19.70%	44.70%	21.10%
Item 37	T praises children for effort towards realizing their potential.	2.60%	13.20%	32.90%	25.00%	26.30%
Item 13	T presents the lesson with a logical flow that moves from simple to more complex concepts.	3.90%	9.20%	27.60%	47.40%	11.80%
Item 6	All Ss are actively engaged in learning.	3.90%	17.10%	27.60%	27.60%	23.70%
Item 33	T creates purposeful activities that engage every S in productive work.	7.90%	10.50%	28.90%	28.90%	23.70%
Item 14	T implements the lesson smoothly moving from one stage to another with well-managed transition points.	3.90%	14.50%	30.30%	34.20%	17.10%
Item 15	T provides sufficient wait time and response strategies to involve all types of learners.	2.60%	17.10%	32.90%	34.20%	13.20%
Item 17	T poses questions which encourage thinking and elicit feedback.	3.90%	18.40%	27.60%	35.50%	14.50%
Item 36	T seeks to engage all Ss in classroom activities.	1.30%	23.70%	27.60%	31.60%	15.80%
Item 16	T gives assignments that stimulate all Ss to active involvement.	2.60%	18.40%	38.20%	23.70%	17.10%
Item 43	There is clarity about what options are available when the Ss finish their assignments.	9.20%	22.40%	30.30%	2.60%	35.50%
Item 18	The length of the pause following questions varies according to the difficulty level of questions	3.90%	17.10%	31.60%	43.40%	3.90%
Item 25	T gives Ss the opportunity to correct their own work.	9.20%	18.40%	31.60%	21.10%	19.70%
Item 34	T's instruction is interactive (lots of questions and answers).	5.30%	21.10%	34.20%	27.60%	11.80%
Item 8	T gives additional opportunities for practice to Ss who need them.	6.60%	26.30%	22.40%	31.60%	13.20%
Item 19	T uses a variety of instructional strategies during the class period	6.60%	21.10%	40.80%	22.40%	9.20%
Item 29	T systematically uses material and examples from the Ss' daily life to illustrate the course content.	6.60%	27.60%	28.90%	26.30%	10.50%
Item 5	Ss communicate frequently with one another on task-oriented issues.	6.60%	31.60%	25.00%	23.70%	13.20%
Item 12	T asks Ss to identify the reasons why specific activities take place in the lesson.	2.60%	27.60%	44.70%	15.80%	9.20%
Item 35	T gives turns to and/or involves those Ss who do not voluntarily participate in classroom activities.	9.20%	30.30%	26.30%	22.40%	11.80%
Item 24	T encourages Ss to ask one another questions and to explain their understanding of topics to one other.	14.50%	28.90%	18.40%	22.40%	15.80%
Item 23	T explicitly provides instruction in problem-solving strategies.	3.90%	26.30%	42.10%	26.30%	1.30%
Item 30	Ss are invited to give their own examples.	11.80%	28.90%	26.30%	22.40%	10.50%
Item 7	T makes a distinction in the scope of the assignments for different groups of Ss.	10.50%	26.30%	39.50%	15.80%	7.90%
Item 21	T invites Ss to use strategies which can help them solve different types of problems.	14.50%	18.40%	38.20%	27.60%	1.30%
Item 20	T uses different, appropriate instructional strategies for different groups of Ss.	7.90%	28.90%	40.80%	18.40%	3.90%
Item 27	T asks the Ss to reflect on the solutions/answers they gave to problems or questions.	13.20%	32.90%	18.40%	31.60%	3.90%
Item 26	T motivates the Ss to think about the advantages and disadvantages of certain approaches.	13.20%	30.30%	23.70%	30.30%	2.60%
Item 22	T invites Ss to explain the different steps of the problem solving strategy which they are using.	11.80%	30.30%	40.80%	17.10%	0.00%
Item 28	T invites the Ss to give their personal opinion on certain issues.	17.10%	32.90%	26.30%	18.40%	5.30%

Appendix VII: Results of the best solution for ISTOF item-based model with standardised coefficients and goodness-of-fit indices



MetaCogn: Meta-cognitive skills teaching

MgtClima: Classroom management and climate

DiffSupp: Differentiation and support

PrntClar: Clarity and logic of presentation

EngagStd: Student engagement

LrnGoal: Strategies to Enhance Learning and Lesson Focus.

Chi-Square=79.85, df=309, P-value=1.00000, RMSEA=0.000

Degrees of Freedom = 309
Minimum Fit Function Chi-Square = 73.85 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 79.85 (P = 1.00)
Estimated Non-centrality Parameter (NCP) = 0.0
90 Percent Confidence Interval for NCP = (0.0 ; 0.0)
Minimum Fit Function Value = 0.98
Population Discrepancy Function Value (F0) = 0.0
90 Percent Confidence Interval for F0 = (0.0 ; 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.0)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
Expected Cross-Validation Index (ECVI) = 5.96
90 Percent Confidence Interval for ECVI = (5.96 ; 5.96)
ECVI for Saturated Model = 10.08, ECVI for Independence Model = 49.87
Chi-Square for Independence Model with 351 Degrees of Freedom = 3686.48
Independence AIC = 3740.48, Model AIC = 217.85, Saturated AIC = 756.00
Independence CAIC = 3830.41, Model CAIC = 447.67, Saturated CAIC = 2015.02
Normed Fit Index (NFI) = 0.98, Non-Normed Fit Index (NNFI) = 1.08, Parsimony Normed Fit Index (PNFI) = 0.86
Comparative Fit Index (CFI) = 1.00
Incremental Fit Index (IFI) = 1.07
Relative Fit Index (RFI) = 0.98
Critical N (CN) = 376.50
Root Mean Square Residual (RMR) = 0.090, Standardized RMR = 0.039
Goodness of Fit Index (GFI) = 0.93, Adjusted Goodness of Fit Index (AGFI) = 0.91, Parsimony Goodness of Fit Index (PGFI) = 0.76

Appendix VIII: Goodness-of-Fit Indices Generated by LISREL 8.72 for ISTOF item-based CFA model

Degrees of Freedom = 390
Minimum Fit Function Chi-Square = 137.59 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 150.30 (P = 1.00)
Estimated Non-centrality Parameter (NCP) = 0.0
90 Percent Confidence Interval for NCP = (0.0; 0.0)
Minimum Fit Function Value = 1.83
Population Discrepancy Function Value (F0) = 0.0
90 Percent Confidence Interval for F0 = (0.0; 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0; 0.0)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
Expected Cross-Validation Index (ECVI) = 7.20
90 Percent Confidence Interval for ECVI = (7.20; 7.20)
ECVI for Saturated Model = 12.40
ECVI for Independence Model = 71.61
Chi-Square for Independence Model with 435 Degrees of Freedom = 5311.03
Independence AIC = 5371.03
Model AIC = 300.30
Saturated AIC = 930.00
Independence CAIC = 5470.95
Model CAIC = 550.11
Saturated CAIC = 2478.79
Normed Fit Index (NFI) = 0.97
Non-Normed Fit Index (NNFI) = 1.06
Parsimony Normed Fit Index (PNFI) = 0.87
Comparative Fit Index (CFI) = 1.00
Incremental Fit Index (IFI) = 1.05
Relative Fit Index (RFI) = 0.97
Critical N (CN) = 250.61
Root Mean Square Residual (RMR) = 0.11
Standardized RMR = 0.051
Goodness of Fit Index (GFI) = 0.88
Adjusted Goodness of Fit Index (AGFI) = 0.86
Parsimony Goodness of Fit Index (PGFI) = 0.74

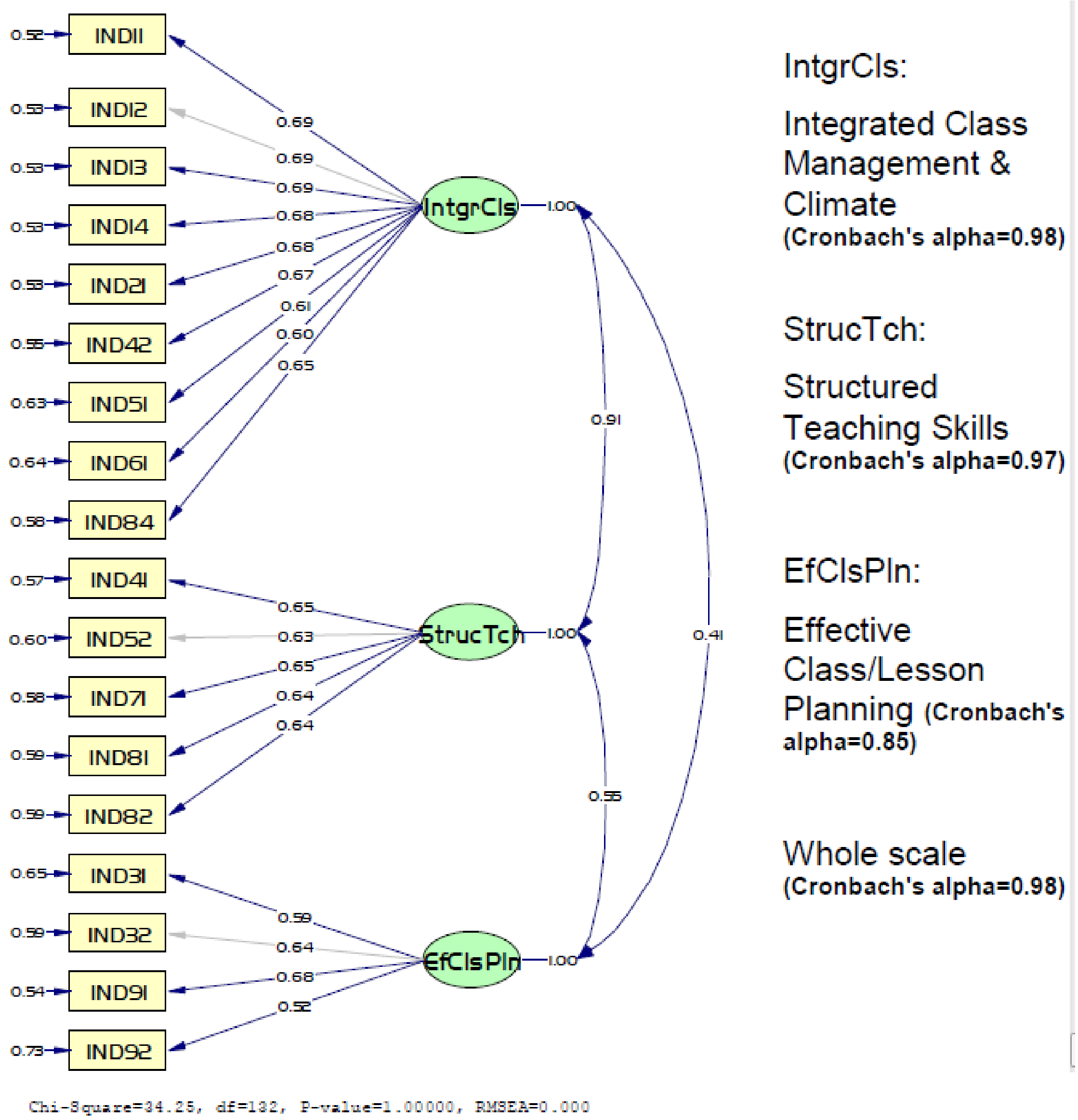
Appendix IX: Descriptive Statistics for QoT Indicators

Indic-ator No.	Indicator Description (T = The teacher; S= student)	Mean	Std. Deviation	Skewness	z-Skewness	Kurtosis	z-Kurtosis
IND43	T gives feedback to Ps	3.36	0.69	-0.85	*3.09**	0.61	1.12
IND84	T ensures effective classroom management	3.28	0.87	-1.07	*3.88**	0.40	0.73
IND41	T gives clear instructions and explanations	3.25	0.66	-0.60	*2.18**	0.70	1.29
IND42	T gives clear explanations of the learning materials and the assignments	3.18	0.74	-0.51	1.86	-0.34	0.62
IND14	T shows respect for the Ps in behaviour and language use	3.14	0.80	-0.92	*3.34**	0.87	1.60
IND83	T uses learning time efficiently	3.13	0.82	-0.84	*3.06**	0.44	0.81
IND81	T ensures the orderly progression of the lesson	3.09	0.82	-0.62	*2.26**	-0.12	0.23
IND11	T ensures a relaxed atmosphere	3.08	0.83	-0.58	*2.11**	-0.26	0.47
IND13	T supports the self-confidence of Ps	3.01	0.90	-0.59	*2.13**	-0.45	0.82
IND31	T clarifies the lesson objectives at the start of the lesson	3.01	0.55	-0.48	1.73	2.30	*4.23**
IND12	T promotes mutual respect	2.99	0.93	-0.58	*2.12**	-0.52	0.96
IND51	T involves all Ps in the lesson	2.99	0.87	-0.22	0.81	-1.11	*2.04**
IND52	T makes use of teaching methods that activate the Ps	2.95	0.88	-0.50	1.83	-0.40	0.74
IND71	T ensures that the teaching materials are orientated towards transfer	2.93	0.82	-0.47	1.70	-0.19	0.35
IND82	T gives a well structured lesson	2.93	0.79	-0.38	1.39	-0.21	0.38
IND24	There is good individual involvement by the Ps	2.92	0.91	-0.50	1.83	-0.48	0.89
IND22	T stimulates the independence of Ps	2.91	0.80	-0.46	1.69	-0.06	0.12
IND21	T ensures cohesion	2.88	0.85	-0.31	1.12	-0.56	1.02
IND72	T stimulates the use of control activities	2.80	0.95	-0.45	1.62	-0.65	1.20
IND73	T provides interactive instruction and activities	2.74	0.91	-0.52	1.89	-0.42	0.77
IND23	T promotes cooperation between Ps	2.70	0.88	-0.08	0.31	-0.73	1.33
IND61	T adapts the instruction to the relevant differences between Ps	2.70	0.83	-0.37	1.34	-0.28	0.51
IND32	T evaluates whether the objectives have been achieved at the end of the lesson	2.68	0.70	-0.21	0.75	0.0003	0.0006
IND62	T adapts the assignments and processing to the relevant differences between Ps	2.61	0.91	-0.10	0.38	-0.75	1.37
IND92	The teaching environment is educational and contemporary	2.32	0.77	0.29	1.04	-0.13	0.24
IND91	T ensures that classroom layout supports the P activities	2.21	0.72	1.00	*3.61**	1.28	*2.35**
IND100	I judge the overall quality of teaching as:	2.93	0.74	-0.72	*2.61**	0.92	1.70

Appendix X: Frequency table for QoT Indicators

Indicator No.	Indicator Description (T = The teacher; S= student)	Predominately weak	More weaknesses than strengths	More strengths than weaknesses	Predominantly strong
IND43	T gives feedback to Ps	1.32%	7.89%	44.74%	46.05%
IND84	T ensures effective classroom management	5.26%	11.84%	32.89%	50.00%
IND41	T gives clear instructions and explanations	1.32%	7.89%	55.26%	35.53%
IND42	T gives clear explanations of the learning materials and the assignments	1.32%	15.79%	46.05%	36.84%
IND14	T shows respect for the Ps in behaviour and language use	5.26%	9.21%	51.32%	34.21%
IND83	T uses learning time efficiently	5.26%	11.84%	47.37%	35.53%
IND81	T ensures the orderly progression of the lesson	3.95%	22.37%	50.00%	23.68%
IND11	T ensures a relaxed atmosphere	3.95%	18.42%	43.42%	34.21%
IND13	T supports the self-confidence of Ps	6.58%	19.74%	39.47%	34.21%
IND31	T clarifies the lesson objectives at the start of the lesson	1.32%	10.53%	73.68%	14.47%
IND12	T promotes mutual respect	7.89%	19.74%	38.16%	34.21%
IND51	T involves all Ps in the lesson	2.63%	30.26%	32.89%	34.21%
IND52	T makes use of teaching methods that activate the Ps	6.58%	21.05%	43.42%	28.95%
IND71	T ensures that the teaching materials are orientated towards transfer	5.26%	21.05%	48.68%	25.00%
IND82	T gives a well structured lesson	3.95%	17.11%	44.74%	34.21%
IND24	There is good individual involvement by the Ps	7.89%	21.05%	42.11%	28.95%
IND22	T stimulates the independence of Ps	5.26%	21.05%	51.32%	22.37%
IND21	T ensures cohesion	5.26%	26.32%	43.42%	25.00%
IND72	T stimulates the use of control activities	11.84%	21.05%	42.11%	25.00%
IND73	T provides interactive instruction and activities	13.16%	18.42%	50.00%	18.42%
IND23	T promotes cooperation between Ps	7.89%	34.21%	38.16%	19.74%
IND61	T adapts the instruction to the relevant differences between Ps	9.21%	26.32%	50.00%	14.47%
IND32	T evaluates whether the objectives have been achieved at the end of the lesson	3.95%	32.89%	53.95%	9.21%
IND62	T adapts the assignments and processing to the relevant differences between Ps	11.84%	32.89%	38.16%	17.11%
IND92	The teaching environment is educational and contemporary	11.84%	51.32%	30.26%	6.58%
IND91	T ensures that classroom layout supports the P activities	9.21%	68.42%	14.47%	7.89%
IND100	I judge the overall quality of teaching as:	5.26%	14.47%	61.84%	18.42%

Appendix XI: Results of a data-driven solution for QoT model with standardised coefficients

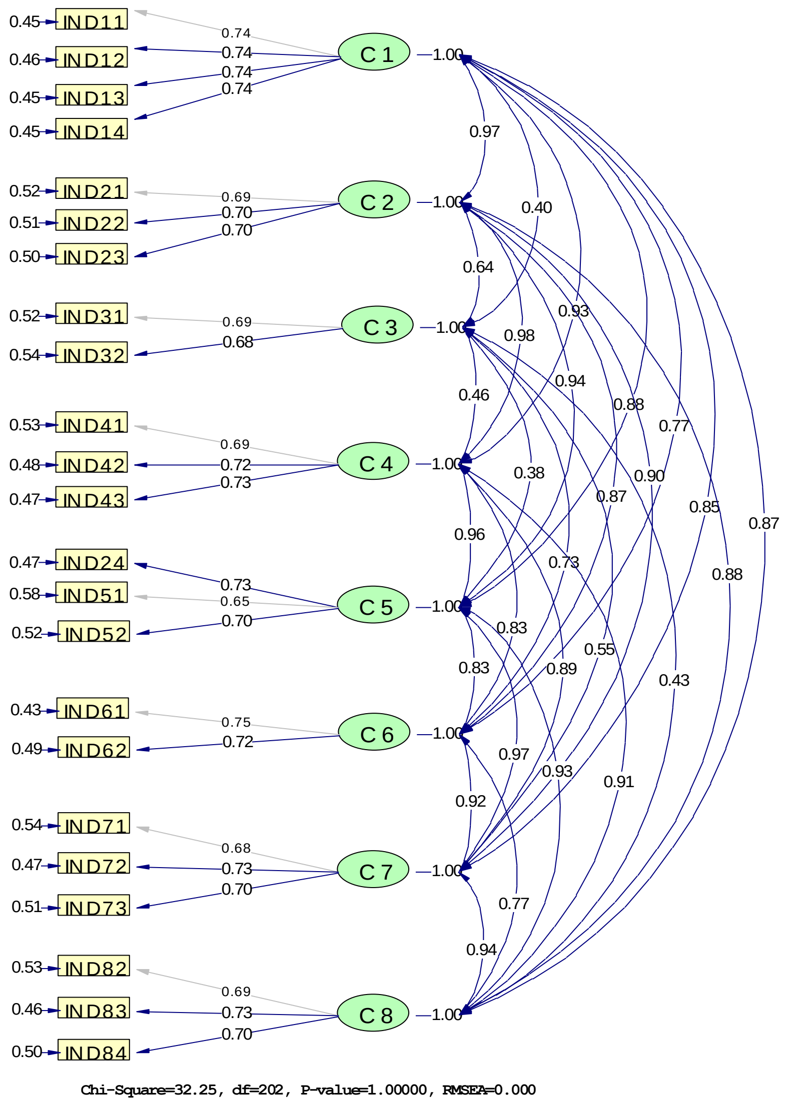


Degrees of Freedom = 132
Minimum Fit Function Chi-Square = 33.95 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 34.25 (P = 1.00)
Estimated Non-centrality Parameter (NCP) = 0.0
90 Percent Confidence Interval for NCP = (0.0; 0.0)
Minimum Fit Function Value = 0.45
Population Discrepancy Function Value (F0) = 0.0
90 Percent Confidence Interval for F0 = (0.0; 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0; 0.0)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
Expected Cross-Validation Index (ECVI) = 2.80
90 Percent Confidence Interval for ECVI = (2.80 ; 2.80)
ECVI for Saturated Model = 4.56, ECVI for Independence Model = 19.11
Chi-Square for Independence Model with 153 Degrees of Freedom = 1397.16
Independence AIC = 1433.16, Model AIC = 112.25, Saturated AIC = 342.00
Independence CAIC = 1493.11, Model CAIC = 242.15, Saturated CAIC = 911.56
Normed Fit Index (NFI) = 0.98, Non-Normed Fit Index (NNFI) = 1.09, Parsimony Normed Fit Index (PNFI) = 0.84
Comparative Fit Index (CFI) = 1.00
Incremental Fit Index (IFI) = 1.08
Relative Fit Index (RFI) = 0.97
Critical N (CN) = 382.52
Root Mean Square Residual (RMR) = 0.089, Standardized RMR = 0.054
Goodness of Fit Index (GFI) = 0.95, Adjusted Goodness of Fit Index (AGFI) = 0.94,
Parsimony Goodness of Fit Index (PGFI) = 0.73

Appendix XII: Fit Indices Generated by LISREL 8.72 for QoT CFA model

Degrees of Freedom = 296
Minimum Fit Function Chi-Square = 62.85 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 72.06 (P = 1.00)
Estimated Non-centrality Parameter (NCP) = 0.0
90 Percent Confidence Interval for NCP = (0.0; 0.0)
Minimum Fit Function Value = 0.84
Population Discrepancy Function Value (F0) = 0.0
90 Percent Confidence Interval for F0 = (0.0; 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0; 0.0)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
Expected Cross-Validation Index (ECVI) = 5.41
90 Percent Confidence Interval for ECVI = (5.41; 5.41)
ECVI for Saturated Model = 9.36
ECVI for Independence Model = 45.60
Chi-Square for Independence Model with 325 Degrees of Freedom = 3368.33
Independence AIC = 3420.33
Model AIC = 182.06
Saturated AIC = 702.00
Independence CAIC = 3506.93
Model CAIC = 365.25
Saturated CAIC = 1871.09
Normed Fit Index (NFI) = 0.98
Non-Normed Fit Index (NNFI) = 1.08
Parsimony Normed Fit Index (PNFI) = 0.89
Comparative Fit Index (CFI) = 1.00
Incremental Fit Index (IFI) = 1.08
Relative Fit Index (RFI) = 0.98
Critical N (CN) = 425.26
Root Mean Square Residual (RMR) = 0.094
Standardized RMR = 0.050
Goodness of Fit Index (GFI) = 0.93
Adjusted Goodness of Fit Index (AGFI) = 0.92
Parsimony Goodness of Fit Index (PGFI) = 0.79

Appendix XIII: A CFA model based on eight theoretical criteria and 23 indicators and modification indices with standardised coefficients (N=76)



C 1 =
Safe and orderly school climate
(Cronbach's alpha=0.98)

C 2 =
Stimulating learning climate
(Cronbach's alpha=0.93)

C 3 =
Clear Objectives
(Cronbach's alpha=0.76)

C 4 =
Clear instruction
(Cronbach's alpha=0.92)

C 5 =
Activating Pupils
(Cronbach's alpha=0.91)

C 6 =
Adaptation of teaching
(Cronbach's alpha=0.93)

C 7 =
Teaching learning strategies
(Cronbach's alpha=0.93)

C 8 =
Effective classroom organization
(Cronbach's alpha=0.93)

Degrees of Freedom = 202
Minimum Fit Function Chi-Square = 30.36 (P = 1.00)
Normal Theory Weighted Least Squares Chi-Square = 32.25 (P = 1.00)
Estimated Non-centrality Parameter (NCP) = 0.0, 90 Percent Confidence Interval for NCP = (0.0 ; 0.0)
Minimum Fit Function Value = 0.40
Population Discrepancy Function Value (F0) = 0.0, 90 Percent Confidence Interval for F0 = (0.0 ; 0.0)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.0),
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00
Expeted Cross-Validation Index (ECVI) = 4.67, 90 Percent Confidence Interval for ECVI = (4.67 ; 4.67)
ECVI for Saturated Model = 7.36, ECVI for Independence Model = 48.74
Chi-Square for Independence Model with 253 Degrees of Freedom = 3609.54
Independence AIC = 3655.54, Model AIC = 180.25, Saturated AIC = 552.00
Independence CAIC = 3732.14, Model CAIC = 426.73, Saturated CAIC = 1471.28
Normed Fit Index (NFI) = 0.99, Non-Normed Fit Index (NNFI) = 1.06
Parsimony Normed Fit Index (PNFI) = 0.79
Comparative Fit Index (CFI) = 1.00
Incremental Fit Index (IFI) = 1.05
Relative Fit Index (RFI) = 0.99
Critical N (CN) = 622.76
Root Mean Square Residual (RMR) = 0.060, Standardized RMR = 0.027
Goodness of Fit Index (GFI) = 0.96, Adjusted Goodness of Fit Index (AGFI) = 0.95,

Appendix XIV: Factor loadings and reliability of a CFA model based on eight theoretical criteria and 23 indicators and modification indices (N=76)

Factor Name	Indicator No.	Indicator Description (T = T ; P = pupil)	Factor Loading	Unstand-ardised Estimate	Square multiple correlation
Safe and orderly school climate	**IND11**	T ensures a relaxed atmosphere	0.74	Fixed at 1	0.55
	IND12	T promotes mutual respect	0.74	1.21	0.54
	IND13	T supports the self-confidence of Ps	0.74	1.13	0.55
	IND14	T shows respect for the Ps in behaviour and language use	0.74	1.77	0.55
Stimulating learning climate	**IND21**	T ensures cohesion	0.69	Fixed at 1	0.48
	IND22	T stimulates the independence of Ps	0.70	1.17	0.49
	IND23	T promotes cooperation between Ps	0.70	0.96	0.50
Clear Objectives	**IND31**	T clarifies the lesson objectives at the start of the lesson	0.69	Fixed at 1	0.48
	IND32	T evaluates whether the objectives have been achieved at the end of the lesson	0.68	0.71	0.46
Clear instruction	**IND41**	T gives clear instructions and explanations	0.69	Fixed at 1	0.47
	IND42	T gives clear explanations of the learning materials and the assignments	0.72	0.74	0.52
	IND43	T gives feedback to Ps	0.73	1.06	0.53
Activating Pupils (Cronbach's alpha=0.91)	*IND24*	*There is good individual involvement by the Ps*	0.73	1.95	0.53
	IND51	T involves all Ps in the lesson	0.65	Fixed at 1	0.42
	IND52	T makes use of teaching methods that activate the Ps	0.70	1.75	0.48
Adaptation of teaching	**IND61**	T adapts the instruction to the relevant differences between Ps	0.75	Fixed at 1	0.57
	IND62	T adapts the assignments and processing to the relevant differences between Ps	0.72	0.87	0.51
Teaching learning strategies	**IND71**	T ensures that the teaching materials are orientated towards transfer	0.68	Fixed at 1	0.46
	IND72	T stimulates the use of control activities	0.73	1.43	0.53
	IND73	T provides interactive instruction and activities	0.70	1.59	0.49
Effective classroom organization	**IND82**	T ensures the orderly progression of the lesson	0.69	Fixed at 1	0.47
	IND83	T uses learning time efficiently	0.73	1.52	0.54
	IND84	T ensures effective classroom management	0.70	1.46	0.50
Reliability of 24 indicators without Indicators 91 and 92: Cronbach's alpha =0.98					

Appendix XV: Correlations between the underlying dimensions of the CVCP (ISTOF) Model and All QoT Models

	Overall quality of teaching	Meta-Cognitive Skills Teaching	Classroom Manage-ment & Climate	Differentiation & Support	Clarity & Logic of Presentation	Student Engage-ment	Strategies to enhance learning and lesson focus	Safe & orderly school climate	Stimulating learning climate	Clear Objectives	Clear instruction	Activating Pupils	Adaptation of teaching	Teaching learning strategies	Effective classroom organi-sation	Efficient classroom management	Safe and stimulating learning climate	Clear Instruction	Adaptation of teaching	Teaching learning strategies
Overall quality of teaching	1.00	.63	.74	.80	.73	.82	.67	.75	.82	.60	.74	.70	.76	.82	.85	.85	.79	.80	.76	.82
Meta-Cognitive Skills Teaching	.63	1.00	.67	.62	.67	.67	.49	.68	.70	.39	.61	.58	.68	.79	.69	.69	.70	.65	.68	.79
Classroom Management & Climate	.74	.67	1.00	.76	.67	.71	.47	.83	.79	.28	.72	.77	.65	.81	.87	.87	.82	.77	.65	.82
Differentiation & Support	**.80**	.62	.76	1.00	.74	.82	.67	.72	.82	.57	.69	.71	.78	.84	.77	.77	.76	.78	.78	.84
Clarity & Logic of Presentation	.73	.67	.67	.74	1.00	.78	.63	.66	.76	.48	.69	.72	.71	.78	.81	.81	.71	.76	.71	.78
Student Engagement	**.82**	.67	.71	**.82**	.78	1.00	.65	.75	.87	.52	.79	.82	.79	.85	.81	.81	.81	.87	.79	.85
Strategies to enhance learning & lesson focus	.67	.49	.47	.67	.63	.65	1.00	.46	.52	.60	.42	.43	.64	.64	.51	.51	.49	.51	.64	.64
Safe & orderly school climate	.75	.68	**.83**	.72	.66	.75	.46	1.00	.90	.34	.85	.78	.70	.79	.80	.80	.98	.85	.70	.79
Stimulating learning climate	**.82**	.70	.79	**.82**	.76	**.87**	.52	**.90**	1.00	.50	.89	.85	.80	.87	.85	.85	.96	.93	.80	.87
Clear Objectives	.60	.39	.28	.57	.48	.52	.60	.34	.50	1.00	.37	.26	.61	.45	.43	.43	.43	.48	.61	.45
Clear instruction	.74	.61	.72	.69	.69	.79	.42	**.85**	**.89**	.37	1.00	.83	.73	.79	.84	.84	.88	.96	.73	.79
Activating Pupils	.70	.58	.77	.71	.72	**.82**	.43	.78	**.85**	.26	**.83**	1.00	.71	.82	.81	.81	.81	.92	.71	.82
Adaptation of teaching	.76	.68	.65	.78	.71	.79	.64	.70	**.80**	.61	.73	.71	1.00	.85	.73	.73	.76	.80	1.00	.85
Teaching learning strategies	**.82**	.79	**.81**	**.84**	.78	**.85**	.64	.79	**.87**	.45	.79	**.82**	**.85**	1.00	.86	.86	.82	.86	.85	1.00
Effective classroom organisation	**.85**	.69	**.87**	.77	**.81**	**.81**	.51	**.80**	**.85**	.43	**.84**	**.81**	.73	**.86**	1.00	1.00	.83	.88	.73	.86
Efficient classroom management	**.85**	.69	**.87**	.77	**.81**	**.81**	.51	**.80**	**.85**	.43	**.84**	**.81**	.73	**.86**	**1.00**	1.00	.83	.88	.73	.86
Safe and stimulating learning climate	.79	.70	**.82**	.76	.71	**.81**	.49	**.98**	**.96**	.43	**.88**	**.81**	.76	**.82**	**.83**	**.83**	1.00	.90	.76	.82
Clear Instruction	**.80**	.65	.77	.78	.76	**.87**	.51	**.85**	**.93**	.48	**.96**	**.92**	**.80**	**.86**	**.88**	**.88**	**.90**	1.00	.80	.86
Adaptation of teaching	.76	.68	.65	.78	.71	.79	.64	.70	**.80**	.61	.73	.71	**1.00**	**.85**	.73	.73	.76	**.80**	1.00	.85
Teaching learning strategies	**.82**	.79	**.82**	**.84**	.78	**.85**	.64	.79	**.87**	.45	.79	**.82**	**.85**	**1.00**	**.86**	**.86**	**.82**	**.86**	**.85**	1.00