Development, validity and reliability testing of the East Midlands Evaluation Tool (EMET) for measuring impacts on trainees’ confidence and competence following end of life care training

B Whittaker, 1 R Parry, 2 L Bird, 3 S Watson, 4 C Faull, 5 on behalf of the EMET project team

ABSTRACT
Objectives To develop, test and validate a versatile questionnaire, the East Midlands Evaluation Tool (EMET), for measuring effects of end of life care training events on trainees’ self-reported confidence and competence.

Methods A paper-based questionnaire was designed on the basis of the English Department of Health’s core competencies for end of life care, with sections for completion pretraining, immediately post-training and also for longer term follow-up. Preliminary versions were field tested at 55 training events delivered by 13 organisations to 1793 trainees working in diverse health and social care backgrounds. Iterative rounds of development aimed to maximise relevance to events and trainees. Internal consistency was assessed by calculating interitem correlations on questionnaire responses during field testing. Content validity was assessed via qualitative content analysis of (1) responses to questionnaires completed by field tester trainers and (2) field notes from a workshop with a separate cohort of experienced trainers. Test–retest reliability was assessed via repeat administration to a cohort of student nurses.

Results The EMET comprises 27 items with Likert-scaled responses supplemented with questions seeking free-text responses. It measures changes in self-assessed confidence and competence on 5 subscales: communication skills; assessment and care planning; symptom management; advance care planning; overarching values and knowledge. Test–retest reliability was found to be good, as was internal consistency: the questions successfully assess different aspects of the same underlying concept.

Conclusions The EMET provides a time-efficient, reliable and flexible means of evaluating effects of training on self-reported confidence and competence in the key elements of end of life care.

INTRODUCTION
Policies in the UK, 1, 2 and internationally, 3–9 strongly endorse staff training as a means to increase both specialist and non-specialist health and social care workers’ competence in end of life care delivery. In England in recent years, there has been a proliferation of training events that vary widely in length, modality, trainees and content. 8 These training events range from half a day to multiple days, and trainees span a broad spectrum of occupational and professional backgrounds, grades, settings and patient groups. Many events are multiprofessional.

The gold standard means of assessing the impact of training is to perform before and after workplace observations of staff and patients interacting. 6, 7 However, this can be highly time-consuming, while evaluation via self-report is far more feasible and economical. Therefore, even though it is known that self-reported confidence and competence do not straightforwardly reflect actual workplace behaviour change, 8, 9 self-report often represents the best available option when resources do not stretch to workplace observations.
A systematic review examined existing self-report tools relevant to assessing end of life care training, and found that most are poorly validated and narrow in scope, and that they largely focus on physical aspects of symptom management. Some tools are also narrowly focused on single professional groupings. This means that current tools have limited usefulness in the current end of life care training environment where, as we noted, end of life care training events are very diverse. Furthermore, evaluations of effects of end of life care education and training interventions published over the past decade all report designing and using individual, project-specific evaluation questionnaires. This adds to the evidence that we lack, and need, established, validated, and broadly relevant tools.

Our questionnaire, referred to as a ‘tool’, is designed to offer a rapid and feasible means to evaluate end of life care training events by measuring changes in trainees’ self-reported confidence and competence. By confidence, we mean the self-awareness of having the competence to complete a task or reach a goal. By competence, we mean having the appropriate skills and behaviours to undertake specific activities. Fortunately, a clear and government-endorsed articulation of core competencies for end of life care exists and (see table 1).

These competencies were developed with the aim of providing a sound framework for the commissioning and design of training programmes recommended in English end of life care policy. They also provide a useful basis for evaluating events.

METHODS

Tool development

The East Midlands Evaluation Tool (known hereafter as EMET) was initially designed by authors BW and CF and project team member Debra Broadhurst on the basis of the English core competences for end of life care (see table 1). It was then refined and developed by the EMET project team (see Acknowledgements) over a year period through five iterative rounds of development. While in practice, all sections of the tool are designed to be used in non-anonymised form; during testing, all completed questionnaires were anonymised by ensuring that each trainee only wrote their unique identifying code on the questionnaire.

The initial design of the tool entailed:

- A literature search and review which found that there were no validated tools available for evaluating the impact of end of life care learning events across a range of roles and care settings.
- Discussions within a project team that included clinical and educational experts in end of life care.
- Translation of the overarching statements of competencies in the Department of Health’s framework into questionnaire items grouped into the same subdomains. This entailed translating broad characteristics into more specific skills, practices or activities. For example, the statement of communication competence: ‘Develop and maintain communication with people about difficult and complex matters or situations related to end of life care’ was reformulated to: ‘I feel confident to talk with a dying person about issues surrounding their death’.

The resulting tool comprised a total of 27 statements to which trainees were asked to respond via Likert-scaled responses. This was supplemented by (1) narrative questions seeking free-text responses to gain trainee views on whether, and how, the training had changed their confidence and competence in the delivery of end of life care.

The next stage of development involved field testing the tool by administering it across a wide range of training events delivered by a total of 13 different organisations. We used the East Midlands Strategic Health Authority and personal networks of the project team to recruit trainers willing to use the tool to evaluate their training events. Thirteen organisations used the tool in 55 different training events involving 1793 trainees. Almost three-quarters (71%) of the training events involved mixed cohorts of registered and non-registered employees from various occupational backgrounds and care settings. Events and trainees are described in table 2.

Trainees returned the questionnaires completed by each trainee to the EMET project team. Additionally, all trainers administering the tool during field testing were asked to complete a brief ‘feedback questionnaire’ which sought both Likert-scaled and free-text responses on the ease of use of the tool, whether it reflected the aims of their own event(s), and whether they thought any items needed adding, altering or removing. This questionnaire is provided in online supplementary file 1.
The needs and requests of trainers and organisations delivering the courses.

New policy initiatives—in particular the UK’s National Institute on Health and Care Excellence (NICE) guidance focused on recognising dying, avoiding inappropriate hospital admissions and initiating conversations about end of life.\(^\text{23}\)

The modifications entailed rewording some questionnaire statements and narrative questions so as to improve their applicability across a wide range of trainees and events. For instance, the initial version of the EMET referred to the care of patients; this was changed so as to replace reference to ‘patients’ with reference to ‘people in my care’. Other modifications made questionnaire items better reflect the competencies being measured by the Likert-scaled items and reported in the narrative questions. For example, some trainees noted the importance of listening to patients, and as a result the initial wording of one item: ‘I feel confident to talk with a dying patient about issues surrounding their death’ was modified to: ‘I feel confident to listen to and talk with a dying patient about issues surrounding their death’.

Table 2 Training events and trainees using the tool

<table>
<thead>
<tr>
<th>Delivering organisation testing the tool (n=13)</th>
<th>Length of training event</th>
<th>Occupation of participants completing the tool</th>
<th>Number of trainees completing the tool (n=1793)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher education institution (n=3)</td>
<td>3 days to 6 weeks</td>
<td>Registered nurses</td>
<td>53</td>
</tr>
<tr>
<td>Community NHS trust (n=2)</td>
<td>1 day</td>
<td>Registered nurses, non-registered carers</td>
<td>201</td>
</tr>
<tr>
<td>NHS regional ambulance service (n=1)</td>
<td>1</td>
<td>Ambulance practitioners</td>
<td>87</td>
</tr>
<tr>
<td>Acute NHS trust (n=1)</td>
<td>1 day</td>
<td>Registered nurses (induction programme)</td>
<td>147</td>
</tr>
<tr>
<td>Voluntary education provider for healthcare professionals (n=1)</td>
<td>1-2 days</td>
<td>Registered nurses, non-registered carers, doctors</td>
<td>81</td>
</tr>
<tr>
<td>Training managers</td>
<td>1</td>
<td>Care home managers, student nurses, end of life care</td>
<td>45, 17, 36</td>
</tr>
<tr>
<td>Chaplains</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| GP, general practitioner; NHS, National Health Service.

Reliability

Two domains were tested: test–retest reliability and internal consistency.

Test–retest reliability

Reliable tools produce the same score or measurement each time they are used if there has been no change to the features being measured.\(^\text{24 25}\) Thus, if the trainee’s self-assessed confidence and competence remain constant, there should be no change in their score on the EMET. Conversely, any change in their score should be due to a change in self-assessed confidence and competence. We examined the degree to which the tool yielded the same results from one administration to the next under the same conditions using a convenience sample of student nurses from a single year group in the Division of Nursing at the University of Nottingham. The students completed the EMET’s 27 Likert-scaled items on two occasions 1 week apart. During that week, they received no end of life care specific teaching and were not exposed to clinical situations in which end of life care would have been observed. The correlation between the two scores was tested using a power calculation of 15% deviation. Scores were tested using a Pearson r test.\(^\text{26}\)

Internal consistency

A reliable multi-item tool will give consistent results when different aspects of the same underlying concept are measured by more than one item.\(^\text{27}\) During field testing, we examined the internal consistency of results across different multi-item sets of questions. Interitem correlations were calculated for each of the five core competence subscales using Cronbach’s α values.\(^\text{26}\)

Validity

Validity refers to whether a tool is measuring what it is designed and claimed to measure. We examined content validity: whether the domains examined by the tool were appropriate, important and sufficient to its purposes.\(^\text{25}\) Content validity is examined through consultation with interested parties rather than via statistical methods. Accordingly, we conducted two separate structured consultations. One of these entailed completion and analysis of trainer feedback questionnaires during field testing as described above (see online supplementary file 1 and table 2). We collated frequencies for the responses to the Likert-scaled statements and analysed responses to the narrative questions via inductive qualitative content analysis.\(^\text{28}\)

Second, we designed and ran a three-part structured workshop during the 2012 conference of the UK National Association of Palliative Care Educators (NAPCE). The workshop began with experienced end of life care trainers testing the tool by completing its pretraining section. Next, they participated in facilitated small group discussions designed to consider positive and negative aspects of the tool. Finally, a...
large group discussion brought together the views of
the small groups. Contemporaneous field notes were
made by authors BW and SW. These were collated and
analysed via inductive qualitative content analysis.28
The process of tool development over a 4-year period
is illustrated in figure 1.

RESULTS
Tool development
The resulting questionnaire ‘tool’ is available as online
supplementary file 2. It comprises three sections:

Pretraining section ‘tool A’
This is completed immediately before the training
begins—usually at the event venue itself, but it can be
sent to each trainee before the event. It gathers brief
background information on the trainee’s work setting,
professional qualifications (if any) and the date and
type of the training event. It then asks the trainee to
rate their confidence and competence across the five
domains of the end of life care competences via the
27 Likert-scaled responses. Additionally, two narrative
questions seek free-text responses from the trainee on
their reasons for attendance and expectations of the
event.

End of training section ‘tool B’
Administered at the end of the training event, usually
at the event venue itself, the trainee rates their confi-
dence and competence via the same 27 statements.
They do so without sight of their pretraining responses. Additionally, narrative questions ask
whether the training has met their expectations, and
ask the trainee to articulate specific actions they plan
to undertake as a consequence of the training. This
question reflects emerging evidence on the value of
action planning (or goal setting) within educational
and behavioural change interventions.29 30

Optional follow-up section ‘tool C’
Designed for postal administration weeks or months
after training, this section repeats the 27 Likert state-
ments and then poses three new narrative questions
which ask the trainee to report any impacts of training
in relation to: (1) recognising dying; (2) avoiding
inappropriate hospital admissions and (3) initiating
conversations about end of life. These questions were
framed in relation to English national end of life care
targets31 and quality standards.32

Reliability
Test–retest reliability
The convenience sample of 112 student nurses com-
pleted the Likert section of the questionnaire on two
separate occasions 1 week apart. The overall total
score at each time point correlated best, with a
Pearson r correlation of 0.84. A score of 1 indicates
a perfect correlation but scores of 0.7 or above are
generally considered to be highly correlated.31 Four of
the five subscales also correlated highly. The score
that correlated least well was ‘overarching values and
knowledge’ (Pearson’s r of 0.56), indicative of a
moderate-to-good correlation. These results in table 3
suggest that the tool has good test–retest reliability
and that changes in score over time can be attributed
to changes in self-assessed confidence and competence
as opposed to the effect of repeat administration.

Table 4 provides the pretraining and post-training
scores for 1793 trainees who completed the assess-
ment tool. In contrast to the test–retest scores which
showed minimal change, these scores showed that trai-
nees’ overall self-perceived confidence and competence
had increased on average by ~13 points
(possible scores range from 2).

Internal consistency
The internal consistency of responses was calculated
for a total of 1793 questionnaires; all components had
acceptable (>0.7) Cronbach’s α values.27 Taking
into account that the subscales comprised a relatively
low number of items, this indicates good reliability.
These results indicate that within the five key com-
ponents the questions successfully assess different aspects of the same underlying concept (see table 5).
Furthermore, the absence of extremely high correla-
tions indicates that there were no redundant ques-
tions, that is, items that were so similar that they
simply asked the same question in marginally different
ways.

Validity
The first structured consultation was via a trainer
feedback questionnaire completed during fieldwork.
This feedback yielded 23 completed responses from 16
trainers representing 10 organisations relating to
23 different training events (separate questionnaires
were completed for each event). Trainer responses to
the Likert-scaled statements are shown in table 6.
Their narrative responses indicated that overall trai-
ners liked the format of EMET, that trainees were able
to complete it, and that it fitted their training’s
content and learning outcomes.

The trainers’ narrative comments provided insights
on the EMET’s positive and limiting features. Four
trainers commented that it was time-consuming for
students to complete, and one reported that some trai-
nees rushed through the post-training section, particu-
larly at the end of the day. One trainer thought the
tool did not match the specific clinical content of their training event. One commented that trainees
could not remember their precourse expectations and
thus could not accurately report on whether expecta-
tions were met. On the other hand, three trainers
commented that the specific focus on end of life care
was particularly useful, and that trainees’ completion
of the tool provided them with particularly valuable
feedback for future training events.

4
Whittaker B, et al. BMJ Supportive & Palliative Care 2017;0:1–8. doi:10.1136/bmjspcare-2016-001100
The second structured consultation took the form of a three-part workshop at the UK NAPCE conference. In total, 22 experienced end of life care trainers participated in the workshop. Group feedback and discussions were collated and the findings are summarised in [Box 1](#).

Overall, a large majority of the 38 trainers consulted during field testing and at the conference workshop considered that the EMET measures what it sets out to measure and is relevant to a range of training events and trainees. Most also remarked that it was practical, flexible to different training events across a range of professional groups and easily administered.

A large majority of trainers reporting on their experiences of using the EMET in 23 training events indicated that they liked the format, that trainees were able to easily complete it, and that it was appropriate for the content of their training. A separate cohort of 22 experienced end of life care trainers consulted via a workshop appreciated its comprehensiveness, its congruence with a nationally defined competence framework, and its capacity to collect both brief responses on a wide range of end of life care competencies and longer free-text responses which could be used to modify and develop future training.

**DISCUSSION**

We developed the EMET to evaluate diverse end of life care training events in England after establishing that there were no existing validated tools that were suitable for the wide range of events and trainees who participate in them. The validated tool provides a comprehensive approach for evaluating training in terms of changes in self-reported confidence and competence across five core areas, reflecting the domains of the English core competencies for end of life care. The tool is suitable for use with a wide range of trainees across a spectrum of end of life care training events. Our small-scale evaluation of its test-retest reliability and a larger scale evaluation of its internal consistency, usability and validity yielded positive results.

The experienced trainers who participated in validity testing also commented on some drawbacks and limitations of the tool. Their responses suggest that the EMET may be too lengthy for administration in training events lasting half a day or less. They raised concerns about possible ceiling effects, that is, that it will not measure changes in trainees who report high levels of confidence and competence pre-training. We acknowledge that anecdotally we know

<table>
<thead>
<tr>
<th>Construct</th>
<th>Mean average at time 1</th>
<th>Mean average at time 2</th>
<th>Correlation value</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication skills</td>
<td>20.920 (SD 3.448)</td>
<td>21.667 (SD 3.084)</td>
<td>0.759</td>
<td>0.000</td>
</tr>
<tr>
<td>Assessment and care planning</td>
<td>21.756 (SD 4.206)</td>
<td>22.826 (SD 3.918)</td>
<td>0.724</td>
<td>0.000</td>
</tr>
<tr>
<td>Symptom management</td>
<td>16.449 (SD 3.265)</td>
<td>17.320 (SD 3.303)</td>
<td>0.705</td>
<td>0.000</td>
</tr>
<tr>
<td>ACP</td>
<td>13.949 (SD 2.701)</td>
<td>14.455 (SD 2.460)</td>
<td>0.669</td>
<td>0.000</td>
</tr>
<tr>
<td>Overarching values and knowledge</td>
<td>19.255 (SD 3.192)</td>
<td>19.310 (SD 3.145)</td>
<td>0.569</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall totals</td>
<td>92.081 (SD 12.497)</td>
<td>94.911 (SD 12.401)</td>
<td>0.840</td>
<td>0.000</td>
</tr>
</tbody>
</table>

ACP, advance care planning.
that trainers varied in how effectively they were able to ensure how delegates completed the tool. Trainers also raised concerns about the validity of the trainees’ responses. This latter concern is congruent with numerous empirical studies that have shown that self-report does not straightforwardly reflect actual skill. We know that in general, self-reports yield larger change scores than evaluation of actual performance. Indeed, it has long been recognised that knowledge about clinical matters is much more easy to assess than actual application of that knowledge to workplace performance. Therefore, where resources allow, multiple assessments including actual workplace performance and patient outcomes should be used. These will potentially yield a more accurate understanding of training’s impacts than can be gleaned from self-reports. However, in situations where resources are limited, assessment of the impacts of training courses on trainees via their self-report, such as provided by the EMET, offers a feasible and economical means of measuring a limited aspect of training’s impacts. Trainers, trainees, managers and commissioners should nevertheless be aware that self-reports provide very limited insights into actual workplace behaviour change. The validated tool provides a baseline in recognising changes in confidence and competence as a starting point to be able to identify impacts of training on clinical practice. Overall, the validated tool fills a recognised gap currently evident in evaluating end of life care training events.

Final V6 was produced on the basis of reliability and validity testing. This version of EMET is freely available and can be downloaded from the Nottingham Centre for the Advancement of Supportive, Palliative and End of Life Care, University of Nottingham (NCARE) http://nottingham.ac.uk/research/groups/src/postgrad-course.aspx. It is currently in use in a range of events internationally in the context of a continued drive to improve end of life care delivery. Our work in developing and validating EMET adds to the understanding of how training can impact on workforce ability to meet patient outcomes in end of life care.

**Strengths and limitations**

We designed an acceptable and useable tool, conducting preliminary testing of the EMET’s validity and reliability. Limitations of the test–retest reliability include the comparatively small and homogeneous sample and the brief time period between administrations. While this reduced the likelihood of confounding factors influencing the second test, it remains conceivable that there was insufficient time for any effects of completing the first test to have dissipated. We acknowledge the comparatively low reliability score for overarching values and knowledge and suggest that this dimension might be particularly sensitive to repeat administration. Further exploration of the factors affecting the overarching values and knowledge score would be valuable. Strengths of this study include the comparatively large data set that was used to test internal consistency and the breadth and range of individuals who contributed to the assessments of validity over a period of time.

**Table 4** Average pretraining and post-training scores for respondents who had completed end of life care training (n=1793)

<table>
<thead>
<tr>
<th>Construct</th>
<th>Pretraining mean average</th>
<th>Post-training mean average</th>
<th>Mean change in score</th>
<th>p Value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication skills</td>
<td>21.448 (SD 4.161)</td>
<td>25.059 (SD 4.343)</td>
<td>3.602 (SD 3.224)</td>
<td>0.000</td>
</tr>
<tr>
<td>Assessment and care planning</td>
<td>14.816 (SD 5.114)</td>
<td>17.410 (SD 5.414)</td>
<td>2.612 (SD 3.224)</td>
<td>0.000</td>
</tr>
<tr>
<td>Symptom management</td>
<td>16.192 (SD 4.199)</td>
<td>19.237 (SD 4.440)</td>
<td>3.077 (SD 3.395)</td>
<td>0.000</td>
</tr>
<tr>
<td>ACP</td>
<td>14.935 (SD 3.017)</td>
<td>17.111 (SD 3.161)</td>
<td>2.160 (SD 2.754)</td>
<td>0.000</td>
</tr>
<tr>
<td>Overarching values and knowledge</td>
<td>19.751 (SD 3.611)</td>
<td>21.483 (SD 3.952)</td>
<td>1.749 (SD 2.996)</td>
<td>0.000</td>
</tr>
<tr>
<td>Overall totals</td>
<td>91.695 (SD 20.894)</td>
<td>104.915 (SD 24.781)</td>
<td>13.220 (17.547)</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Wilcoxon signed rank test.
ACP, advance care planning.

**Table 5** Reliability test results: internal consistency via Cronbach’s α values using Pearson’s r

<table>
<thead>
<tr>
<th>Component</th>
<th>Cronbach’s α</th>
<th>Number of Likert scale items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications skills</td>
<td>0.845</td>
<td>6</td>
</tr>
<tr>
<td>Assessment and care planning</td>
<td>0.863</td>
<td>7</td>
</tr>
<tr>
<td>Symptom management</td>
<td>0.863</td>
<td>5</td>
</tr>
<tr>
<td>Advance care planning</td>
<td>0.823</td>
<td>4</td>
</tr>
<tr>
<td>Values and knowledge</td>
<td>0.819</td>
<td>5</td>
</tr>
<tr>
<td>27 items</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CONCLUSION

We advocate use of this freely available validated evaluation tool of self-reported confidence and competence (see online supplementary file 2), to assess impacts of end of life care training and to gather feedback on training events. Where feasible, additional observational assessment of performance will provide more direct evaluation of training’s impact on practice and quality of service provision within the context of end of life.

Ethics

None required. All trainees and facilitators were informed that data would be used as part of the tool development and testing.

Acknowledgements

The EMET project was initially funded by the East Midlands Strategic Health Authority and the National End of Life Care Programme, UK up to 2013. The EMET project team comprised: Debra Broadhurst, LOROS Hospice Leicester, early questionnaire design; Professor Jane Seymour, Sue Ryder Care Centre for the Study of Supportive, Palliative and End of Life Care, project support; Gavin Narayanasamy, Melanie Narayanasamy and Sarah Todd, data input and analysis; Dr Anthony Arthur, early advice on statistics methods.

Contributors

BW, SW and CF were responsible for project planning. BW and SW were responsible for conduct of the study. BW, SW, LB, CF and RP contributed to the reporting and discussion. BW has overall responsibility for the content.

Competing interests

None declared.

Provenance and peer review

Not commissioned; externally peer reviewed.

REFERENCES

1 Leadership Alliance for Care of Dying People. One chance to get it right: improving people’s experience of care in the last few days and hours of life. London: NHS England, 2014.

The EMET was designed in the context of policy and practice in England, and while its comprehensive coverage relates to end of life care delivery nationally, we suggest that it could be usefully applied in other countries. Such application should ideally be accompanied by testing of reliability and validity across different national and cultural contexts.


