A systematic review of 7 years of research on entrustable professional activities in graduate medical education, 2011–2018

Emily O'Dowd,[1,2] Sinéad Lydon,[1,3] Paul O'Connor,[1,2] Caoimhe Madden[1,2] & Dara Byrne[1,3]

PURPOSE This review aimed to synthesise some of the extant work on the use of entrustable professional activities (EPAs) for postgraduate physicians, to assess the quality of the work and provide direction for future research and practice.

METHOD Systematic searches were conducted within five electronic databases (Medline, Scopus, Web of Science, PsycINFO and CINAHL) in September 2018. Reference lists, Google Scholar and Google were also searched. Methodological quality was assessed using the Quality Assessment Tool for Studies with Diverse Designs (QATSDD).

RESULTS In total, 49 studies were included, classified as *Development of EPAs* (n = 37; 76% of total included), *Implementation and/or assessment of EPAs* (n = 10; 20%), or both (n = 2; 4%). EPAs were described for numerous specialties, including internal medicine (n = 14; 36%), paediatrics (n = 8; 21%) and psychiatry (n = 4; 10%). Of the development studies, 92% utilised more than one method to generate EPAs. The two most commonly used methods were developing initial EPAs in a working group, (n = 27; 69%) and revising through deliberation (n = 21; 54%). Development papers were of variable quality (mean QATSDD score = 20, range 6–41). Implementation and assessment studies utilised methods that included observing trainee performance (n = 6; 50%) and enrolling trainees in competency-based curricula, which included EPAs (n = 4; 33%). The methodological quality of these implementation studies varied (mean QATSDD score = 19.5, range = 6–32).

CONCLUSIONS This review highlighted a need for: (i) consideration of best practice guidelines for EPA development; (ii) focus on the methodological quality of research on EPA development and of EPAs, and (iii) further work investigating the implementation of EPAs in the curriculum.

Medical Education 2019
doi: 10.1111/medu.13792

[1]Irish Centre for Applied Patient Safety and Simulation, National University of Ireland, Galway, Ireland
[2]Discipline of General Practice, School of Medicine, National University of Ireland, Galway, Ireland
[3]School of Medicine, National University of Ireland, Galway, Ireland

Correspondence: Paul O'Connor, Discipline of General Practice, 1 Distillery Road, National University of Ireland Galway, Co. Galway, H91 TK33 Ireland. Tel: 00 353 91 492 897; E-mail: paul.oconnor@nuigalway.ie

INTRODUCTION

The abilities of newly graduated trainee physicians upon entering practice often fall below the expectations of senior colleagues.[1–3] This disparity between the expectations and realities of physician competence is regarded as a patient safety issue[4] and has led to increased interest in competency-based medical education (CBME). CBME is concerned with outcomes of the education experience, independent of time spent in education.[5] Although CBME has been generally well received,[6] there remains debate surrounding its use. One alternative to traditional CBME may be the use of entrustable professional activities (EPAs).

Entrustable professional activities are an effort to bridge the gap between the theory of competencies and practical clinical work.[7] An EPA is an essential unit of work incorporating one or more core competencies, is observable, and can be entrusted to trainees at different levels, ranging from not being entrusted to being entrusted to supervise others in the activity.[8] The entrustment of an EPA applies only to that activity, in that context, and eliminates uncertainty on the part of the learner and the supervisor regarding supervision requirements.[9] The EPA framework allows supervisors to make evidence-based decisions regarding the ability of a trainee to conduct an activity safely and competently, and simplifies CBME by integrating competencies into the assessment of EPAs.[10] Moving from a focus on the 'person-descriptors' of competencies to the 'work-descriptors' of EPAs enables a more systematic implementation of CBME.[10] EPAs formalise a framework of entrustment and in turn impact patient safety.[11,12]

Entrustable professional activities are becoming increasingly common,[13–15] with the Netherlands, Canada and Australia among countries adopting EPA frameworks in graduate medical education (GME). However, as of yet, there have been few efforts to examine the current state of research on the development and implementation of GME EPAs internationally. Pilot studies in undergraduate medical education (UME) are ongoing.[14] However, there remains a need to consider the research evidence and data surrounding GME EPAs.[16]

The aim of the current systematic review was to synthesise some of the extant research and knowledge on the development and use of EPAs in GME, to assess the methodological quality of the identified studies, and to provide direction for future research and work on EPAs. Conducting a systematic review on EPAs in GME at this point will be of value in guiding future research and for those in clinical practice developing or using EPAs.

METHOD

Design

A systematic review was conducted. Systematic reviews, which collate data from multiple studies in a research area, are advocated as an efficient way of keeping up to date with literature, particularly in health research,[17] a more objective method of reviewing and synthesising evidence than other review types,[18] and as the 'best evidence' in the hierarchy of evidence.[19] As per best practice for systematic reviews, this review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines.[20]

Search strategy

We conducted searches using five electronic databases: Medline, CINAHL, Scopus, PsycINFO and Web of Science. Searches were conducted in September 2018. 'Entrustable professional activit* OR Entrustability' was searched across databases and altered as necessary to meet the entry formatting required (for further detail, see Appendix S1). To increase the likelihood of identifying grey literature, Google and Google Scholar were also searched. The terms 'entrustable professional activities' and 'entrustability' were entered and the first 1000 returns were screened for each of these search engines. Finally, the reference lists of all included studies were screened.

Study selection

Inclusion criteria

To meet our inclusion criteria, papers or documents were required to explicitly report a focus on 'entrustable professional activities'. Beyond this, papers had to: (i) be written in English; (ii) describe the development or expansion of one or more EPAs, or report on the implementation outcomes of one or more EPAs, and (iii) report specifically on EPAs for postgraduate trainees (i.e.

physicians who have completed their undergraduate medical degree and are pursuing further training). Both peer-reviewed and grey literature documents were eligible for inclusion.

Exclusion criteria

We excluded papers if the EPAs were intended for use with undergraduates or for the evaluation of newly graduated students entering GME. Similarly, papers describing EPAs for other professions were excluded (e.g. physician assistants, nurses and pharmacists). Other reasons for exclusion were as follows: (i) a focus on related concepts within GME, including competencies, capabilities or milestones, but not on EPAs; (ii) the availability of an abstract only; (iii) an absence of original data regarding the development or implementation of, or trainee assessment using, EPAs, and (iv) the use of EPAs by trainees to assess their own capabilities or the application of EPAs in assessing hypothetical trainees.

Screening process

Titles and abstracts returned during the search process were screened by one author (EO'D), and papers that did not meet the criteria for inclusion were excluded. If these provided insufficient information to make a decision, the full text was accessed and a decision regarding inclusion was made. If any uncertainties arose at this point, a decision was made in consultation with another author (SL).

Categorisation of studies

Included papers were assigned to one of two categories by one author (EO'D). The first category was concerned with the *development of EPAs*, and studies that reported on the development or expansion of one or more EPAs for postgraduate physicians were assigned to this category. The second category was related to the *implementation of and/or assessment of trainees using EPAs* and comprised studies that reported on how EPAs were implemented within clinical settings and/or used in practice to assess trainee physicians. In some cases, it was deemed appropriate to assign studies to both of these categories.

Data extraction

There were minor differences in the data extraction process for the categories. For the *development of EPAs* category, data were extracted on year of publication, country of study, medical specialty, number of EPAs developed, content of EPAs, grade of physician, method of development of EPAs and methodological quality. As studies provided great detail regarding their method of development, it was necessary to code the methods to facilitate synthesis. Two researchers developed codes to capture development methods employed across these studies. This was an iterative process, reducing variation to produce a discrete list of codes (for more information, see Appendix S2), at the same time retaining all information provided by studies. For example, 'committees', 'working groups' and 'panels' were all considered sufficiently similar to be coded as 'working groups'.

For the *implementation and/or assessment of EPAs* category, data were extracted on year of publication, country of study, method of implementation and assessment, medical specialty, number of EPAs, number and grade of trainees included, outcomes of implementation and assessment, and methodological quality. The methods of implementation and assessment described within the papers were coded as above.

Data extraction and coding were carried out independently by two researchers ((EO'D and CM). Agreement between these researchers was found to be 98% across this process, with any disagreements resolved through discussion between the two researchers until consensus was achieved.

Quality assessment and data synthesis

The methodological quality of included papers was assessed using the Quality Assessment Tool for Studies with Diverse Designs (QATSDD).[21] The QATSDD is intended for use in the assessment of the methodological quality of studies with varying research designs. It is a 16-item tool developed for use by health services researchers. Items are rated on a 4-point scale (0–3), with a higher score indicating greater methodological rigour. Scores on the QATSDD can range from 0 to 42 (qualitative and quantitative studies) or 48 (mixed-methods studies). This assessment tool has been shown to produce good agreement[21] and has been used in a number of different reviews pertaining to health services and medical education research.[22–24] The QATSDD was applied to studies within both categories by two researchers independently and any disagreements were resolved through discussion.

RESULTS

A total of 2894 records were screened, with 49 articles[25–73] deemed eligible for inclusion (for PRISMA diagram, see Fig. 1). Of these, 37 papers (76%) were assigned to the *development of EPAs* category alone and 10 (20%) were assigned to the *implementation and/or assessment of EPAs* category alone. Two studies (4%)[25,46] were assigned to both categories, as they each described both the development and assessment of at least one EPA. The number of studies published in the area increased year upon year (for a graph presenting this, see Fig. 2).

Papers describing the development of EPAs

Key characteristics of these papers are summarised in Table 1 (for detail on individual studies, see Appendix S3).

Characteristics of included studies

Most studies were conducted in the USA (n = 23; 59%). The next most commonly reported study location was Canada (n = 6; 15%). EPAs were developed across a range of specialties (see Table 1). The most common specialties in which EPAs were developed were internal medicine (n = 14; 36%), paediatrics (n = 8; 21%) and psychiatry (n = 4; 10%). Papers reported the development of EPAs for graduate learners in different postgraduate years; most were described by authors as having been developed for 'Residents' (n = 17; 44%).

As can be seen in Table 1, the majority of papers (n = 36; 92%) applied multiple distinct methods in the development of EPAs reported (median = 3, range = 1–5). Table 2 provides information on the definitions, frequency and examples of the various methods that authors reported using during EPA development processes. Prior to drafting an initial list of EPAs, 12 papers (31%) reported a literature search to gain insight into the specialty. Most, but not all, authors who conducted a literature review explicitly used this review of the literature to inform the development of their draft EPAs (n = 10; 26%). Other methods employed to develop initial draft EPAs included working groups (n = 27; 69%) and interviews or focus groups (n = 6; 15%). After the EPAs were drafted, most research teams then revised or refined them iteratively (n = 31; 79%), using methods detailed in Table 2.

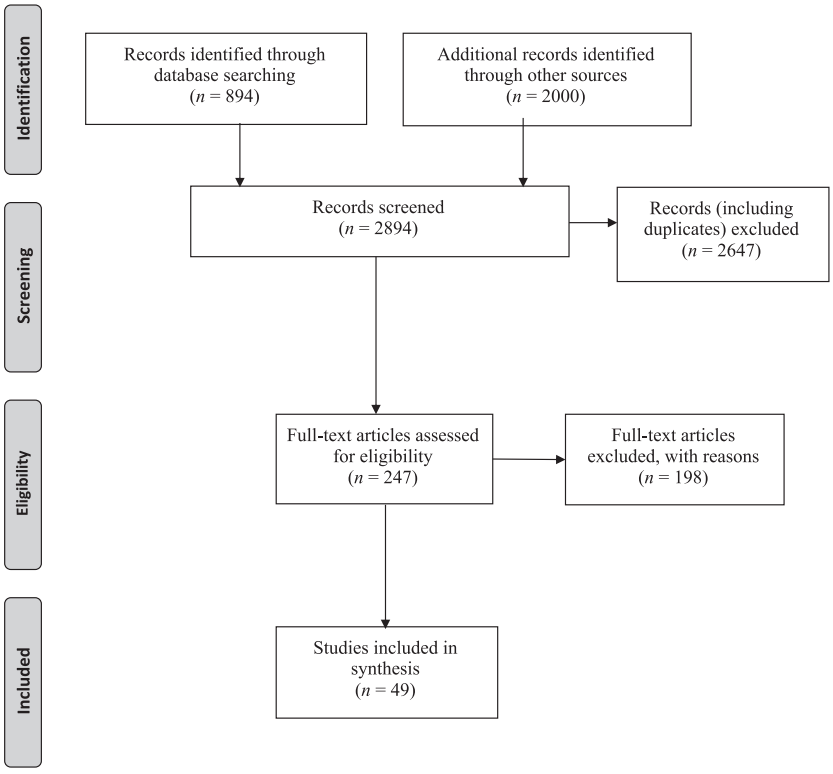

Figure 1 PRISMA flow diagram

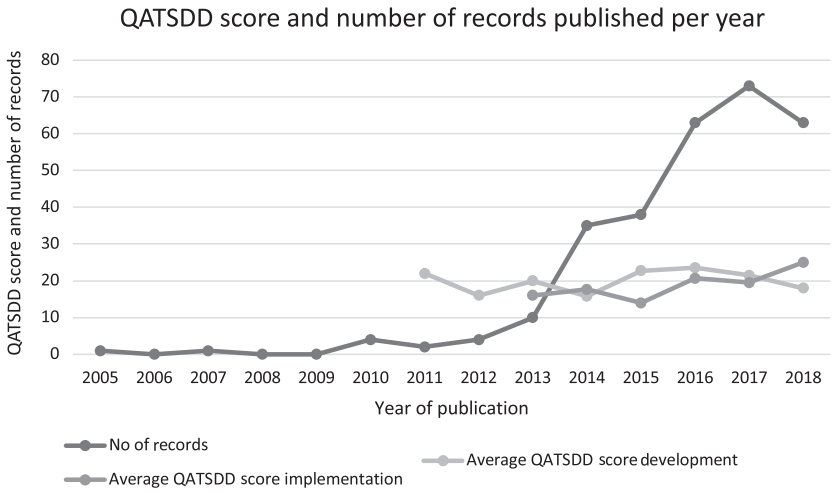

Figure 2 Web of Science results for 'Entrustable Professional Activit* OR entrustability' and the average Quality Assessment Tool for Studies with Diverse Designs (QATSDD) scores of included development and implementation papers by year of publication. First included development paper was published in 2011, and first included implementation and assessment paper was published in 2013

Methodological quality

Development papers had a mean QATSDD score of 20 out of a possible 48 (SD = 7.6; range = 6–41). The included studies scored well on items relating to the use of a particular theoretical framework, providing clear and explicit aims in the report, and by describing in detail the procedure for collecting data. Studies scored poorly on other items, such as those relating to justifying selection of a method of analysis, demonstrating evidence that sample size was considered in terms of the analysis, and reasoning why a particular data collection method was used. Appendix S4 provides a detailed table of how development papers scored on various aspects of the QATSDD. The QATSDD scores per year can be seen presented in a graph in Fig. 2.

Papers describing the implementation outcomes and/or assessment of EPAs

Key characteristics of these papers are summarised in Table 1 (for detail on individual studies, see Appendix S5). Studies described the implementation of EPAs only (n = 1; 8%), assessment of trainees using an EPA framework only (n = 5; 42%), or both (n = 6; 50%).

Characteristics of included studies

Papers emerged most frequently from the USA (n = 5; 42%), followed by Canada (n = 3; 25%). The remaining papers (n = 4; 32%) came from four different countries or regions (see Table 1). Studies within this category reported EPAs across eight different specialties. Only internal medicine and family medicine were the focus of more than one study. The grade of physicians that were the focus of the implementation or assessment efforts varied across studies (see Table 1). 'Residents' and 'Fellows' were how participants were most often described (n = 7; 59%, and n = 3; 25%, respectively). Table 2 provides specific information on the definitions, frequency and examples of the various methods used during EPA implementation and assessment.

The implementation of EPAs was discussed by 50% of papers (n = 6). As can be seen in Table 2, implementation of EPAs frequently took the form of enrolling trainees in new or existing competency-based curricula to support the EPAs (n = 4; 33%) or trainees observing faculty members or more senior clinical staff on the ward during the early stages of entrustment (n = 2; 17%), along with a number of other methods.

The vast majority (n = 11; 92%) of papers in this category described assessment of trainees using EPAs, which took the form of discussing either the methods of assessment or the specific assessment tools used. The most common method of assessment was through observation of the trainees performing the EPA in practice (n = 7; 58%). A variety of tools were used to assess the trainees' performance on an EPA, including the global entrustment scale devised by ten Cate and Scheele[10] (n = 3; 25%) and number of errors per

Table 1 Key characteristics of included studies

Characteristics	No and % of development studies	No and % of implementation and assessment studies
Study location		
USA	23 (59)	5 (42)
Canada	6 (15)	3 (25)
Australia and New Zealand	5 (13)	1 (8)
The Netherlands	2 (5)	1 (8)
Germany	2 (5)	0 (0)
India	1 (3)	1 (8)
Singapore	0 (0)	1 (8)
Specialty*		
Internal medicine (Total)	14 (36)	3 (25)
Unspecified	8 (21)	2 (17)
Rheumatology	2 (5)	0 (0)
Gastroenterology	2 (5)	0 (0)
Pulmonary and critical care	1 (3)	0 (0)
Geriatrics	1 (3)	0 (0)
Nephrology	0 (0)	1 (8)
Paediatrics (Total)	8 (21)	1 (8)
Unspecified	5 (13)	1 (8)
Neonatology	1 (3)	0 (0)
Paediatric emergency medicine	1 (3)	0 (0)
Developmental-behavioural	1 (3)	0 (0)
Psychiatry	4 (10)	1 (8)
Emergency medicine	3 (8)	0 (0)
Family medicine and primary care	3 (8)	2 (17)
Radiology	2 (5)	0 (0)
Anaesthesiology	2 (5)	0 (0)
General surgery	2 (5)	1 (8)
Hospice and palliative medicine	2 (5)	0 (0)
Pathology (Total)	2 (5)	1 (8)
Unspecified	1 (3)	0 (0)
Histopathology and/or cytopathology	1 (8)	1 (8)
Obstetrics and gynaecology	1 (3)	1 (8)
Physical medicine and rehabilitation	1 (3)	0 (0)
Haematology and oncology	1 (3)	0 (0)
Public health and preventative medicine	1 (3)	0 (0)
Orthopaedics	0 (0)	1 (8)
Multiple specialties (unspecified)	1 (3)	1 (8)
Grade of physician		
Residents	17 (44)	7 (59)
Fellows	7 (18)	3 (25)
Interns	3 (8)	1 (8)
Not specified	13 (33)	0 (0)
Attendings	0 (0)	1 (8)
No of methods used to develop EPAs		NA
1	3 (8)	-
2	16 (41)	-
3	15 (38)	-
4 or more	5 (13)	-
Type of paper	NA	
Implementation and assessment	-	6 (50)
Assessment only	-	5 (42)
Implementation only	-	1 (8)

*Percentages do not total to 100% as several papers reported on EPAs relating to more than one specialty.
NA = not applicable; EPA, entrustable professional activities.

chart, when the EPA involved writing a chart ($n = 1$; 8%).

Methodological quality

Quality Assessment Tool for Studies with Diverse Designs scores for implementation and assessment papers varied substantially. Mean score on the QATSDD was 19.5 out of a potential 48

Table 2 Codes, frequencies and examples from included studies

Development process	Description of process	Papers using method (*n*; % of development papers)	Example from included studies
Literature review	A review of related research literature was conducted prior to developing draft EPAs. Authors do not explicitly state that the EPAs developed were developed from this review	12 (31)	Literature review performed to identify papers related to physician handovers[25]
Initial EPAs drafted	Discrete list of EPAs drafted	37 (95)	
From literature review	Authors explicitly state that EPAs were drafted based on review extant literature	10 (26)	List of paediatric EPAs identified using literature review[28]
By working group	EPAs were drafted iteratively by a working group	27 (69)	Set of EPAs identified by curriculum committee[26]
From interviews and focus groups	EPAs were drafted using qualitative methodologies	6 (15)	Findings from interview and focus group study used to produce EPAs[33]
EPAs revised and refined	List of EPAs edited and streamlined	31 (79)	
Delphi method	A Delphi process with stakeholders was used to revise and refine initial EPAs	10 (26)	Multi-round Delphi panel validates and refines EPAs[52]
Survey	An online or offline questionnaire was sent to stakeholders and used to refine EPAs	8 (21)	Paediatric residents completed an online survey[37]
Stakeholder deliberation	An iterative, unstructured discussion by stakeholders and working groups were used to refine EPAs	20 (51)	Authors presented the draft list to a steering committee of internal medicine stakeholders[29]
EPAs mapped to milestones and competencies	The concordance between EPAs and other related CBME frameworks was examined after the development of EPAs	11 (28)	A total of 15 relevant milestones mapped to EPA[30]
Curriculum objectives developed	The desired outcomes of the CBME and EPA curriculum were established as EPAs were developed	2 (5)	Ensured the EPAs developed covered the skills residents needed by end of training[55]
EPAs grouped by specialty	The EPAs were reviewed and grouped by specific specialties, when developed for more than one specialty	1 (3)	The EPAs were pooled according to specialty, as multiple specialty EPAs were developed[36]
Steps recommended by ten Cate for EPA development[11] consulted	The steps recommended by ten Cate[11] for EPA development were consulted to ensure EPA(s) fit criteria	4 (10)	Process of EPA development was based on steps outlined by ten Cate and Young[49]
EPAs tested for completeness in clinical setting	The draft EPAs were applied within clinical setting to identify any gaps or flaws	1 (3)	The EPAs were tested for completeness in an out-patient centre for 18 months[57]
EPAs merged with previously developed sets of EPAs	The EPAs that were developed are then merged with existing EPAs	1 (3)	Two lists of EPAs reconciled into a single set[48]

(SD = 6.36; range = 6–32), and a graph visualising QATSDD scores per year can be found in Fig. 2. Studies scored well on items relating to the discussion of strengths and limitations, stating their aims clearly in the report, and describing the setting in which the study was conducted. Studies

Table 2 (Continued)

Development process	Description of process	Papers using method (*n*; % of development papers)	Example from included studies
EPAs benchmarked with other sets of EPAs	The scope of each EPA was compared with other sets of EPAs	1 (3)	Benchmarked with other specialties' EPAs to define aims and processes for EPA development[42]

Implementation and assessment codes	Description of codes	No of papers (% of implementation papers)	Example from included studies
Implementation			
Teaching sessions	The EPAs were explained and taught to trainees in the context of teaching sessions	2 (17)	Teaching sessions delivered on how to write discharge summaries and the EPA framework[66]
Observation of faculty members on the ward	The EPAs were introduced by having trainees observe faculty members performing the EPA	2 (17)	Trainees observe faculty members conducting EPA 'chairing multidisciplinary rounds' on three occasions[64]
Enrolment in competency-based curriculum	The trainees were enrolled in a wholly competency-based programme of learning, which included EPAs	4 (33)	To assess trainee competencies, summative EPAs were introduced into the curriculum, which was already focused on developing competency[67]
Peer feedback	Trainees' performance on EPAs was discussed in groups and feedback from peers was given on performance	1 (8)	Interactive session for interns to review peers' discharge summaries and give feedback[66]
EPA performance recorded by trainee in portfolio	A trainee records and reflects on their performance on an EPA in a portfolio	2 (17)	Authors incorporated the 36 EPAs for family medicine into existing generic field notes[70]
Assessment (methods)			
EPA observed and assessed in practice by senior faculty member	Trainees' performance of EPAs was assessed in clinical practice by staff observing them	6 (50)	Trainee encounters with patients reviewed and evaluated by senior staff when the patient was in the clinic[73]
Non-clinical performance	Trainees were assessed on their performance of an EPA in a simulated setting	4 (33)	EPAs assessed using OSCEs/ Sawbone model[65]
Portfolio review	Trainees' records of their performance on specific EPAs were assessed by faculty members	2 (17)	Programme director reviews portfolio on request of trainee for entrustment[72]

Table 2 (Continued)

Implementation and assessment codes	Description of codes	No of papers (% of implementation papers)	Example from included studies
Chart-based audit	Teaching faculty member conducted an audit of trainees' charts to assess their performance on an EPA	2 (17)	Descriptive, retrospective, chart-based audit conducted to assess performance of fine-needle aspiration biopsy EPA[68]
Comparison with control group	Trainees were assessed on performance and compared on their results with a control group not using EPAs	1 (8)	The mean global rating score on each EPA was compared between the two groups[65]
Written examination	Trainees' learning of EPAs was assessed via a written and multiple-choice examination	1 (8)	Content knowledge was assessed using a multiple-choice examination[71]
Assessment (tools and measures)			
Assessment form	Non-standardised assessment form was used by study to measure trainee performance on EPAs	2 (17)	Levels of entrustment on new scale were assigned to fellows by teaching faculty member[69]
Standardised rubric	A standardised rubric was used in the study to measure performance of trainees on EPAs	1 (8)	Attending physician assesses entrustability based on a standardised rubric[66]
Global entrustment scale	The ten Cate global entrustment scale was used to assess trainee entrustment level on an EPA	3 (25)	Assessed at baseline, 3 and 6 months[64]
Number of errors	The number of trainee errors in a chart was used to establish their level of entrustment on an EPA	1 (8)	No of deficiencies assessed every 6 months[73]
Time to competency	The time it took trainees to reach a suitable level of performance was assessed	1 (8)	Time taken to achieve adequacy rate of 85% assessed[68]

Percentages do not sum to 100, as more than one code could be assigned to an individual study.
CBME, competency-based medical education; EPA, entrustable professional activities.

scored poorly on items relating to sample size consideration, information on participant recruitment, and involving users or stakeholders in the study design. For more detail on the scores of individual studies on aspects of the QATSDD, see Appendix S4.

DISCUSSION

The use of EPAs in medical education has become increasingly widespread. This systematic review aimed to clarify the current state of knowledge relating to the development and implementation of EPAs in GME. We synthesised the evidence from 49 papers (published 2011–2018). Key findings include the range of methods used to develop EPAs, and importantly the variability in the quality, rigour and scope of studies describing the development of EPAs. The knowledge gained regarding development of EPAs from this review has implications for educators and researchers who wish to develop new EPAs that are valid for assessing trainees. Regarding the implementation and assessment of EPAs in clinical practice, this review highlighted a lack of studies to date. However, from

the small number of implementation studies identified, it is clear that implementation offers the opportunity to determine whether EPAs are a valid form of assessment.

The review found substantial variation in the EPA development process. It is clear from the included studies that a standardised approach to the development of EPAs has not emerged. Methods used to develop EPAs included Delphi groups, literature reviews and focus groups.[25–45,47–52,54–58,60–63] Although no apparent 'best' method could be determined, this systematic review is nonetheless an important overview of the various processes that can contribute to EPA development in the future. The use of standardised guidelines may also be a means of improving the development process. The suggested template for developing EPAs follows three steps: initial development, expansion and validation.[11] However, only four studies included in our review reported adhering to this guideline. Future research should explore how this template may impact on the overall quality and validity of the EPAs developed. Another development guideline[9] suggests a maximum of 20–30 EPAs for GME curricula. Many included papers far exceeded this number, as detailed in the results section (range = 1–76). Often, the papers that exceeded this recommended maximum developed EPAs that were specific, with a narrow scope. This approach to EPA development risks reintroducing the issues with CBME. (e.g. too much paperwork for educators).[74] Broad EPAs that in turn link to multiple competencies and reduce the complexity of the curriculum enable a holistic view of the learner and therefore lend themselves better to implementation.[7,75] The information collated in this review on development methods may inform the content of guidelines for developing EPAs in a standardised manner. A systematic development process will be key to successful implementation of these frameworks.[76] Establishing a set of standards for the development of EPAs, and ensuring that the development is informed by these standards, will potentially increase the likelihood of EPAs being a valid and effective assessment of trainees in clinical practice.

The variability in the quality scores of the development studies is interesting. Although it can be difficult to balance methodological quality and practical success,[77] it is important that researchers developing EPAs give consideration to the quality of their approach. Because the current review includes studies that were published over the space of 7 years, this may account for some of the variability in quality. Although a comparison of papers published across these 7 years may be somewhat inappropriate, with more recent work learning from and potentially improving on earlier publications (e.g. ten Cate building his definition of EPAs a few years after proposing them[11]), this review learned lessons from examining the overall body of literature on EPAs in GME, new and old. Also, although publication date may have played a role, other reviews in medical education have identified similar variability in quality.[23,78] This variability may indicate a need to broadly examine methods in medical education research or to develop methodological quality assessment tools better suited to the field of research. Variable developmental quality may impact on the quality of resulting EPAs and in turn their implementation. However, this issue has been infrequently examined; most of the included studies did not assess the quality of their EPAs. Researchers have recognised the importance of producing high-quality EPAs, with two tools for measuring EPA quality available.[79,80] Linking scores on these tools to the methods used could provide more insight into the preferred methods for development, and in turn, the case for implementing EPAs will be strengthened if greater consideration is given to quality.

The current review has demonstrated that the implementation and assessment of EPAs have been reported infrequently. Further, the implementation studies that were found varied considerably in methods used. These studies used multiple tools (e.g. global rating scale[64] and peer feedback[66]) to implement and assess EPAs, perhaps highlighting the lack of knowledge on best practice for implementation. However, the implementation papers to date do give some insight into the potential of various assessment methods, which could be explored further. The role of simulation in the assessment of EPAs warrants further exploration, given its successful use in a number of the included studies.[46,65] Similarly, the use of a portfolio to track trainee progress on EPAs prior to a summative entrustment decision being made shows some promise. The portfolio could potentially be accessed by supervisors at critical moments to provide an empirical basis for entrusting a learner with a task in practice.[66] Recent discussion in the literature on entrustability scales, such as that by Rekman et al,[81] could be integrated with future implementation studies to help ground decisions for supervisors of residents and ensure the validity of assessments. Pilot studies in GME to date

have been on a small scale with existing data on implementing EPAs limited to the included studies. At this point in the work on EPAs, implementation is the most crucial aspect. Further data on validity, feasibility and utility of EPAs in GME are necessary, which could be collected by comparing and correlating performance on EPAs with other variables such as patient outcomes. It is essential that going forward, studies do not just develop EPAs, but also report on these aspects. Failure to do so will inhibit the progression of the field, and our understanding of whether or not EPAs improve on existing CBME principles in GME, will not be clear.

Limitations

This review has several limitations. First, the search process required studies to explicitly state a focus on EPAs. This excluded by default the United Kingdom (UK) construct of 'foundation professional capabilities'.[82] This decision to exclude 'foundation professional capabilities' was considered defensible, as although it may be conceptually similar, the capabilities fall outside of ten Cate's definition of an EPA.[8] There is also recognised difficulty in the language used in CBME,[83] and stringent inclusion criteria for this review were intended to reduce this difficulty. Second, because of resource limitations, only papers published in English were included. However, there is some evidence to show that restricting to one language does not inherently bias a review.[84] Third, the grey literature search was limited to the first 1000 results on Google and Google Scholar, out of a total of 59 000 and 3510 results, respectively (based on searches on 4 September 2018). However, previous research has recommended that the first 200–300 results on Google Scholar are used for grey literature searches for systematic reviews,[85] which this paper exceeded. Previous research has also critiqued the lack of best practice guidelines for identifying grey literature relevant to systematic reviews.[86] Next, the review focused only on GME. This decision was made because, at this level, a lack of competence in the clinical environment can seriously impact on patient safety.[7] Junior doctor rotations can also vary substantially[87] and are often less structured in their curriculum than undergraduate teaching. Future research could explore whether including the full continuum from UME to GME would improve our understanding of EPAs in general. Finally, the highly dynamic nature of research and development work on EPAs may be considered a limitation of a systematic review conducted at this point in time. With new publications emerging rapidly and increasing year upon year (see Fig. 2), a systematic review conducted in a year or thereafter could produce different conclusions. However, systematic reviews contribute importantly to the advancement of a field,[19] and this review offers an important overview of much of the work to date along with indicating where future research and work must progress.

Recommendations and future research

Our synthesis suggests several recommendations for future research relating to EPAs. First, the resources required for implementing EPAs must be considered in terms of the benefits. Although economic assessment is widely applied in health care, it is uncommon in medical education.[88] However, the cost of an educational intervention is an important consideration. Accurate estimations of the costs of introducing EPAs could determine their value. The issue of increased paperwork upon introduction of EPAs has been raised in existing implementation studies[73] and must be examined further to ensure the benefits of EPAs would outweigh the difficulties. An assessment of the impact of EPA introduction on the workload of both trainees and assessors[14] is essential to clarify the practicalities of introducing EPAs.

Secondly, there is potential for the use of EPA quality assessment tools to determine the best method of development for EPAs in GME, and for establishing the validity of EPAs. Although this review found that studies that scored highly on the QATSDD often used more than one method of development for their EPAs, it could not answer the question as to what methods are the most effective for developing high-quality EPAs. To answer this question, it is necessary to assess the quality of EPAs that are developed, and this would be an interesting and important area for future research, providing insight into what methods should be prioritised, and included in a potential standardised template for developing EPAs. Future researchers could use the EQual[80] or the QUEPA[79] tools to determine the quality of EPAs, in tandem with the QATSDD[21] measure of methodological quality to explore the relationship between these variables. In turn, the validity of EPAs could be assessed by looking at the data from quality assessment tools such as those previously mentioned and linking the score on the tool with the number and type of methods used to develop the EPAs.

Finally, the studies included in this review come from multiple specialties and countries. However,

some specialties had more than one set of EPAs developed across different countries.[26,60] This raises a question for future research into EPAs internationally. Should specialties develop one overarching set of international EPAs? Several alternate viewpoints on this have been presented within the research literature.[89,90] For example, each country may have unique requirements for a particular issue. This would justify developing independent EPAs. However, others have noted that physician migration is common,[90] which could cause challenges. One potential solution is to have flexible international core EPAs for each specialty, which can be added to, as appropriate, in local contexts.

CONCLUSION

This review comprises the first attempt to synthesise some of the research pertaining to EPAs within GME. Research has focused on the development of EPAs, with considerably less attention devoted to the implementation of EPAs. There is a clear need for: (i) establishment of some consistency in the EPA development process, consideration of how this process is related to EPA quality, and whether development methods should differ depending on the healthcare context; (ii) increased consideration of the methodological quality of research reporting the development of EPAs; (iii) a focus on measuring the quality of EPAs using tools such as QUEPA or EQual, and (iv) more research investigating the implementation of EPAs in clinical settings, the validity of EPAs in practice and the tools used to assess trainees. This review, through synthesising the research to date on EPAs within GME and identifying the gaps in the field, will enable the advancement of EPA research and ensure that this field of research progresses fruitfully. It also provides a rationale for future work to focus on ascertaining the feasibility, validity and utility of EPA frameworks in GME by synthesising efforts to implement EPAs to date.

Contributors: (EO'D) searched the databases, led the data extraction, coding, quality analysis and synthesis, and wrote the initial draft of the manuscript. SL planned the study, contributed to data extraction and assisted with writing the manuscript. PO'C planned the study and contributed to writing and redrafting the manuscript. CM contributed to data extraction and coding, and to redrafting the manuscript. DB planned the study and drafted and redrafted the manuscript. All authors (EO'D, SL, PO'C CM, DB) have seen and approved the submitted version of the manuscript.
Acknowledgements: none.
Funding: this research was partially supported by the Health Services Executive (HSE) National Doctors Training and Planning (NDTP), Dublin, Ireland.
Conflicts of interest: none.
Ethical approval: this paper does not report research with human subjects, so ethical approval was not required.

REFERENCES

1 Langdale L, Schaad D, Wipf J, Marshall S, Vontver L, Scott C. Preparing graduates for the first year of residency: are medical schools meeting the need? *Acad Med* 2003;**78** (1):39–44.
2 Lypson ML, Frohna JG, Gruppen LD, Woolliscroft JO. Assessing residents' competencies at baseline: identifying the gaps. *Acad Med* 2004;**79** (6):564–70.
3 Bell RH Jr, Biester TW, Tabuenca A, Rhodes RS, Cofer JB, Britt LD, Lewis FR Jr. Operative experience of residents in US general surgery programs: a gap between expectation and experience. *Ann Surg* 2009;**249** (5):719–24.
4 Ten Cate O, Hart D, Ankel F *et al*. Entrustment decision making in clinical training. *Acad Med* 2016;**91** (2):191–8.
5 Frank JR, Snell LS, ten Cate O *et al*. Competency-based medical education: theory to practice. *Med Teach* 2010;**32** (8):638–45.
6 Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach* 2007;**29** (7):648–54.
7 Van Loon KA, Driessen EW, Teunissen PW, Scheele F. Experiences with EPAs, potential benefits and pitfalls. *Med Teach* 2014;**36** (8):698–702.
8 Ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ* 2005;**39** (12):1176–7.
9 ten Cate O. Nuts and bolts of entrustable professional activities. *J Grad Med Educ* 2013;**5** (1):157–8.
10 ten Cate O, Scheele F. Viewpoint: competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med* 2007;**82** (6):542–7.
11 ten Cate O, Young JQ. The patient handover as an entrustable professional activity: adding meaning in teaching and practice. *BMJ Qual Saf* 2012;**21** (1 Suppl): i9–12.
12 Hanson JL, Bannister SL. To trust or not to trust? An introduction to entrustable professional activities. *Pediatrics* 2016;**138** (5):1–4.
13 Mulder H, ten Cate O, Daalder R, Berkvens J. Building a competency-based workplace curriculum around entrustable professional activities: the case of physician assistant training. *Med Teach* 2010;**32** (10): e453–9.
14 Lomis K, Amiel JM, Ryan MS, Esposito K, Green M, Stagnaro-Green A, Bull J, Mejicano GC, AAMC Core

EPAs for Entering Residency Pilot Team. Implementing an entrustable professional activities framework in undergraduate medical education: early lessons from the AAMC core entrustable professional activities for entering residency pilot. *Acad Med* 2017;**92** (6):765–70.
15 Scheele F, Teunissen P, Luijk SV *et al.* Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach* 2008;**30** (3):248–53.
16 Lomis KD, Obeso VT, Whelan AJ. Building trust in entrustment: pursuing evidence-based progress in the core entrustable professional activities for entering residency. *Acad Med* 2018;**93** (3):341–2.
17 Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;**7** (1):10.
18 Oxman AD, Guyatt GH. The science of reviewing research. *Ann N Y Acad Sci* 1993;**703** (1):125–34.
19 National Health and Medical Research Council. NHMRC levels of evidence and grades for recommendations for developers of clinical practice guidelines. Canberra, Australia: National Health and Medical Research Council 2009.
20 Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med* 2009;**6** (7): e1000100.
21 Sirriyeh R, Lawton R, Gardner P, Armitage G. Reviewing studies with diverse designs: the development and evaluation of a new tool. *J Eval Clin Pract* 2012;**18** (4):746–52.
22 Hardy M, Johnson L, Sharples R, Boynes S, Irving D. Does radiography advanced practice improve patient outcomes and health service quality? A systematic review. *Br J Radiol* 2016;**89** (1062):20151066.
23 Walton M, Harrison R, Burgess A, Foster K. Workplace training for senior trainees: a systematic review and narrative synthesis of current approaches to promote patient safety. *Postgrad Med J* 2015;**91** (1080):579.
24 Lydon S, Cupples ME, Murphy AW, Hart N, O'Connor P. A systematic review of measurement tools for the proactive assessment of patient safety in general practice. *J Patient Saf* 2017. [Epub ahead of print.]
25 Aylward M, Nixon J, Gladding S. An Entrustable Professional Activity (EPA) for handoffs as a model for EPA assessment development. *Acad Med* 2014;**89** (10):1335–40.
26 Boyce P, Spratt C, Davies M, McEvoy P. Using entrustable professional activities to guide curriculum development in psychiatry training. *BMC Med Educ* 2011;**11** (1):96.
27 Brown CR, Criscione-Schreiber L, O'Rourke KS, Fuchs HA, Putterman C, Tan IJ, Valeriano-Marcet J, Hsieh E, Zirkle S, Bolster MB. What is a rheumatologist and how do we make one? *Arthritis Care Res* 2016;**68** (8):1166–72.
28 Carraccio C, Englander R, Gilhooly J, Mink R, Hofkosh D, Barone MA, Holmboe ES. Building a framework of entrustable professional activities, supported by competencies and milestones, to bridge the educational continuum. *Acad Med* 2017;**92** (3):324–30.
29 Caverzagie KJ, Cooney TG, Hemmer PA, Berkowitz L. The development of entrustable professional activities for internal medicine residency training: a report from the Education Redesign Committee of the Alliance for Academic Internal Medicine. *Acad Med* 2015;**90** (4):479–84.
30 Chan B, Englander H, Kent K, Desai S, Obley A, Harmon D, Kansagara D. Transitioning toward competency: a resident-faculty collaborative approach to developing a transitions of Care EPA in an internal medicine residency program. *J Grad Med Educ* 2014;**6** (4):760–4.
31 Chang A, Bowen JL, Buranosky RA, Frankel RM, Ghosh N, Rosenblum MJ, Thompson S, Green ML. Transforming primary care training - patient-centered medical home entrustable professional activities for internal medicine residents. *J Gen Intern Med* 2013;**28** (6):801–9.
32 Deitte LA, Gordon LL, Zimmerman RD, Stern EJ, McLoud TC, Diaz-Marchan PJ, Mullins ME. Entrustable professional activities: ten things radiologists do. *Acad Radiol* 2016;**23** (3):374–81.
33 El-Haddad C, Damodaran A, McNeil HP, Hu W. A patient-centered approach to developing entrustable professional activities. *Acad Med* 2017;**92** (6):800–8.
34 Fehr F, Weiß-Becker C, Becker H, Opladen T. Entrustable professional activities in post-licensure training in primary care pediatrics: necessity, development and implementation of a competency-based post-graduate curriculum. *GMS J Med Educ* 2017;**34** (5):1–7. English version.
35 Fessler HE, Addrizzo-Harris D, Beck JM, Buckley JD, Pastores SM, Piquette CA, Rowley JA, Spevetz A. Entrustable professional activities and curricular milestones for fellowship training in pulmonary and critical care medicine report of a multi society working group. *Chest* 2014;**146** (3):813–34.
36 Graafland M, ten Cate O, van Seventer JP, Schraagen JMC, Schijven MP. Mapping the demand for serious games in postgraduate medical education using the entrustable professional activities framework. *Games Health J* 2015;**4** (5):381–6.
37 Hamburger EK, Lane JL, Agrawal D, Boogaard C, Hanson JL, Weisz J, Ottolini M. The referral and consultation entrustable professional activity: defining the components in order to develop a curriculum for pediatric residents. *Acad Pediatr* 2015;**15** (1):5–8.

38 Hauer KE, Kohlwes J, Cornett P, Hollander H, ten Cate O, Ranji SR, Soni K, Iobst W, O'Sullivan P. Identifying entrustable professional activities in internal medicine training. *J Grad Med Educ* 2013;**5** (1):54–9.

39 Hsu D, Nypaver M, Fein DM *et al*. Essentials of PEM fellowship part 2: the profession in entrustable professional activities. *Pediatr Emerg Care* 2016;**32** (6):410–8.

40 Klein MD, Schumacher DJ, Sandel M. Assessing and managing the social determinants of health: defining an entrustable professional activity to assess residents' ability to meet societal needs. *Acad Pediatr* 2014;**14** (1):10–2.

41 Kwan J, Crampton R, Mogensen LL, Weaver R, van der Vleuten CPM, Hu WCY. Bridging the gap: a five stage approach for developing specialty-specific entrustable professional activities. *BMC Med Educ* 2016;**16** (1):117.

42 Landzaat LH, Barnett MD, Buckholz GT *et al*. Development of entrustable professional activities for hospice and palliative medicine fellowship training in the United States. *J Pain Symptom Manage* 2017;**54** (4):609–16.e1.

43 Leipzig RM, Sauvigné K, Granville LJ, Harper GM, Kirk LM, Levine SA, Mosqueda L, Parks SM, Fernandez HM, Busby-Whitehead J. What is a geriatrician? American Geriatrics Society and Association of directors of geriatric academic programs end-of-training entrustable professional activities for geriatric medicine. *J Am Geriatr Soc* 2014;**62** (5):924–9.

44 Mallow M, Baer H, Moroz A, Nguyen VQC. Entrustable professional activities for residency training in physical medicine and rehabilitation. *Am J Phys Med Rehabil* 2017;**96** (10):762–4.

45 McCloskey CB, Domen RE, Conran RM *et al*. Entrustable professional activities for pathology: recommendations from the College of American Pathologists Graduate Medical Education Committee. *Acad Pathol* 2017;**4**. https://doi.org/10.1177/2374289511714283

46 McMurray L, Hall AK, Rich J, Merchant S, Chaplin T. The nightmares course: a longitudinal, multidisciplinary, simulation-based curriculum to train and assess resident competence in resuscitation. *J Grad Med Educ* 2017;**9** (4):503–8.

47 Meade LB, Suddarth KH, Jones RR, Zaas AK, Albanese T, Yamazaki K, O'Malley CW. Patients, nurses, and physicians working together to develop a discharge entrustable professional activity assessment tool. *Acad Med* 2016;**91** (10):1388–91.

48 Moloughney B, Moore K, Dagnone D, Strong D. The development of national entrustable professional activities to inform the training and assessment of public health and preventative medicine residents. *Can Med Educ J* 2017;**8** (3):e71.

49 Moore D, Young CJ, Hong J. Implementing entrustable professional activities: the yellow brick road towards competency-based training? *ANZ J Surg* 2017;**87** (12):1001–5.

50 Myers J, Krueger P, Webster F *et al*. Development and validation of a set of palliative medicine entrustable professional activities: findings from a mixed methods study. *J Palliat Med* 2015;**18** (8):682–90.

51 O'Keeffe M. Clinical competence in developmental-behavioural paediatrics: raising the bar. *J Paediatr Child Health* 2014;**50** (1):3–10.

52 Parker T, Guiton G, Jones M Jr. Choosing entrustable professional activities for neonatology: a Delphi study. *J Perinatol* 2017;**37** (12):1335.

53 Pugh D, Cavalcanti RB, Halman S, Ma IW, Mylopoulos M, Shanks D, Stroud L. Using the entrustable professional activities framework in the assessment of procedural skills. *J Grad Med Educ* 2017;**9** (2):209–14.

54 Rose S, Fix OK, Shah BJ, Jones TN, Szyjkowski RD. Entrustable professional activities for gastroenterology fellowship training. *Neurogastroenterol Motil* 2014;**26** (8):1204–14.

55 Schultz K, Griffiths J, Lacasse M. The application of entrustable professional activities to inform competency decisions in a family medicine residency program. *Acad Med* 2015;**90** (7):888–97.

56 Sharma P, Tanveer N, Goyal A. A search for entrustable professional activities for the 1st year pathology postgraduate trainees. *J Lab Physicians* 2018;**10** (1):26.

57 Shaughnessy AF, Sparks J, Cohen-Osher M, Goodell KH, Sawin GL, Gravel J Jr. Entrustable professional activities in family medicine. *J Grad Med Educ* 2013;**5** (1):112–8.

58 Shumway NM, Dacus JJ, Lathrop KI, Hernandez EP, Miller M, Karnad AB. Use of milestones and development of entrustable professional activities in 2 hematology/oncology training programs. *J Grad Med Educ* 2015;**7** (1):101–4.

59 Taylor DR, Park YS, Smith CA, Karpinski J, Coke W. Creating entrustable professional activities to assess internal medicine residents in training. *Ann Intern Med* 2018;**168** (10):724–9.

60 Weiss A, Ozdoba A, Carroll V, DeJesus F. Entrustable professional activities: enhancing meaningful use of evaluations and milestones in a psychiatry residency program. *Acad Psychiatry* 2016;**40** (5):850–4.

61 Wisman-Zwarter N, van der Schaaf M, ten Cate O, Jonker G, van Klei WA, Hoff RG. Transforming the learning outcomes of anaesthesiology training into entrustable professional activities. *Eur J Anaesthesiol* 2016;**33** (8):559–67.

62 Wölfel T, Beltermann E, Lottspeich C, Vietz E, Fischer MR, Schmidmaier R. Medical ward round competence in internal medicine - an interview study towards an interprofessional development of an Entrustable Professional Activity (EPA). *BMC Med Educ* 2016;**16** (1):174.

63 Young JQ, Hasser C, Hung EH, Kusz M, O'Sullivan PS, Stewart C, Weiss A, Williams N. Developing end-of-training entrustable professional activities for psychiatry: results and methodological lessons. *Acad Med* 2018;**93** (7):1048–54.

64 Ng LB, Ng Joo Ming M. Entrustable professional activities to enhance continuity of care. *Med Educ* 2014;**48** (11):1115.

65 Dwyer T, Wadey V, Archibald D, Kraemer W, Shantz J, Townley J, Ogilvie-Harris D, Petrera M, Ferguson P, Nousiainen M. Cognitive and psychomotor entrustable professional activities: can simulators help assess competency in trainees? *Clin Orthop Relat Res* 2016;**474** (4):926–34.

66 Hauer KE, Soni K, Cornett P, Kohlwes J, Hollander H, Ranji SR, ten Cate O, Widera E, Calton B, O'Sullivan PS. Developing entrustable professional activities as the basis for assessment of competence in an internal medicine residency: a feasibility study. *J Gen Intern Med* 2013;**28** (8):1110–4.

67 Jurd S, de Beer W, Aimer M, Fletcher S, Halley E, Schapper C, Orkin M. Introducing a competency based Fellowship programme for psychiatry in Australia and New Zealand. *Australas Psychiatry* 2015;**23** (6):699–705.

68 Kumar HB, Khan AM, Arora VK, Singh N. Fine needle aspiration biopsy: an entrustable professional activity in cytopathology postgraduate training. *J Cytol* 2017;**34** (2):84–9.

69 Mink RB, Schwartz A, Herman BE, Turner DA, Curran ML, Myers A, Hsu DC, Kesselheim JC, Carraccio CL, Steering Committee of the Subspecialty Pediatrics Investigator Network (SPIN). Validity of level of supervision scales for assessing pediatric fellows on the common pediatric subspecialty entrustable professional activities. *Acad Med* 2018;**93** (2):283–91.

70 Schultz K, Griffiths J. Implementing competency-based medical education in a postgraduate family medicine residency training program: a stepwise approach, facilitating factors, and processes or steps that would have been helpful. *Acad Med* 2016;**91** (5):685–9.

71 Steiman J, Sullivan SA, Scarborough J, Wilke L, Pugh C, Bell RH, Kent KC. Measuring competence in surgical training through assessment of surgical entrustable professional activities. *J Surg Educ* 2018;**75** (6):1452–62.

72 Van Loon KA, Teunissen PW, Driessen EW, Scheele F. The role of generic competencies in the entrustment of professional activities: a nationwide competency-based curriculum assessed. *J Grad Med Educ* 2016;**8** (4):546–52.

73 Yuan CM, Prince LK, Zwettler AJ, Nee R, Oliver JD 3rd, Abbott KC. Assessing achievement in nephrology training: using clinic chart audits to quantitatively screen competency. *Am J Kidney Dis* 2014;**64** (5):737–43.

74 Touchie C, ten Cate O. The promise, perils, problems and progress of competency-based medical education. *Med Educ* 2016;**50** (1):93–100.

75 Carraccio CL, Englander R. From flexner to competencies: reflections on a decade and the journey ahead. *Acad Med* 2013;**88** (8):1067–73.

76 Ten Hoor GA, Plasqui G, Schols AM, Kok G. Development, implementation, and evaluation of an interdisciplinary theory-and evidence-based intervention to prevent childhood obesity: theoretical and methodological lessons learned. *Front Public Health* 2017;**5**:352.

77 Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;**3** (1):28.

78 Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, Prideaux D. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. *Med Teach* 2006;**28** (6):497–526.

79 Post J, Wittich C, Thomas K, Dupras D, Halvorsen A, Mandrekar J, Oxentenko AS, Beckman TJ. Rating the quality of entrustable professional activities: content validation and associations with the clinical context. *J Gen Intern Med* 2016;**31** (5):518–23.

80 Taylor DR, Park YS, Egan R, Chan M-K, Karpinski J, Touchie C, Snell LS, Tekian A. EQual, a novel rubric to evaluate entrustable professional activities for quality and structure. *Acad Med* 2017;**92** (11 Suppl):S110–7.

81 Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med* 2016;**91** (2):186–90.

82 The UK Foundation Programme Curriculum. *The Foundation Programme Curriculum 2016*. Birmingham, UK: UKFPO 2016.

83 Englander R, Frank JR, Carraccio C, Sherbino J, Ross S, Snell L. Toward a shared language for competency-based medical education. *Med Teach* 2017;**39** (6):582–7.

84 Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, Mierzwinski-Urban M, Clifford T, Hutton B, Rabb D. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care* 2012;**28** (2):138–44.

85 Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PLoS ONE* 2015;**10** (9):e0138237.

86 Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. *Syst Rev* 2015;**4** (1):138.

87 Mehta NB, Hull AL, Young JB, Stoller JK. Just imagine: new paradigms for medical education. *Acad Med* 2013;**88** (10):1418–23.

88 Lin Y, Cheng A, Hecker K, Grant V, Currie GR. Implementing economic evaluation in simulation-based medical education: challenges and opportunities. *Med Educ* 2018;**52** (2):150–60.
89 Miccinesi G, Fischer S, Paci E, Onwuteaka-Philipsen BD, Cartwright C, van der Heide A, Nilstun T, Norup M, Mortier F, EURELD consortium. Physicians' attitudes towards end-of-life decisions: a comparison between seven countries. *Soc Sci Med* 2005;**60** (9):1961–74.
90 Gouda P, Kitt K, Evans DS *et al*. Ireland's medical brain drain: migration intentions of Irish medical students. *Hum Resour Health* 2015;**13** (1):11.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1. Table of search terms, dates of searches, and number of results by database.
Appendix S2. Tables of codes developed for synthesis of data.
Appendix S3. Data extraction table: development papers.
Appendix S4. QATSDD breakdown for included studies.
Appendix S5. Data extraction table: implementation and assessment papers.

Received 31 May 2018; editorial comments to authors 5 November 2018; accepted for publication 20 November 2018

