Golightly, David and Dadashi, Nastaran (2016) The characteristics of railway service disruption: implications for disruption management. Ergonomics . ISSN 1366-5847

Eliciting variation within railway disruption management

Golightly, D.*, Dadashi, N.

Human Factors Research Group

Faculty of Engineering

University of Nottingham

Nottingham

UK

NG7 2RD

Email - {david.golightly, nastaran.dadashi}@nottingham.ac.uk

Telephone - + 44 115 9514070

* Corresponding author

Acknowledgments

This work was supported by EU FP7 under the On-time project (Grant agreement FP7 - SCP0 – GA – 2011 – 265647)

Abstract

Rail disruption management is central to operational continuity and customer satisfaction. Disruption is not a unitary phenomenon - it varies by time, cause, location and complexity of coordination. Effective, user-centred technology for rail disruption must reflect this variety. A repertory grid study was conducted to elicit disruption characteristics. Construct elicitation with a group of experts (n=7) captured 26 characteristics relevant to rail disruption. A larger group of operational staff (n=28) rated 10 types of rail incident against the 26 characteristics. The results revealed distinctions such as business impact and public perception, and the importance of management of the disruption over initial detection. There were clear differences between those events that stop the traffic, as opposed to those that only slow the traffic. The results also demonstrate the utility of repertory grid for capturing the characteristics of complex work domains.

Practitioner Summary

The aim of the paper is to understand how variety in rail disruption influences socio-technical design. It uses repertory grid to identify and prioritise 26 constructs, and group 10 disruption types, identifying critical factors such as whether an incident stops or merely slows the service, and business reputation.

Keywords

Railways, disruption, repertory grid, expertise

1. Introduction

Delay and disruption on the railways has a financial cost in terms of lost patronage and delay attribution penalties (National Audit Office, 2008). The high priority placed by travellers on reliability (Thomas et al., 2006) means delay both undermines existing passenger confidence and inhibits potential mode shift from the car to the train. Tackling disruption is therefore a critical challenge if the railways are to offer a viable, sustainable travel option (EU, 2011).

The following paper presents a repertory grid approach to explore and identify the characteristics of railway disruption with groups of railway incident practitioners. This was with the aim of contributing to a more nuanced understanding of rail disruption for an EU FP7 project (On-time) developing new tools for rail capacity management. The specific study objectives were (1) to understand the major characteristics of rail disruption as perceived by operational staff and (2) determine any variations or regularities of these characteristics over different disruption types. By doing so, the intention was to understand the variety (Flach, 2009) inherent within events such as rail disruptions, so that this variety can be adequately reflected in new technology design, new processes or new organisational configurations. As such, this study contributes new data on disruption management within rail, and transport generally, and provides further insight into the personal, cognitive work of those people involved in disruption management (Farrington-Darby et al., 2006). The study also serves as an illustration of both a method and outputs that allow variety to be captured in a manner appropriate for cognitive work systems such as rail control.

2. Background

2.1 The nature of rail disruption management

Transport disruption can be defined as an event that leads to significant re-planning of a transport service (Pender et al., 2012). Specifically, for this paper, the focus is on unplanned disruption, as opposed to planned engineering. Unexpected events such as train failures or overhead line dewirements, or widespread restrictions due to high-winds or flooding, are a daily occurrence on the railways. Events such as these may take hours to resolve, and cause significant delay and frustration to passengers and freight customers. While this paper focusses on the experience of the railways of Great Britain, such disruption is common to all forms of heavy and light railway, globally (Pender et al., 2012), and some form of disruption is a reality for all forms of transportation system including aviation (Clausen, 2007) and highways (Koorey et al., 2008).

Various approaches are proposed to support the management of incidents and disruption on the railways. One is to prevent disruptions, for example by predicting potential infrastructure faults (Dadashi et al., 2014) or through preventing fatality and trespass (Ryan, 2013). It is not feasible, however, to anticipate all forms of potential failures or incidents. An alternative approach is therefore to manage and mitigate disruption more effectively as it occurs. Observational (Farrington-Darby et al., 2006) and incident analyses (Golightly et al., 2013a) have captured the critical activities of rail disruption management. These include, first, notification and containment of an event, then coordinating and mobilising any trackside response such as maintenance staff or emergency services, re-planning the service to maintain continuity during the event including informing passengers, restoration of infrastructure to service and, finally, return to normal. All of this is conducted within the key performance criteria of safety, punctuality and cost effectiveness (Millen et al., 2011). In Great Britain, this process is led by incident controllers who have central responsibility for coordinating response. These responsibilities include making operational decisions to manage the rectification of disruption and to organise and disseminate alternative service arrangements (further detail on the orchestration of incident control is presented in Farrington-Darby et al., 2006). These decisions are acted upon, and informed by, signallers, train crew, station staff, maintenance staff and sometimes external parties such as the emergency services. Naturally, a major set of stakeholders in the disruption are the passengers themselves, delayed in trains or stations. Two of the key roles, signallers (Golightly et al., 2010) and incident controllers (Farrington-Derby et al., 2006) are physically remote from the disruption and, in the case of incident controllers, have no direct interface or means of control, acting through others via phone calls, emails and Information Technology. While roles vary for other countries (Golightly et al., 2013b), the major functions embedded within disruption management systems are typical for any rail system.

Technology forms a central part of the strategy for the effective management of disruption events. Proposed solutions include traffic re-planning and operational decision-making tools (Kauppi et al., 2006; Pasquier, Rezillon and Pomerol, 2000), complemented by support tools for short-term crew and rolling stock re-planning (Jespersen-Groth et al., 2009). Re-planning tools such as predictive, interactive train graphs (Kauppi et al., 2006), could not only help individual decision makers to plan alternatives, but can act as a visual, shared representation so that a number of relevant stakeholders can see the proposed plan (Male and Baber, 2014). Technology also applies to passengers and the need to provide accurate information on the duration of delays and potential alternatives. This kind of information is transmitted through traditional means such as station staff, but also through more recent forms of technology such as mobile travel applications (Lyons, 2006) and social media (Pender et al., 2014; Golightly and Durk, accepted).

2.2 Variation in disruption

Disruption is not, however, a unitary phenomenon. First, incidents and disruptions are not the same thing. Incidents, or events such as infrastructure failure, do not always lead to disruption and this can be down to the way the incident is managed, minimising or avoiding altogether any impact on timetabled services. Therefore, while the overall aim of the study was to understand the factors, decisions and processes that may or may not lead to disruption, the starting point for the elicitation was the potential cause of that disruption – termed here as an 'incident', in keeping with common GB rail nomenclature (for example, events are categorised and recorded on the 'Control Centre Incident Log' [Golightly et al., 2013a]). Incidents and resulting disruptions are dynamic, with different stages requiring different types of information and coordination amongst various stakeholders. For rail disruption, the opening stages, where the event is first noticed and immediate containment action is applied, may be brief (in the order of a few minutes) but critical if the situation is not to escalate (Belmonte et al., 2011). However, tools such as crew and rolling stock rescheduling may only be relevant to a later phase of re-planning to put alternative service plans in place. Also, railway disruptions do not always follow a single, linear process. Major incidents such as a break in the overhead electrification may take many hours to resolve. During the course of such incidents there may be many cycles of investigation, rectification and adaption of plans before normal service is restored (Golightly et al., 2013a).

Another consideration is that incidents vary – their causes, timing and location can all influence the choice of effective strategy. For example, at the busiest times on the busiest parts of the network the most adaptive course of action may be to apply temporary (though safe) repairs to keep a partial service running, and complete the fix overnight when there are fewer train services.

The handling of incidents involving the railways requires close coordination between many parties across different organisations, such as different train operating companies, some of whom have conflicting priorities (Steenhuisen et al., 2009). For the most severe incidents and emergencies, these groups are rapidly configured between parties who do not regularly cooperate, such as the emergency service or air ambulance (Smith and Dowd, 2000) and may often be physically distributed, communicating primarily through telephones. These types of incidents may draw on different skills and competence compared to ‘routine’ disruption. For example, even highly experienced control staff feel least confident when dealing with third party emergency services during incidents (NAO, 2008; Cheng and Tsai, 2011).

2.3 Expressing complexity

All of these different sources of variation, and context, are acknowledged to make rail disruption very unpredictable. Any contingency plan, whether that is provided by technology or some other means (paper-based contingency plans are still the norm in many rail control centres), can only be a template. Contextualising decision support to reflect the specifics of an event is a major challenge (Lenior et al., 2006). In the naturalistic observation of incident controllers, Farrington-Darby et al. (2006) found that the work required significant elements of problem-solving and learning, as well as knowledge of the social context (people and responsibilities) that can put plans into effect.

From a cognitive systems engineering perspective (Hollnagel and Woods, 2005; Flach, 2009) such complexity can be viewed as variety within the domain to be controlled. Effective control can only be achieved through understanding the nature of this variety and making it available and salient through the control system. Therefore new technology, processes and organisational configurations (which often change at the same time [Wilson and Norris, 2006]), should reflect this variety, making it salient to agents, both human and automated, involved in the decision making process.

Feltovich et al. (2004) identify a number of categories of complexity in socio-technical systems. These include the system being dynamic and non-linear, comprising interactive and heterogeneous components, and being conditional depending on context. The review of rail disruption above highlights that all of these aspects of complexity are present, but it would be useful to understand the specifics of this complexity in a more regular or structured manner. This would help to define the variables that might need to be expressed within a tool if it is to provide sufficient support, or set bounds on the type of disruption that realistically can be supported by any given approach.

One means is through understanding current operational approaches to managing disruption, and complexity perceived by operational staff. This would be with a view to seeing where regularities lie. These regularities might be in both the overt characteristics of the event (e.g. its causes, location etc.) and also in the adaptive strategies and heuristics used to manage the event (Flach, 2009). Those strategies are dependent on factors associated with both the cause and means of managing a disruption. Eliciting these factors was the aim of the study described in the rest of this paper.

2.4 Repertory grid

While observation (Farrington-Darby et al., 2006), Critical Decision Method (CDM) (Dadashi et al., 2013) and post-hoc incident analysis (Smith and Dowd, 2000; Golightly et al., 2013a) can contribute to our understanding of rail incidents, they have certain limitations. Observation is dependent on being present while an incident takes place, CDM is extremely time consuming if the intention is to cover multiple incident types with multiple experts, and post-hoc analyses to date have looked at organisational rather than individual factors, as communication records and incident reports rarely give insight into individual strategies. The requirement for this study was to find a structured technique to elicit, compare and contrast characteristics of multiple disruption types, as the people at the front line of incident management perceive them.

Repertory grid technique is rooted in Kelly’s Personal Construct Theory (Bannister and Fransella, 2002). The theory postulates that people have a set of core hypotheses or expectations about the world that shape their interpretation of the environment and events. It is assumed that these expectations, termed ‘constructs’, are bi-polar with judgements of similarity between the world and the constructs being made along a continuum (Edwards et al., 2009). Also, Personal Construct Theory proposed that constructs are organised as a hierarchical system, and by examining groups of related constructs it may be possible to find superordinate constructs that shape a person’s understanding of a domain. The aim of the repertory grid technique, therefore, is to elicit a set of constructs and, through comparison, identify higher order beliefs, values or rules.

The technique has been transposed from the personal and therapeutic setting, for use in knowledge elicitation and subsequently applied in the human factors arena for uses such as capturing product design characteristics (Baber, 1996; Kuru and Erbug, 2013), human computer interaction (Hassenzahl & Wessler, 2000), road and driving characteristics (Riemersma, 1988) and for identifying the characteristics of workload in rail signalling (Pickup et al., 2010).

The aims of the repertory grid technique, as applied in this study, were to understand constructs relevant to the perception and management of rail disruption with the intention of determining similarities between different incident types. Repertory grid was chosen as the appropriate technique for this elicitation study, specifically because it allowed comparison across multiple types of event, which could draw out regularities and idiosyncrasies across several critical types of disruption.

3. Method overview

The repertory grid technique is applied as follows. The first stage, construct elicitation, explores elements in groups of three. These elements are exemplars for a given domain. In the current case, the set of elements comprised a number of different incident types, presented in Table 1.These incidents were selected from series of discussions with experts working in rail disruption from across the rail sector. The incidents represent a range of cases that typically have a major impact on rail operations either because of their immediate severity (e.g. in the case of an overhead line dewirement), or because of their incremental effect on rail system performance (e.g. on-going passenger loading issues across a region because of icy platforms).

[Table 1 about here]

The researcher then presents the elements to participants and asks them to comment on a characteristic, or 'construct' linking two of the elements, but differentiating the third. To give an example, when presented with a fatality, a points failure and passenger loading issues, a participant may comment that the first two have a specific location, whereas passenger loading issues may be widespread during poor weather. Thus 'disruption location' has been elicited as a construct, with 'specific' and 'widespread' identified as the two poles. The process is repeated, theoretically until all combinations of three elements have been presented and all possible constructs have been elicited, though pragmatically the elicitation may stop once the participant is unable to provide any new constructs (Baber, 1996).

The second stage consists of rating the constructs against each of the elements, either as a binary 'yes' or 'no' as to whether the construct is relevant to that element or a scaled rating of elements for each construct. Statistical analysis, through approaches such as cluster analysis or factor analysis can be used to identify groups of constructs that together indicate higher order or underpinning constructs. The analysis can also identify variation between element types over these constructs, for example to identify which product rates most highly for comfort or appearance (see Edwards et al., 2009 for a useful review of the different variants of repertory grid application).

Due to the time and availability of the railway experts that participated in this study, it was not possible to perform the first and second stage (construct elicitation and construct rating) with one group of participants. Also, there was a desire to capture the opinion of railway practitioners from a number of roles to increase the applicability of the results, using a ‘fixed grid’ (Edwards et al., 2009). This two stage approach to applying repertory grid has a precedent in the ergonomics literature (Riemersma, 1988). These two stages are reported separately in the following sections. The ethics panel at the Faculty of Engineering, University of Nottingham, approved both of these stages.

4. Construct Elicitation

4.1 Method

4.1.1 Participants

Seven members of staffs from the Great Britain railways agreed to take part in the study. These experts were from various roles including signallers (two participants), incident controllers (three participants) and station managers (two participants), with an average of 20 years of experience. Participants were approached via contact with a senior manager at the railway infrastructure manager who facilitated recruitment.

4.1.2 Apparatus

The ten incidents selected as elements (see Table 1) were presented individually on cards. Identified constructs were recorded in an Microsoft Excel™ spread sheet in real-time during the interviews. iTalk™ digital voice recorder was used to record participants' comments. As one of the participants declined to be recorded, there were six recordings that were subsequently transcribed.

4.1.3 Procedure

After agreeing to take part in the study and signing a consent form, participants were presented with groups of three cards, representing incident types. Participants were asked to group two of the three over a criterion of their choice, to describe that criterion, and then offer a rationale for their decision. The criterion was then recorded as a construct in a Microsoft Excel™ spread sheet. A second set of three incidents was presented, and the process repeated until the participant was no longer able to offer a new construct. Typically this took around 30 minutes per participant.

4.2 Results

A total of 142 bipolar constructs were identified from the excel spreadsheets, supported by transcripts. Some constructs were noted by a number of participants, for example six participants stated 'the number of people involved (i.e. multi-agent control/single-agent control)' as one of the constructs that defines a railway incident. Also, in some cases a construct was identified more than once by the same participant despite the intention that each of their constructs should be unique. Removing duplications within and between participants led to 26 constructs. These, as well as their contrasting poles, are presented in Table 2. The table also presents the number of participants who mentioned the constructs, total references to those constructs and an example of incident trio that would elicit the construct. These examples are drawn from the participant transcripts.

[Table 2 about here]

From this list it was possible to build a fixed grid of the 26 constructs by 10 incident types, which served as the materials for the next stage of the study: construct rating.

5. Construct Rating

5.1 Method

5.1.1 Participants

The second group of participants comprised 28 railway operational staff attending a workshop on new traffic management technology, facilitated by the rail infrastructure manager. Participants had an average of 20 years of experience with various responsibilities including signalling (10 participants), local operations management (9 participants), incident control managers (6 participants), an electrical control room operator (1 participant), a senior modelling specialist (1 participant), and a station manager (1 participant).

5.1.2 Apparatus

The 26 constructs identified during the previous stage were presented as a paper-based matrix with 26 constructs down the side, and 10 incident types across the top. The matrix was printed and could be completed using a pen. The matrix also gave an example of how to complete the form, as well as ethical and consent information. It should be noted that both poles were presented for each construct with one scored as ‘1’ and the other as ‘9’, and participants required to score on or between these two poles. The pole presented as ‘1’ was generally the negative construct. For clarity, Figure 1 presents an extract from the matrix as it was presented to participants.

5.1.3 Procedure

The purpose of the survey was explained to the participants attending the workshop. The survey was then distributed and participants were requested to rate each incident type (in other words, to write a value in each cell of the matrix) against the constructs with a relevance rating from 1 to 9. Participants were not required to rank incidents, that is, they did not have to put a unique score for each incident for a given construct, and were allowed to put duplicate values. For example, a participant might give both points failure *and* signal failure a rating of 1 for ‘time of day’ indicating that time of day was not relevant to either incident type. Completing the survey took around 15-20 minutes and all responses were collected at the end.

5.2 Results

Table 3 shows participants mean scores (Min=1 and Max=9) for each of the constructs in relation to various incident types, as well as total mean and standard deviations for element and construct. Constructs are ranked top (greatest) to bottom in terms of mean score, with elements ranked left (greatest) to right. NB These rankings are only illustrative. For constructs, in particular, scores do not always follow 1 = ‘least / smallest / least important’ etc. and 9 = ‘best / biggest / most important’ convention, and the reader should refer the left hand column of Table 3.

[Table 3 about here]

Analysis of Variance (ANOVA) was conducted to facilitate the review and comparison of the 28 grids with each other. This was to confirm that the participants’ collective opinion was not contradictory and the values are representative (Edwards et al., 2009). The comparison between the constructs in the grids (F (324,26)=12.15) and the correlation between construct angles is (r^2=0.92) indicates participants have common opinions with regards to the constructs associated with the railway incidents.

One means to understand groupings of constructs is hierarchical cluster analysis. Pairs of similar items are grouped based on a metric such as Euclidean distances. These pairs may in turn be paired with other items or pairs, forming higher order clusters. A dendogram (see figures 2 and 3) represents the stages of pairing. Typically, further pairing is ignored when very large distances are being grouped together, as indicated by a large distance across the x axis of the dendogram or the agglomeration schedule generated during the cluster process (Clatworthy et al., 2005). Therefore, the point at which further pairing is discarded, and the interpretation of the meaning of clusters, is somewhat through the interpretation of the investigator (Edwards et al., 2009).

Hierarchical cluster analysis was used to identify groups of incident types that shared similar scores according to the element ratings. This was conducted in SPSS 20.0 using squared Euclidean distances and between groups linkage method. The dendogram is shown in Figure 2. Visual inspection of the agglomeration schedule suggested a three cluster solution. These were

Cluster 1 – Points failure, track circuit failure, signal failure and power failure.

Cluster 2 – Fatality, OHL dewirement.

Cluster 3 – Station overrun, freight adhesion, speed restriction, passenger loading.

[Figure 2 about here]

Hierarchical cluster analysis was also used to identify groups of the 26 characteristic constructs, by how they were similarly scored against incident types, using squared Euclidean distances and between groups linkage method. The dendogram is shown in Figure 3. Visual inspection of the agglomeration schedule suggested a five cluster solution was the most appropriate. These clusters were

Cluster 1 – Delay, impact, incident duration, delay estimation, investigation, number of people, access, danger

Cluster 2 – Business reputation, time of day, social media, geographical distribution, variability, importance of timetable knowledge, visibility of features, likelihood, technological aids

Cluster 3 – Human behaviour, driver behaviour, weather, seasons

Cluster 4 – Diagnosis, handling, noticeability, handling location

Cluster 5 – Effect on service

[Figure 3 about here]

6. Discussion

6.1 Constructs

The final list of 26 constructs demonstrates a range of factors considered relevant to disruption by participants. The interpretation of these constructs has been aided by referring to the transcripts from participants at stage 1. Relatively few of the constructs are concerned with the external characteristics of the event such as location and time of day or causal factors such as human behaviour and driver behaviour. Instead, many factors are concerned with the management of the incident and the organisation of incident response. In terms of models of disruption, it supports Farrington-Darby et al., (2006) by suggesting a greater overhead on decision-making (and potentially a greater need for support) at the incident management and replanning stage as opposed to initial response phase. As such, this supports and elaborates on Lenior et al., (2006) by demonstrating the range of factors, beyond just the external, observable facts of the disruption, that operational staff use in their assessment of an disruptive event.

Several of the characteristics of the constructs are striking. First, despite being identified by only one participant at stage one, business reputation came out as an important consideration at the rating stage. This result highlights the importance of more commercial factors, not just in terms of strategy, but in terms of the practical decision making during disruption. Similarly, social media is included as a construct, and is further evidence to support Pender et al., (2014) in the growing role of social media in disruption management. It is worth noting that since this study took place (in early 2013), almost all train operating companies in Great Britain have moved their social media teams into the rail control centre to maximise the flow of information, in both directions, between passenger and disruption management (Golightly and Durk, accepted).

The construct clusters should be treated with some caution. The aim of the study was to capture variability, whereas clustering is an attempt to collapse that variability. That said, the clusters point to some interesting patterns. First, ‘effect on service’ (whether an incident stops or slows the service) is in a cluster of its own (Cluster 5). Rather than being an outlier, this suggests the importance of this construct on whether it is possible to maintain some service (see below for how ‘effect on service’ is reflected in the incident clustering). Also, constructs around danger, access and diagnosis (which all relate to trackwork), are clustered with constructs relating to delay and duration (Cluster 1). This correlates with the findings in Golightly et al., (2013a) and Cheng and Tsai (2001) where access to track greatly increases the complexity and demands of the incident due to the need to ensure trackworkers or evacuating passengers are safe from train movements and that electrical supplies have been isolated . Both human and driver behaviour and weather factors also form a cluster (Cluster 3), as these are all potential causal factors.

Also, apparently similar constructs have subtly different implications For example, ‘time of day’, and ‘knowledge of timetable’, which at a surface level both seem to be the same, reflect different aspects of incidents when inspecting the ratings. Freight adhesion is shaped by knowledge of the timetable as much as other disruption events, but ‘time of day’ is not rated as important, presumably because there are no passengers and therefore of whether it is a peak period is less of a consideration. Similarly, the ‘need for investigation’ is not the same as the complexity or ‘ease of diagnosis’ as they were rated differently and assigned to different clusters. It is likely that ‘need for investigation’ is related to whether troubleshooting is required or not whereas ‘diagnosis’ is how easy it is to understand the cause of the problem. For design and process change, identifying small but subtle differences is important if designers and developers, particularly from outside the railways, are to avoid oversimplifying the task of disruption management (Feltovich et al., 2004).

6.2 Incidents

Turning to the incidents themselves, it is clear they vary in severity over a number of factors. This might be in terms of ‘effect’ (whether they stop or merely slow the service), or whether they can be handled by the operational staff in the control centre or require people on site, which also increases ‘danger’.

Clustering suggests three groups – those that stop the service (OHL and fatality), infrastructure faults (signal, points, track circuit and power failures) that may have a variable effect depending on location and severity, and then a number of disruptions (station overrun, speed restriction, freight adhesion and passenger loading) that are minor in their effect, and can normally be managed remotely by the signaller without significant multi-party involvement. These clusters also suggest that other disruption events, not included in this analysis, may have similar and therefore predictable profiles. For example, a major train failure, that stops the service, may show a similar profile and have similar control requirements to a Fatality or OHL dewirement. While only one construct explicitly related to the multi-agent nature of incident, there was much variation between incident types. This adds further support, from a different perspective, to Golightly et al (2013a) emphasising that events like fatality (7.46) or OHL (7.18) require close coordination with parties on track, whereas freight adhesion (2.79) or station overrun (2.29) may only require communication between the signaller and the driver.

There are some results that warrant further investigation. ‘Driver behaviour’ scores highly for fatality, when it would seem that there would be little scope, particularly in high speed situations for the driver to do much to avert an accident. Fatalities also cover level crossings and trackworker incidents, and this different scope may change perceptions of this factor. As well as requiring further discussion and validation with experts, this point generally illustrates the very specific use of terminology and the difficulties of interpretation for domains such as the railways.

6.3 Methodology

In terms of the quality of the data, there was a high degree (over 80%) of duplication during the initial elicitation. This suggests a level of concordance between participants in terms of the factors they considered relevant to rail disruption. At the rating stage, clusters of both incident types and constructs include items that would intuitively appear to form natural groups (e.g. a cluster of all infrastructure failures, a cluster relating to diagnosis and handling) that suggests face validity.

In terms of lessons learned regarding the methodology, repertory grid has proved a successful tool in highlighting not only the variety inherent in disruption, but also how these vary over incident types. There have also been some interesting artefacts of the manner it has been applied. By using different samples between the capture and the rating phase in the manner of Riesmerse (1988), constructs that seemed only marginally relevant at elicitation phase (e.g. business reputation), turned out to be important at the construct rating phase. While obvious duplications were removed at the elicitation phase, some overlap between terms was permitted for the rating phase. This proved to be useful as seemingly similar terms (e.g. 'timetable' and 'time of day'; 'investigation' and 'diagnosis') were given different patterns of rating depending on the incident type, or were assigned to different clusters.

6.4 Limitations

One limitation of the study is that the study is very much focussed on rail control in Great Britain. While experience on EU projects has shown there are many generalities between different countries (Golightly et al., 2013b), there are specific factors that only apply locally. For example, Sweden has a national incident co-ordination centre whereas an incident management in GB is typically conducted at the regional level. It would be useful to replicate this study elsewhere, and the matrix is available from the authors on request. Also, there is a limitation with scales used and therefor data collected. The analysis of participant responses in the second phase has broadly assumed that ratings for poles go from smaller (1) to larger (9) and therefore, implicitly, from less severe / impactful, to more severe / impactful. While this holds for the majority of construct ratings, it is maybe more tenuous for others (mode of handling: handled remotely (1), handled onsite (9)). While in this case the intention was to stay loyal to the constructs and poles elicited by the seven experts, a future variant of the fixed grid matrix may adapt construct poles to make rating scales more consistent.

A final limitation is that the study used different roles. All were experienced in disruption management, but it is possible that the signaller (with responsibilities that include safe access for trackworkers, and changing route setting to accommodate late and delayed trains) may see events differently from incident control roles who have a more strategic role in service continuity. A future study may choose to study different roles separately and in greater detail.

7. Conclusions

The aim of the study presented in this paper was to elicit characteristics relevant to decision making within rail incident management. In addition, there was an aim to understand how incident types might vary in accordance with these characteristics, and to understand if there were any groupings of incidents with shared characteristics. These aims had a practical goal of understanding factors to feed into requirements for technology or process change. In terms of knowledge, the study aimed to expand on studies such as Farrington-Darby et al. (2006) and Golightly et al. (2013a) by bringing to light the constraints and considerations that impinge on cognitive activity during disruption management.

First, this paper has identified 26 characteristics that create variety in rail disruptions. The most marked difference is around whether an incident slows or stops the service, and also there were unexpected factors that moved beyond purely operational considerations, to cover business reputation or the importance of social media. In terms of the incident types themselves, there was a clear grouping for events such as fatality and overhead line dewirement that caused a complete blockage to the service. Infrastructure faults were viewed in a similar manner to each other, with a third grouping of service disruption events that slowed but did not necessarily stop the service.

Any technology or process designer is invited to consider the potential relevance of these 26 characteristics when developing or reviewing new approaches to disruption management.

Importantly, these might be operational tools, but these 26 characteristics will also ultimately influence how the incident is managed and, as a result, how it might be perceived by passengers. Therefore, these characteristics also have relevance to the design of information and processes as they are presented to passengers or freight customers. Overall, these characteristics reflect concerns around management and prediction of incident impact, rather than capturing initial data around the response, and are more marked when the service is stopped rather than slowed.Future models of the railway control, such as the abstraction hierarchy of Millen et al (2012) need to make explicit this performance shaping constraint.

There is clear evidence incidents are different, which may be self-evident in the rail sector, but needs to be considered by those (software developers for example) approaching the sector for the first time. Importantly, extending on the organisational analysis of Golightly et al. (2013) these data show that while incidents vary, they can be grouped and offers a profile of the factors (constructs) relevant to these incident types.

The work here therefore should help in reducing the ‘reductive tendency’ of designers of complex sociotechnical systems for rail (Feltovich et al., 2004). It is hoped this paper adds weight to the argument that future transport control technology needs to reflect and accommodate variety, not just in terms of external disruption awareness, but in terms of communication and organisation (and therefore adaption) to event characteristics.

Also, the repertory grid has once again proved an effective and flexible method for eliciting valuable domain knowledge, and is an important tool for understanding complex systems and for capturing user requirements. While this work has identified a set of factors, a useful next step would be to link these factors to decisions, depending on the role of the actor involved. This would both validate the work presented here, and make it clearer how decisions are informed by disruption factors, which could in turn lead to design and process recommendations. It is not common to make use of the qualitative commentary that participants give during construct elicitation, but in our case there are several hours of detailed rationale from our experts as to *why* factors are relevant in different types of disruption. To that end, the next step for this work will be to reanalyse the qualitative date to capture some initial decision making models that can be further elaborated, probably in new interview work, to produce decision making models and requirements for design.

References

Baber, C. 1996. Repertory grid theory and its application to product evaluation. In *Usability Evaluation in Industry*, ed. P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland, pp. 157-165. Taylor and Francis, London

Bannister, D., & Fransella, F. (2002). *Inquiring man: Theory of personal constructs*. London: Psychology Press.

Belmonte, F., Schön, W., Heurley, L., & Capel, R. 2011. Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model:

An application to railway traffic supervision. *Reliability Engineering & System Safety*, *96*(2), 237-249.

Cheng, Y. H., & Tsai, Y. C. (2011). Railway-controller-perceived competence in incidents and accidents. *Ergonomics*, *54*(12), 1130-1146.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. British journal of health psychology, 10(3), 329-358.

Clausen, J. (2007). Disruption management in passenger transportation-from air to tracks. In 7th Workshop on Algorithmic Methods and Models for Optimization of Railways (pp. 30-47).

Dadashi, N., Golightly, D., & Sharples, S. 2013. Requirements elicitation for disruption management support. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, 399 - 406.

Dadashi, N., Wilson, J. R., Golightly, D., & Sharples, S. 2014. A framework to support human factors of automation in railway intelligent infrastructure. *Ergonomics*, 57(3), 387-402.

Edwards, H. M., McDonald, S., & Young, S. M. 2009. The repertory grid technique: Its place in empirical software engineering research. Information and Software Technology, 51(4), 785-798.

Farrington-Darby, T., Wilson, J. R., Norris, B. J., & Clarke, T. (2006). A naturalistic study of railway controllers. *Ergonomics*, *49* (12-13), 1370-1394.

Feltovich, P. J., Hoffman, R. R., Woods, D., & Roesler, A. (2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *Intelligent Systems*, IEEE, 19(3), 90-94.

Ferguson, E., & Cox, T. 1993. Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment*, *1*(2), 84-94.

Flach, J. M. 2009. The dynamics of experience: a search for what matters. In *Proceedings of the European conference on cognitive ergonomics (ECCE 2009). Helsinki, Finland* (pp. 11-18).

Friman, M., & Gärling, T. (2001). Frequency of negative critical incidents and satisfaction with public transport services. II. *Journal of Retailing and Consumer Services*, *8*(2), 105-114.

Golightly, D, & Durk, J. Accepted. Social media on the railways: human factors and complex control. To appear in proceedings of Ergonomics and Human Factors Conference, April, 2016.

Golightly, D., Dadashi, N., Sharples, S., & Dasigi, M. (2013a). Disruption management processes during emergencies on the railways. *International Journal of Human Factors and Ergonomics*, *2*(2-3), 175-195.

Golightly, D., Sandblad, B., Dadashi, N., Andersson, A. W., Tschirner, S., & Sharples, S. (2013b). A socio-technical comparison of rail traffic control between GB and Sweden. In *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction* (pp. 367-376).

Golightly, D., Wilson, J.R., Lowe, E. & Sharples, S. (2010) The role of Situation Awareness for understanding signalling and control in rail operations. Theoretical issues in Ergonomics Science, 11-12

Hassenzahl, M., & Wessler, R. (2000). Capturing design space from a user perspective: The repertory grid technique revisited. *International Journal of Human-Computer Interaction*, 12(3-4), 441-459.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press.

Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G., & Nielsen, M. (2009). Disruption management in passenger railway transportation. *Robust and Online Large-Scale Optimization*, 399-421.

Kauppi, A., Wikström, J., Sandblad, B., & Andersson, A. W. (2006). Future train traffic control: control by re-planning. Cognition, Technology & Work, 8(1), 50-56.

Koorey, G., McMillan, S., Nicholson, A. 2008. The effectiveness of incident management on network reliability. Land Transport NZ Research Report 346, Wellington, New Zealand. Avaialbe from http://ir.canterbury.ac.nz/handle/10092/2728/. [Last accessed: 19/09/15]

Kuru, A., & Erbuğ, Ç. (2013). Explorations of perceived qualities of on-body interactive products. *Ergonomics*, *56*(6), 906-921.

Lenior, D., Janssen, W., Neerincx, M., & Schreibers, K. (2006). Human-factors engineering for smart transport: Decision support for car drivers and train traffic controllers. *Applied Ergonomics*, 37(4), 479-490.

Lyons, G. 2006. The role of information in decision-making with regard to travel. In Intelligent Transport Systems, IEE Proceedings (Vol. 153, No. 3, pp. 199-212). IET.

Male, T., & Baber, C. 2014. Designing visual analytics for collaborative activity: a case study of rail incident handling. In *Contemporary Ergonomics and Human Factors 2014: Proceedings of the international conference on Ergonomics & Human Factors 2014, Southampton, UK, 7-10 April 2014* (p. 327). CRC Press.

Millen, L., Edwards, T., Golightly, D., Sharples, S., Wilson, J., & Kirwan, B. 2011. Systems change in transport control: applications of cognitive work analysis. *International Journal of Aviation Psychology*, 21 (1). 62-84

National Audit Office (NAO) 2008. Reducing Passenger Rail Delays by Better Management of Incidents: Report, Together with Formal Minutes, Oral and Written Evidence, Vol. 655, HMSO, London.

Pasquier, L., Brézillon, P., & Pomerol, J. C. 2000. From representation of operational knowledge to practical decision making in operations. *Decision Support through Knowledge Management*, 301-320.

Pender, B., Currie, G., Delbosc, A., & Shiwakoti, N. 2012. Planning for the unplanned: An international review of current approaches to service disruption management of railways. *In Australasian Transport Research Forum* (ATRF), 35th.

Pender, B., Currie, G., Delbosc, A., & Shiwakoti, N. 2014. Social media use during unplanned transit network disruptions: A review of literature. Transport Reviews, 34(4), 501-521.

Pickup, L., Wilson, J., & Lowe, E. 2010. The operational demand evaluation checklist (ODEC) of workload for railway signalling. *Applied ergonomics*, *41*(3), 393-402.

Riemersma, J. B. J. 1988. An empirical study of subjective road categorization. *Ergonomics*, *31*(4), 621-630.

Ryan, B. 2013. Reducing the risk of suicide or trespass on railways: developing better interventions through understanding behaviours of people. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *227*(6), 715-723.

Smith, W., & Dowell, J. 2000. A case study of co-ordinative decision-making in disaster management. *Ergonomics*, *43*(8), 1153-1166.

Steenhuisen, B., Dicke, W., & de Bruijn, H. 2009. “Soft” Public Values in Jeopardy: Reflecting on the Institutionally Fragmented Situation in Utility Sectors. Intl Journal of Public Administration, 32(6), 491-507.

Thomas, L. J., Rhind, D. J., & Robinson, K. J. 2006. Rail passenger perceptions of risk and safety and priorities for improvement. Cognition, Technology & Work, 8(1), 67-

Wilson, J. R., & Norris, B. J. 2006. Human factors in support of a successful railway: a review. *Cognition, Technology & Work*, *8*(1), 4-14.

Tables

Table 1 – Incident types used as elements for the repertory grid technique

Incident type	Description
Fatality	Fatality, used in this study, refers to death of any member of rail staff, passenger or public (for example through trespass or suicide).
Signalling / signalling systems failure	Refers to failure of signalling equipment used to grant movement authority to rail traffic.
Point failure	Failure of equipment for track points and switches
Track circuit failure	Failure for equipment for track circuits (equipment that detects train within a track section, triggering interlocking).
OHL (overhead line) dewirement	Refers to damage to overhead catenary wire supplying power to traction. As well as requiring repair, the unavailability of overhead power limits the type of traction that can enter the affected area, and makes it difficult to extract trapped trains (Golightly et al., 2013a).
Station overrun	The event in which a train which the driver is attempting to bring to a stand at a booked station stop proceeds beyond the designated stopping point such that any door intended to be available for public use at that station is no longer on the level platform.
Passenger loading	Passenger loading refers to problems associated with passenger flow, unexpected crowd (e.g. football match, etc.), or extended access / egress times due to poor platform adhesion for passengers (e.g. during snow).
Power failure	Refers to loss of power supply to signalling, stations, depots, and to third rail (DC) powered traction.
Speed restriction	Any problem that leads to imposing speed alteration to trains, these problems can be caused by weather, such as high winds, or temporary changes to the infrastructure.
Freight adhesion issues	Delay to a freight service due to a selection of issues such as leaf fall, length and weight of freight.

Table 2 – Unique constructs, with contrasting poles, counts and an example of a triad where the item in italics is selected as different from the other two

	Construct	Poles	Count (participants)	Count (total)	Example
1	Number of people involved	Single person; many people	6	17	*Fatality*, signal failure and track circuit failure. Fatality requires several groups from within the railway and external organisations to communicate with each other to manage this incident.
2	Impact	Low impact; high impact	6	17	*Freight adhesion issues*, track circuit failure and signal failure. Freight adhesion leads to lower impact to the service.
3	Distribution	Isolated; geographically dispersed	5	7	Fatality, signal failure and *freight adhesion*. Freight adhesion is often a result of weather which means the effects cannot be localised.
4	Effect on service	Stops the service; slows the service	5	11	*Fatality*, speed restriction issues and signal failure. Fatality leads to a complete blockage. The others may not.
5	Importance of human behaviour	Human behaviour important; human behaviour not important	5	10	Station overrun, fatality and *track circuit failure*. Track circuit failure is the different because it is not caused by human behaviour.

	Construct	Poles	Count (participants)	Count (total)	Example
6	Requires investigation	No investigation is required; further investigation required	4	4	Fatality, OHL and *track circuit failure*. Track circuit failure requires investigation to understand the problem.
7	Weather related	Weather not important; weather important	4	5	Station overrun, freight adhesion issues and *signal failure*. Signal failure is not usually caused by weather.
8	Length of incident	Short-term; long-term	3	5	*OHL*, track circuit failure and point failure. OHL takes a lot longer to handle.
9	Delays attributed	Short delays; long delays	3	10	Passenger loading, *OHL* and point failure. OHL most likely leads to long delays.
10	Importance of timetable knowledge	Timetable knowledge is not necessary; is necessary	3	6	Passenger loading, *signal failure* and freight adhesion issues. Signal failure usually requires good knowledge of timetable in order to regulate and re-route other trains around it.
11	Visibility of features	Not visible; very visible	3	4	*Station overrun*, point failure and signal failure. Station overrun is different as most of the features of the event are known without further investigation.
12	Importance of seasons	Not important; important	3	5	*Speed restriction issues*, point failure and passenger failure. A speed restriction issue tends to be seasonal.

	Construct	Poles	Count (participants)	Count (total)	Example
13	Variability of incident	Not variable; highly variable	3	7	Power failure, signal failure and *fatality*. Fatality is not as variable as the other two in terms of time its impact or time required to return to service.
14	Noticeability of incident	Easy to notice; difficult to notice	2	2	*Station overrun*, signal failure and power failure. Station overrun is usually difficult to notice by a signaller or control function. The other two have specific alarms.
15	Estimation of delay	Easy; difficult	2	3	*Station overrun*, signal failure and point failure. Station overrun is different as it is easy to estimate the delay associated with it once the cause is known.
16	Driver behaviour	Driver behaviour not important; important	2	2	Station overrun, freight adhesion and *signal failure*. Signal failure is different as driver behaviour is not important.
17	Diagnosis	Easy to diagnose; difficult	2	5	Signal failure, power failure and *freight adhesion issues*. Freight adhesion issues are different as it is easy to diagnose.
18	Mode of handling	Reactive; proactive	2	2	*Speed restriction issues*, OHL and fatality. Speed restriction issues can be handled proactively.
19	Location of handling	Requires access to site; can be handled remotely	2	5	*Signal failure*, station overrun and freight adhesion issues. Signal failure as it requires going to the site to manage it.

	Construct	Poles	Count (participants)	Count (total)	Example
20	Access to the site	Easy access; difficult	2	4	*Station overrun*, power failure and signal failure. Station overrun is easy to access and handle.
21	Technological aids	Technical aids not available; available	2	4	*Fatality*, track circuit failure and point failure. Fatality as there are no technological aids available alerting or supporting that type of incident.
22	Danger	Not very dangerous; very dangerous	1	1	*OHL*, point failure and track circuit failure. OHL can be potentially dangerous to passengers or staff (eg in terms of getting people off trains, access for staff).
23	Likelihood of occurrence	Not very likely to happen; very likely to happen	1	1	Points failure, track circuit failure and *freight adhesion issues*. A freight adhesion issue is not as likely as the other two incidents.
24	Social media	Social media is not important; important	1	2	Fatality, passenger loading and *signal failure*. Signal failure is different as social media is not a consideration.
25	Business reputation	Business reputation not important; very important	1	1	Signal failure, points failure and *fatality*. Fatality is different as it does not impact on the business reputation.
26	Time of day	Not important; very important	1	2	Signal failure, a track circuit failure and *freight adhesion issues*. Freight adhesion issue as the time of day is not important.

Table 3 – Mean ratings for element against constructs. Rows / constructs are ranked by overall mean ranking (highest at the top). Columns / elements are ranked by total mean rankings (highest on the left).

Construct (with poles)	**Fatality**	**OHL**	**Signal failure**	**Power failure**	**Points failure**	**Track circuit failure**	**Speed restriction issue**	**Freight adhesion issues**	**Station overrun**	**Passenger loading**	**Mean**
Business reputation (1 = not important; 9 = important)	7.25	7.21	7.32	7.18	7.04	6.25	5.5	4.93	6	6.25	6.49
Time of day (not important; important)	7.18	7.18	6.96	7.07	7	6.86	5.68	1.89	5.89	6.54	6.23
Variability of incident (not variable; very variable)	7.07	6.43	6.86	6.43	5.75	5.71	5.21	5.14	4.07	6	5.87
Importance of timetable knowledge (not important; important)	7.11	6.75	6.96	6.93	6.61	5.54	4.75	4.89	3.64	4.68	5.79
Likelihood (not likely; very likely)	5.96	4.61	5.64	5.07	6.54	6.39	5.96	5.64	5.21	6.79	5.78
Visibility of features(not visible; very visible)	6.68	5.25	6.82	6.5	6.32	6.68	4.5	4.5	4.25	3.82	5.53

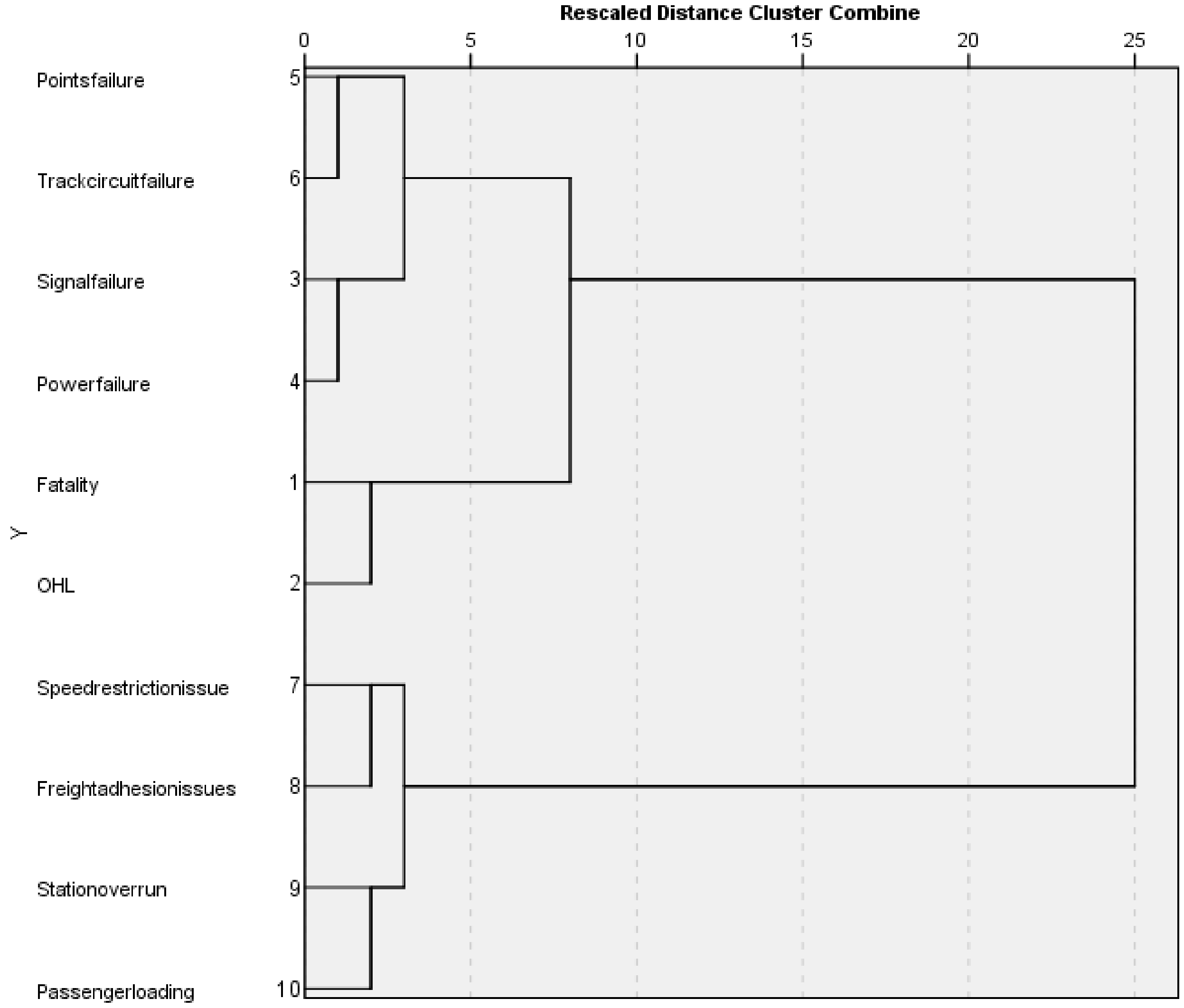
Weather(not important; important)	4.86	6.61	5.11	4.71	5.18	5.39	5.04	7.18	6.32	3.43	5.38
Human behaviour(not important; important)	8.04	5.07	4.29	4.25	4.14	4.04	4.07	5.29	6.68	7	5.29
Driver behaviour (not important; important)	7.04	5.57	4.46	3.21	3.86	4.11	5.29	5.89	7.89	4.71	5.2
Delay (short; long)	8.29	8.04	6.75	6.39	5.04	4.57	3.79	4.07	2.43	2.5	5.19
Impact (low impact; high impact)	7.96	7.71	6.61	6.46	5.11	4.93	3.71	3.68	2.39	2.68	5.13
Social media (not important; important)	6.64	5.71	5.54	5.29	5.18	4.89	4.25	3.68	4.25	5.36	5.08
Delay estimation (easy; difficult)	6.86	6.93	7.04	6.71	5.21	4.86	3.25	3.96	2.54	2.71	5.01
Effect on service (stops; slows)	1.68	1.64	3.39	3.46	4.71	5.64	7.39	6.82	7.32	7.75	4.98
Geographical distribution (localised; widespread)	5.68	5.54	5.57	5.57	4.86	4.5	4.89	4.71	4.07	3.79	4.92
Incident duration (short;	7.07	7.46	5.86	5.93	4.32	4.11	4.32	4.14	2.5	2.61	4.83

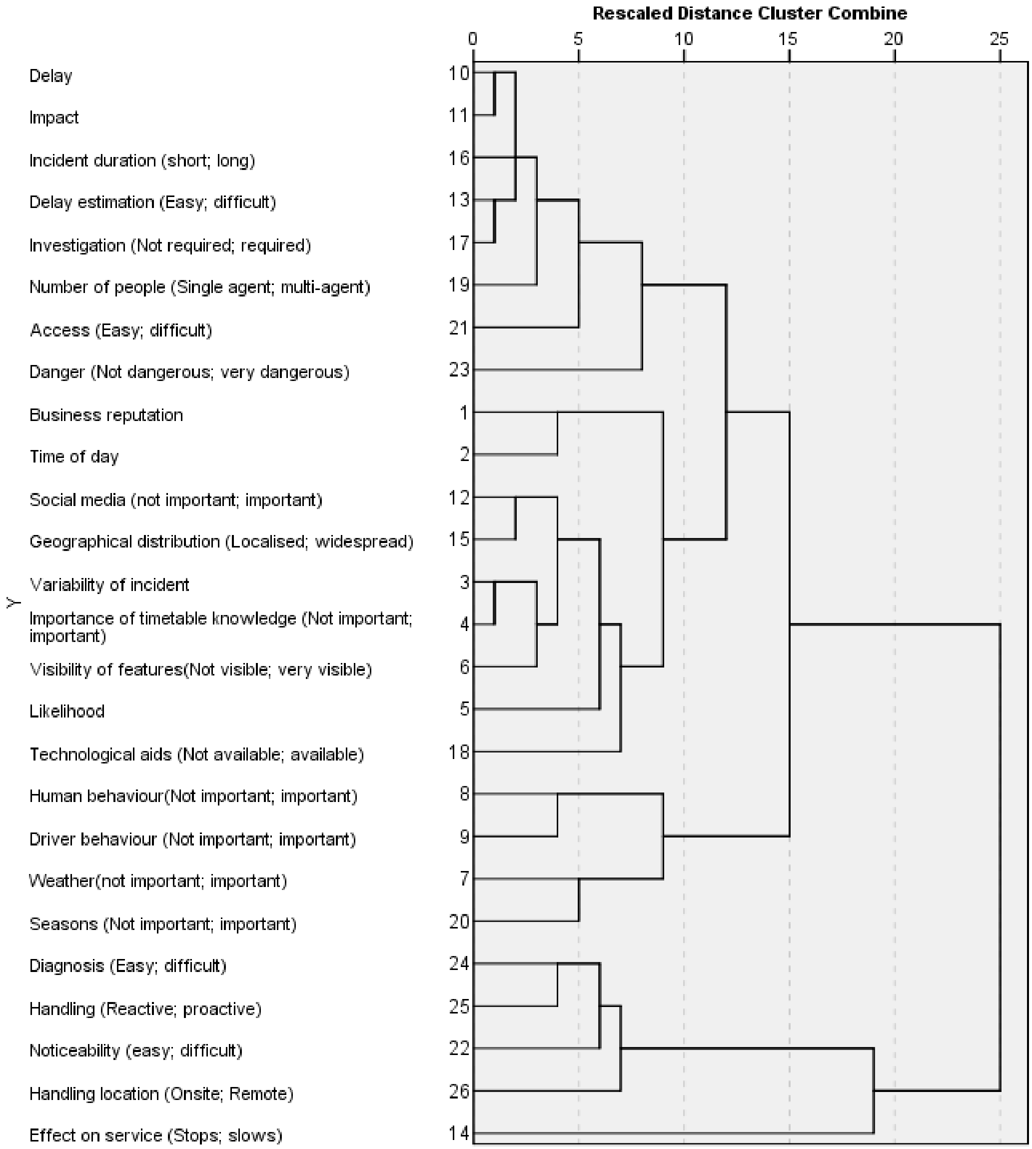
long)											
Investigation (not required; required)	6.18	7.21	6.54	6.11	4.75	4.82	3.18	3.32	3.57	2.29	4.8
Technological aids (not available; available)	3.25	4.18	5.39	5.21	5.5	5.32	5.07	4.5	3.93	4	4.64
Number of people (single agent; multi-agent)	7.46	7.18	5.32	5.04	4.25	4.18	3.18	2.79	2.29	2.46	4.41
Seasons (not important; important)	3.68	5.39	3.21	3.46	4.46	4.18	3.96	6.18	6.14	3.36	4.4
Access (easy; difficult)	5.71	5.25	4.68	4.61	4.39	4.75	4.04	4.11	2.46	2.25	4.23
Noticeability (easy; difficult)	4.14	3.71	3	3.5	2.89	3.54	5.18	5.36	5.5	5.32	4.21
Danger (not dangerous; very dangerous)	6.79	6.86	4.07	3.82	3.11	3	2.18	2.64	3.68	3.75	3.99
Diagnosis (easy; difficult)	3.75	3.54	4	4.68	3.36	3.64	3.96	3.21	2.64	3.64	3.64
Handling (reactive; proactive)	2.46	2.25	3.32	3.04	3.71	3.36	4.36	4.07	3.46	5.5	3.55
Handling location (onsite;	1.93	1.54	2.54	3.54	1.89	1.75	4.43	2.96	5.5	4.93	3.1

Remote)										
	5.8	5.57	5.28	5.16	4.81	4.73	4.51	4.45	4.41	4.39

Figures

ITEM	CONSTRUCTS	INCIDENTS									
		Fatality	OHL dewirement	Passenger loading	Points failure	Power failure	Signals/Signalling system	Station overrun	Track circuit failure	Speed restriction	Freight adhesion issues
1	The length of incident (from start of the failure to back to normal timetable) 1=Short term 9=Long term										
2	The number of people involved with the handling of the incident. 1=Single person 9=Many people										


Rescaled Distance Cluster Combine
0
5
10
15
20
25
Y
Pointsfailure 5
Trackcircuitfailure 6
Signalfailure 3
Powerfailure 4
Fatality 1
OHL 2
Speedrestrictionissue 7
Freightadhesionissues 8
Stationoverrun 9
Passengerloading 10


Rescaled Distance Cluster Combine
0
5
10
15
20
25
Y
Delay 10
Impact 11
Incident duration (short; long) 16
Delay estimation (Easy; difficult) 13
Investigation (Not required; required) 17
Number of people (Single agent; multi-agent) 19
Access (Easy; difficult) 21
Danger (Not dangerous; very dangerous) 23
Business reputation 1
Time of day 2
Social media (not important; important) 12
Geographical distribution (Localised; widespread) 15
Variability of incident 3
Importance of timetable knowledge (Not important; important) 4
Visibility of features(Not visible; very visible) 6
Likelihood 5
Technological aids (Not available; available) 18
Human behaviour(Not important; important) 8
Driver behaviour (Not important; important) 9
Weather(not important; important) 7
Seasons (Not important; important) 20
Diagnosis (Easy; difficult) 24
Handling (Reactive; proactive) 25
Noticeability (easy; difficult) 22
Handling location (Onsite; Remote) 26
Effect on service (Stops; slows) 14