



Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Scarff, CE; Bearman, M; Chiavaroli, N; Trumble, S

Title:

Trainees' perspectives of assessment messages: a narrative systematic review

Date:

2019-03-01

Citation:

Scarff, C. E., Bearman, M., Chiavaroli, N. & Trumble, S. (2019). Trainees' perspectives of assessment messages: a narrative systematic review. *Medical Education*, 53 (3), pp.221-233. <https://doi.org/10.1111/medu.13775>.

Persistent Link:

<https://hdl.handle.net/11343/285356>

DR. CATHERINE ELIZABETH SCARFF (Orcid ID : 0000-0002-7717-950X)

DR. MARGARET BEARMAN (Orcid ID : 0000-0002-6862-9871)

MR. NEVILLE G CHIAVAROLI (Orcid ID : 0000-0003-1488-9747)

PROF. STEPHEN C TRUMBLE (Orcid ID : 0000-0002-8097-5785)

Article type : Research Papers

TITLE PAGE

Trainees' perspectives of assessment messages. A narrative systematic review.

Catherine E Scarff MBBS, MMed, MHPE, FACD; Margaret Bearman, PhD; Neville Chiavaroli BAppSc, BA (Hons), MEd, MPhil; Steve Trumble MBBS, MD, FRACGP.

Dr Catherine E Scarff, C/O Department of Medical Education, Melbourne Medical School, Room E728, Level 7 East, Medical Building, Grattan Street, University of Melbourne Vic 3010, Phone: +61 3 9035 3270. Email: scarffc@student.unimelb.edu.au

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/medu.13775](https://doi.org/10.1111/medu.13775)

This article is protected by copyright. All rights reserved

Abstract

PURPOSE

To conduct a narrative systematic literature review on medical specialist trainees' perspectives of the assessment messages they receive in the context of clinical performance assessments. The aim of the study was to determine if the trainees value the information received through the formats which are designed to promote their development, and if not, the reasons for this.

METHODS

The authors searched the ERIC, Embase, Ovid Medline and PsycINFO databases for articles published up to 16 June 2018 which presented original data on trainees' perspectives of the assessment messages they receive in the context of work based assessments (WBA) and in-training assessments (ITA) used within their training programme. After screening by title by one author, all authors screened the total of 938 abstracts and 139 full-text articles were assessed. Descriptions of quantitative data and thematic analysis of qualitative data were used to present the opinions of the trainees.

RESULTS

33 articles met inclusion criteria. Twenty-six articles (79%) described trainees' perspectives in the context of WBA with the remaining describing ITA formats. Wide-ranging opinions were reported and the analysis categorised these into three themes: trainees value developmental assessment messages; trainees become disengaged when assessment messages aren't developmental; trainees' views depend on the environment, the assessor and themselves. Some trainees reported the assessment messages were valuable and provide input on their performance to guide their development; but many disagreed. In particular, the trainee's own level of engagement with the assessments influenced perspectives on the messages received.

CONCLUSIONS

Trainees do not universally perceive that clinical performance assessments provide them with the valuable developmental input on their performance as they were

designed. Factors related to the trainee, the assessor and the environment influence their perspectives.

Author Manuscript

Keywords

Evaluation/assessment of clinical performance, personal characteristics/attitudes, postgraduate training.

Author Manuscript

Introduction

Clinical performance assessments – defined here as work based assessments (WBA), in-training assessments (ITA) or in-training evaluations (ITER) – are designed to provide medical trainees with holistic information about the quality of their practice. They recognise that “[t]he highest level of evidence of attainment of a skill is the demonstration of the behaviour in routine daily practice, i.e. performance in vivo.” (1). The workplace setting also allows the assessment of additional attributes (2) and provides trainees with the opportunity to receive qualitative and quantitative judgements on their performance from multiple assessors during the course of a placement. Whether primarily formative or summative in design, the information provided is intended to promote improvement. However, trainees and their assessors sometimes hold very different views regarding the performance information given and received in general (3). It is important to understand trainees’ views on the assessment messages they receive in the workplace in order to know if they are achieving the aim of providing guiding information from the trainee’s perspective. If these assessment messages are not helpful, then reasons for that need to be explored.

We use the term “assessment message” to emphasise consideration of quantitative information, in addition to written or verbal comments, as the former is sometimes not included in feedback research. We also use the term to deliberately distinguish the information provision that is of interest here from more recent versions of feedback, which have broadened beyond the “teacher-specific behavioural principles” of information delivery (eg timely, specific) (4) to include the learner’s barriers and reactions to feedback and the relationship between the teacher and learner (4). Indeed feedback can be seen as a matter of co-construction between teacher and learner (5). Our focus here is on the initial part of the feedback process – the message – comprising both ratings and comments that are delivered to the trainee.

To date, scholarly papers have evaluated many aspects of performance assessments, including their validity or reliability (6), their educational impact and feasibility (7,8) and evidence of their impact on doctors’ future work (2). Many of these have studied this from the perspective of the assessor, however, none have determined trainees’ perspectives on the information received from the two types of clinical performance assessment together. This is important as perspectives may be relevant to

both formats. Bing-You, Hayes, Varaklis, Trowbridge, Kemp and McKelvy recently conducted a scoping review of what is broadly known about feedback in health professions education. Though high-level evidence was lacking, repeated and immediate feedback was seen to be beneficial for learning and the importance of additional aspects including emotions, credibility of the source and demographics was highlighted (9). Massie and Ali (10) performed a critical narrative literature review on medical trainee and assessors' views of WBAs in training, finding widespread negativity towards them, particularly around their perceived purpose and associated logistics, such as time available to perform them and training for those involved. Poor quality feedback to trainees was a common outcome reported.

The goal of our study was to extend these works and bridge the gap between assessment and feedback literature by focusing specifically on medical trainees' perspectives of the assessment messages that they receive in the context of clinical performance assessments. In order for these assessments to achieve their aims, trainees need to value the assessment information they receive. We therefore sought to collate published reports on trainee's perceptions of their worth. The question for this systematic review is "What are medical specialist trainees' perspectives of the assessment messages they receive in clinical performance assessments?"

Methods

Definitions and explanations

“Medical specialist trainees” refers to qualified doctors in a medical specialty training programme: for example, those who are training to become surgeons, physicians or general practitioners. In North American, these doctors are known as residents, while in the UK, Australia and New Zealand, they are termed registrars. For simplicity, we refer to them as “trainees” from here.

“Perspective”, refers to “a way of regarding situations, facts etc, and judging their relative importance” (11). We use this term as it was commonly used in the research question or aims of included articles; it also encompasses a broad range of ideas by including the concepts of attitudes, outlooks and views.

An “assessment message” describes the information provided on the assessment form (ratings, comments) and any verbal information delivered to the trainee by the assessor about their performance.

“Clinical performance assessments” refers to assessments of trainees conducted in authentic settings, by an assessing clinician with the delivery of an assessment message to the trainee. They occur within usual clinical practice, rather than a formal examination or simulation setting, and the assessment content is ‘natural’ and not predetermined and so varies between learners. Clinical performance assessments therefore do not include formats such as Objective Structured Clinical Examinations (OSCE) or similar.

This review was performed with an interpretive stance. That is, we sought to interpret others’ interpretations by inductively grouping results from studies into themes of similar ideas. Ethical approval for this literature review was not required.

Search strategy and information sources

The review followed the PRISMA (12) guidelines for systematic reviews.

The databases used for the search were Ovid Medline, Embase, PsycINFO, and ERIC which were selected in consultation with library staff as being the most likely to yield references relevant to the research question. In the search design phase, databases were explored for appropriate subject headings relating to the question. The list of

participant and interest search terms (English language only) were combined by 'and' to generate the search yield (table 1).

Author Manuscript

(Table 1 near here)

Database searches were undertaken on 26 October 2016, with an update performed on 16 June 2018 (no limit was placed on the start date for the search), yielding 9248 articles. Citations were managed in Endnote X7 (Thomas Reuters, New York, New York). Duplicated papers (n=3663) and non-English papers were removed (n=260). Exclusion of articles by title (n=4391) was conducted by CS. The remaining articles (n=938) were imported into Covidence (Melbourne, Australia). A double abstract review was undertaken (CS for all and ST, NC and MB approximately one third of the articles each) with any disagreements resolved by discussion and reference to the inclusion and exclusion criteria (table 2). Four additional articles were obtained through hand searching of the reference lists of the 32 articles identified for data extraction and one was included for data extraction.

(Table 2 near here)

Author Manuscript

Full-text review for inclusion or exclusion was performed by CS with all other authors double reviewing approximately 34% (n=47/139) of the articles. Double review was performed where there was uncertainty as to an article's inclusion status. Thirty-three articles met inclusion criteria and data extraction and quality assessment was conducted by CS. Double review of quality assessment on 14/33 (42%) articles was performed by MB, ST and NC. Consensus on disagreements was achieved through discussion. See figure 1.

(Figure 1 near here)

Inclusion and exclusion criteria

Articles were included if they presented qualitative and/or quantitative data describing trainee reports of perspectives on the assessment messages received in a clinical performance setting in current usage in their training program. This included logistical aspects, such as the time spent getting feedback and opinions on its quantity, quality and usefulness. Articles that reported pilot assessments were not included as perspectives on feedback in these likely differ from those in actual use. Articles that noted trainees' satisfaction with or general perspectives on the assessment process only were not included.

In addition to ITA, the WBAs in included articles are:

- mini-CEX
- Direct Observation of Procedural Skills (DOPS)
- Case based Discussions (CbD)
- Multisource Feedback tools (MSF)
- Procedure Based assessments (PBA)
- Objective assessment of Surgical and Technical Skills (OSATS)
- Multiple Consultant Report (MCR)

Many of the studies included used the term "feedback" to describe what we have defined as the "assessment message" and we have not modified their expression. The reader is alerted to this and asked to be mindful that "feedback" in this context relates to the "assessment message".

Data extraction

Data extraction from included articles' results section was performed by CS. Each article was read carefully and data were entered into a spreadsheet in Microsoft Excel for Mac Version 15.30 (Microsoft). Data extracted included the author details, year and country of publication, assessment type, research aim, data collection method, number and type of participants, and the qualitative and quantitative data on trainees' perspectives of assessment messages. Supporting quotes were recorded. Quality assessment was performed using the MERSQI tool (13) for quantitative data and the Walsh and Downe tool (14) for qualitative data.

Data analysis and synthesis

Data analysis was performed through an interpretive paradigm and thematic analysis was the method employed for the synthesis. The data extracted from the results of included studies were organised into open data-derived codes and all instances of codes were collated together; in this way, all studies contributed data to the thematic analysis (15). Similar codes were clustered together into themes and CS determined three categories that ran across the themes. The qualitative analysis from the articles with the highest quality assessment (16–23) informed this process. In this way, the tension between the analytic work done by authors and our own analytic work was balanced (24). Additional themes were inductively coded and included in an overall theme as required. MB and CS met to iteratively discuss codes, categories and themes. MB re-read the original data of six of the top papers to review the final categories. Quantitative data were included in this analysis and presented in a descriptive form with the theme to which they relate.

While we did not use the analyses from studies with lower results on the two quality assessment tools used (level determination described in Appendix 2), we did include their reported data. This decision was made post hoc. While we determined that lower quality articles had a valuable contribution to make, they had fewer interpretive approaches to the data and therefore contributed less to our own interpretive processes.

Results

Study Characteristics

The 29 articles included in this systematic review are summarised in Appendix 1.

Most studies were from the UK (n=23,70%), with others from Canada (n=3, 9%), Ireland (n=2, 6%), the Netherlands (n=2, 6%) and one (3%) each from Argentina, Australia and New Zealand and India. Twenty-six (79%) articles covered WBA (mini-CEX n=4, 12%, CbD n=4, 12%, MSF n=2, 6%, PBA n=3, 9%, OSATS n=1, 3%, MCR n=1, 3%, mixed n=11, 33%) and seven (21%) focused on ITA. Fourteen (42%) of the studies were primarily quantitative in design, 13 (39%) were primarily qualitative and six (18%) were mixed methods. Where studies included qualitative and quantitative data, a quality assessment for each part was conducted. Eight studies achieved ten or more 'yes' scores overall on the Walsh and Downe scale while 11 scored less (14). Seventeen studies scored 7 or more on the MERSQI scale (13), with 3 scoring less than 7.

Thematic analysis

There were three main categories derived from thematic analysis which addressed the research question. These are:

- **trainees value developmental assessment messages;**
- **trainees become disengaged when assessment messages aren't developmental; and**
- **trainees' views depend on the environment, the assessor and themselves.**

These are outlined in turn.

Trainees value developmental assessment messages

Many trainees highly valued the assessment messages they obtained from the clinical performance assessments. They reported that the assessment messages they received were fair (25) and useful (25–28), and promoted deeper understandings and learning (29,30):

"if it is done well you can get an awful lot of learning opportunities out of it and ...it gives you an awful lot of feedback" (22).

They described the assessment messages as beneficial by being real-time and objective (31), though some recognised they could be subjective (25). They noted the assessment messages indicated and documented their performance (30,32,33) so increasing awareness of their strengths and weaknesses (31):

"The feedback was useful, either for confirming what I had done well and for suggesting other things I could have done" (30).

■ Trainees often ascribed different values to the individual components of the assessment messages, with qualitative information valued above quantitative:

"I think the most important thing about all of these things is the feedback. It's not the mark you get at the end of the day" (23)

For example, one study on the MSF reported that 89% of trainees agreed the comments were useful, compared with 38% agreeing the numeric report was useful (34). Trainees particularly valued the verbal delivery of performance information (35); in Sabey and Harris, 94% of trainees ascribed their verbal feedback as useful compared with 74% for written comments (20). This was especially when in-depth conversations occurred (19) and areas of weakness were identified (18,29,31,36):

"... You look for assessors that you know are knowledgeable and where you get something out of it, a good dialogue or really learn something. Not just marks on a sheet of paper ..." (33).

Others highlighted that verbal feedback was more effective in achieving outcomes (such as by addressing specific learning points or self-directed learning (28)).

Written narrative information, with detailed examples, was highly valued, especially when it explained ratings (21) or was coupled with verbal input as to how to work to achieve competencies (30). These high values placed on written comments and verbal discussions with assessors were noted for both the WBA (18,23,25,32,36) and ITA formats (19):

"you can actually figure out why you didn't meet [expectations] or why you were doing well so you can keep doing it" (21).

Some trainees wished to avoid quantitative scores altogether (34) or at least to have a better understanding of scales (19), consistency in their use (20) or a form redesign to focus on developmental feedback (37).

Evidence of an effect on learning was limited and was generally based on self-report alone. For example, between 20-29% of trainees felt they improved practice because of feedback (35,38), with those who reported receiving good quality feedback being more likely to agree (35). Some felt that the assessment messages, often coupled with dialogue, helped their development of reflective skills (25,37,39) and self-assessment skills (19).

Trainees highlighted additional aspects of the assessment messages they valued such as their breadth including coverage of nontechnical skills (18) or giving a more holistic picture of performance (23,25,34):

"What I think is more useful is when you get feedback about how you're performing as a trainee in general" (25)

Some acknowledged the ability of the message to provide encouragement (30) or reassurance (36) and to convey the feeling of being appreciated (20), and some noted that feedback being affirming was a potential outcome in itself (16). Some trainees reported the assessment messages led to an increased confidence with aspects of their work (31):

"It's nice that they have approved that you can do such and such. You get confidence - I know I have been able to do this correctly" (33).

Others highlighted the limitations of positive feedback:

"positive feedback is nice but it's not always going to be productive for you" (22).

In addition, some acknowledged the potential difficulties for assessors with giving negative feedback :

"people always say they want negative feedback, but people don't take it as well... puts the evaluator in a very awkward position" (21).

Trainees become disengaged when assessment messages are not developmental

By contrast, many trainees reported highly unfavourable views of their assessment messages, often noting "widespread dissatisfaction" (21) with the process.

Between 33-41% of trainees reported they rarely or never received verbal feedback immediately after a WBA (27, 38), with 16% receiving no feedback in one study (40):

"But the majority of consultants will just sort of go "Oh God, you want one of these again". Fill it in really quickly, give you no feedback whatsoever and it is just a tick box" (22).

Wrongly or rightly, this left some trainees to infer from their feedback:

"...if nobody has criticised you or had a go at you then you know you are alright" (22)

When trainees did receive feedback, not all reported it to be useful or helpful (36,41):

"...it's a box ticking exercise. I only do it because I have to. I don't do it because I think it benefits me in any way." (22)

Trainees reported variability in rating (17,18) and that assessors vary in their level of leniency (18).

Many trainees noted frustration at a lack of explanation of written or verbal comments (36):

"They don't mention anything in the notes or just one line and that doesn't show you a great deal as in which areas there are that you should improve and that sort of leaves you in a state of why did they mark you poor?" (42).

and poor quality comments were also noted (21,28):

"It's all just platitudes" and "They'll write something a bit vague and a bit meaningless" (20)

Some trainees doubted the ability of the assessments to reflect their progress (43) or to detect struggling trainees (17,18,23) both in the ITA, where 39% of urology trainees in one UK study believed they would not identify a failing trainee (44) and WBA formats, where only 45% of GP trainees in the UK agreed that WBAs could detect struggling trainees (20).

"Generally they are a waste of time for a good doctor. They won't be useful to differentiate good from bad doctors" (17).

Some trainees noted that assessors could concentrate on unrepresentative episodes (21) or be deficit-focused (16), noting that feedback delivered poorly can be “demotivating” (30) and if too negative, counterproductive (42).

These perspectives lead to a view of such assessments as ‘form-filling’ (35) or ‘tick box’ initiatives (16,17,22,39,41,45):

“Now it is just so people can tick the boxes. The inspiration and pride in our profession has gone” (17).

Trainees’ views depend on the environment, their assessors and themselves

The reviewed articles highlighted factors that influence the trainee’s perspectives on the assessment messages they have received. These included the environment, the assessors and themselves.

Environment

The time available for feedback influenced trainee perspectives. Many noted a lack of dedicated time (28) and the consequent potential for poor quality feedback with “limited educational value” (25). They reported that feedback usually took less than five minutes (for the CbD (27), PBA (40) and WBA in general (38)) but in some studies, only the minority (14%) felt time for feedback was insufficient (36), while other studies reported this varied with the trainees’ level. For example, the majority (67%) of senior trainees in Tsagakataki and Choudhary’s study felt that feedback time with the OSATS was insufficient (41).

Perspectives also depended on timeliness of receipt of assessment messages. While some trainees noted completion of the form and delivery of feedback was mostly timely (35), others disagreed, reporting delays in receiving assessment messages (19,27,28,38) or difficulties with getting the forms completed in full (17) or at all (42):

“The main obstacle or issue I find is from higher levels and not actually from me getting them achieved...finding consultant input to sit down and go through them with to sign them off was a difficulty...it's so difficult to get them...” (23)

The setting was important, with trainees reporting the negative impact of interruptions and inability to speak freely (25) on feedback occurring in a public place.

Assessor

Many trainees felt that the “style” of trainer impacted heavily on the assessment message provided (39). The background of the assessor was relevant: for example, GP assessors were seen to be more skilled than hospital based ones (20). The seniority of the assessor was contentious; some reported that senior registrars provided better feedback than consultants (45), while others reported that consultant feedback was more likely to be constructive (43) and honest:

“[r]egistrar feedback tends to always be positive, but consultants will tell you how it is” (39)

Assessor enthusiasm in the process was very influential, with almost 80% of trainees in one study reporting that “the enthusiasm of their trainer for PBAs determined the extent of benefit to be gained from using them” (35). This engagement was highlighted in many studies (19,21,25,39) and trainees reported that feedback from an engaged assessor is more valuable :

“If you choose someone who wants to teach and can teach, it makes a big difference. If they just fill the paperwork in then you're not going to get anything from it” (25).

Unfortunately, large numbers of trainees felt that assessors were not engaged in assessments (17,20):

“[t]he trainers aren't engaged...the majority...they'll just help you in order to get by” (22)

The assessor’s skills in giving feedback were reported as very important for feedback to result in learning and improvement (30). Some trainees reported concerns about assessor motivation and how personal preferences or biases (32,43), victimisation (36) and retribution (25) could affect the assessment message.

Many trainees raised concerns about the level of honesty in their assessment messages. This stemmed from the influence of interpersonal relationships (20,42) and from poor assessment practices such as retrospective completion (16–18) or completion of parts of the form related to unwitnessed tasks (19,20):

“... they just go through and fill them out anyway” (19)

Half the participants in one study reported they had been told to complete their *own* forms (23). Some trainees doubted the honesty of messages delivered in face to face settings such as mini-CEX and CbD (20) and were aware of assessor reluctance to document deficiencies (22):

"What you want from an assessment is someone to actually tell you what your fault is" (28).

Themselves

Perspectives on the assessment messages were further influenced by trainees' own characteristics and behaviours. Training level influenced views of feedback by tool. For example, in one study, 52% of senior trainees felt CbDs were more useful at junior levels (39); while the MSF was reported to be useful in all years of training (34). Some senior trainees highlighted their desire for different types of feedback:

"I am wanting to be assessed more globally. A sort of overall impression of "am I almost ready for consultant practice or not" " (23).

Many studies showed that trainees influenced their assessment messages strategically. Those who didn't just want "marks on a sheet of paper" (33) chose assessors they knew were serious about assessments (33), and whom they anticipated would give them quality feedback (18) and take time to go through the assessment with them immediately afterwards (35). Others focused on ratings only (20) or "gamed" the system (18,23), seeking out assessors known to be lenient in their assessments (22):

"If I know a person wouldn't rate me high, even for a teaching thing, I would rather find someone else. In terms of real development, that's not how it should be" (37).

Others avoided assessors they felt gave negative feedback especially when the assessment carried implications such as supporting progress decisions (22):

"Like I don't get certain consultants to do things particularly because arguably they're more critical on paper and then in an ARCP environment it then looks bad on me" (22).

Some strategically chose cases (18,20,22,23,25), with between 2 - 11% of trainees agreeing they often /always avoided a "complex or difficult cases owing to the fear of receiving negative feedback" (27,38):

"I actively pick a procedure that I know I did well in and ask someone to rate me on it" (22)

Some acknowledged the consequences of this on the assessment message:

"I rarely bring forward patients that I did badly with and so rarely get negative feedback. So it's useful but there's not a lot of constructive criticism to go on most of the time" (25).

Discussion

Summary of findings

This review collates the perspectives of medical trainees on the assessment messages they receive through clinical performance assessments. While these assessments have a well-established position in the program of assessments undertaken in many speciality training programs and are designed to provide trainees with valuable input on their performance, this review shows that trainees do not universally agree that the assessment messages they receive achieve this aim. Instead, it shows that trainees' perspectives can vary widely even within the same study (17,25) and depend on multiple factors.

Their own and their assessor's engagement in the assessment process were seen to be major drivers on perspectives on the assessment messages received. The assessors' enthusiasm and engagement was fundamentally important (19,21,25,35,39) and predicted the quality of assessment messages. Unfortunately, many trainees cited examples of disengaged behaviour by their assessors (19,20,22,27,38), and so had poor opinions of the resultant assessment messages. However, in contrast, where assessors were engaged, trainees often valued their messages highly (33).

The assessor's engagement often predicted the trainee's engagement, and further influenced perspectives on the messages received. Three studies (18,22,42) highlighted the polarisation of trainees into two groups – those who valued formative feedback and sought quality information to promote their learning and development

(learning goal oriented); and those who were results-focused, primarily interested in receiving positive comments and high marks (performance goal oriented). The former group clearly expressed their desire for more information – for explanation of their assessor's views on their performance (21) and their frustrations when this does not occur (42). They recognised a lack of honesty (20), calling for more truth in the assessment messages provided (28,39), though they acknowledged the difficulties of doing this (21). Reluctance to deliver negative messages is well studied in the psychology literature (46). The performance-orientated group focused on their goal of just getting the assessment ticked off (22) rather than deriving any educational benefit. This aligns with work from Dijksterhuis, Schuwirth, Braat, Teunissen and Scheel who distinguished performance-orientated from mastery-orientated behaviours in assessors and trainees in general formative assessment situations (47).

These different trainee behaviours may not stem solely from innately different trainee types. Some trainees may be fundamentally performance goal orientated, while the circumstances of the assessments may alter the orientation of others. Ongoing resentment to the introduction of these assessments has been noted (23), and challenging environmental circumstances exacerbate this. Lack of time and an appropriate location for feedback (25) promote a cursory approach to them, as may imposing quotas for completion (23); most assessors and trainees do not think that increasing numbers of assessments leads to improvement in clinical skills (48). These factors coupled with assessor disengagement in the assessments would likely challenge the most enthusiastic learning goal orientated trainee to become more performance orientated.

Implications for practice

Identification of the problems with assessment messages from clinical performance assessments is the first step is to ensuring they function as intended.

The lack of engagement, in part related to ongoing resentment at their being “dropped on trainees and [assessors]” (23), may be reduced by reorientation to the assessments, with the close involvement of trainees and assessors. This should include their input in defining the purpose of the assessments in each training context and so minimise a tick-box view of them (16,17,22,39,41,45). Addressing trainee desires for written and verbal comments on their performance and increasing emphasis on

constructive comments is important (49), and others have also reported that feedback is often lacking (50–52). Comments could be provided alone, as introduced in some settings (53). Training assessors in assessment and feedback is clearly necessary for production of quality assessment messages, but similarly, training of trainees in how to incorporate assessment messages into practice may be required. That is, while many of the notions about co-construction through feedback (5) are evident in the data (particularly through third theme Trainees' views depend on the environment, the assessors and themselves), many assessors simply write a message for trainees and both assessors and trainees should understand how to maximise the value of this performance information.

This study highlights that environmental aspects of these assessments are equally important to address and neglecting to do so will limit their potential. The training organisation (49) must provide more support, the number of assessments to be completed must be balanced and justified (for example (48)) and dedicated time to conduct quality assessments provided. Some have implemented ways to address this, for example with dedicated WBA clinics (54), but time and service pressures remain a significant challenge for most and other options need to be developed.

In addition to optimising environmental aspects, recognition of the difficulties of delivering negative assessment messages is needed. The pervasive human tendency of reluctance to provide negative information (46) needs to be acknowledged, and emphasising the educational aspects of these tools may encourage assessors to deliver a clear assessment message, even when it is 'negative'. Some authors have recognised how the culture of medicine and "reluctance to reveal errors and shortcomings" (22) may also impact on the delivery of assessment information and this is an important area for further research from an educational perspective.

The use of clinical performance assessments in a training program is only one step in the process of providing trainees with assessment messages that contribute meaningfully to their development. This review combines the assessment and feedback literatures, and has shown that without commitment to and engagement in the assessments by the trainee and their assessor and the provision of environmental support, resulting assessment messages will be of little value. Attending to the formative purpose of the assessments, by not only providing timely and specific but also quality information based on performance, that leads on to a feedback exchange, is key

to their benefits. Without this, the message will be disregarded at best and invalid at worst, with potential ongoing harmful effects, as trainees who are unaware of their true level of performance may not improve and subsequently fail a future examination or - of more concern - cause harm to a patient. Promoting such engagement and attention to the intended outcomes of these assessments is essential if they are to function as designed and help trainees to develop into the best doctors they can be.

Limitations

This review has several limitations. The inclusion criteria limited the search to studies published in English, meaning relevant articles in other languages were missed. This was done for logistical and financial reasons. The focus on medical specialist trainees excluded others, including junior doctors not in training programs and medical students. This was done as the consequences of assessments are different for each, and therefore perspectives on them may differ. Similarly, although four databases were searched, it is possible that relevant articles only listed in other databases were missed. Exclusion of articles by title and by only one researcher had the potential to overlook relevant articles. Similarly, data extraction by a single researcher is a further limitation and may have missed relevant information. Most of the articles included were from the United Kingdom and this “UK-centricity” of works on WBA in particular has been noted in previous studies (10).

Analysing and presenting qualitative and quantitative data together is challenging and different authors may have emphasized different areas. As highlighted, eight articles formed the framework for the analysis and while these are of higher quality, this may have unduly influenced the review.

As noted in several of the included studies, surveys of people’s perspectives tend to attract responses from those with polarised views and the data subsequently included in this study may be overly representative of such groups. However, opinions at all extremes were noted. Social desirability bias encourages socially acceptable answers from participants and is of particular relevance for data extracted from focus group or interview settings.

Conclusion

This systematic review collates the perspectives of medical trainees in speciality training programs about the assessment messages they receive in clinical performance assessment contexts. We believe this is the first collation of such opinions and they are crucial to consider in relation to further developments of these important assessment formats. Some trainees report their assessment messages are valuable, but some report they are not. This appears to stem from both their and their assessors' engagement, and importantly, environmental factors. Addressing aspects of all three will be needed for these assessments to realise their potential benefits.

Author Manuscript

References

1. Whitehouse A. Pilot study of 360 degree assessment of personal skills to inform record of in training assessments for senior house officers. *Hosp Med.* 2002;63(3):172–5.
2. Saedon H, Salleh S, Balakrishnan A, Imray CH, Saedon M. The role of feedback in improving the effectiveness of workplace based assessments: a systematic review. *BMC Med Educ.* 2012;12(1):25.
3. Sender Liberman A, Liberman M, Steinert Y, McLeod P, Meterissian S. Surgery residents and attending surgeons have different perceptions of feedback. *Med Teach.* 2005;27(5):470–2.
4. Bing-you RG, Trowbridge RL. Why medical educators may be failing at feedback. *JAMA.* 2009;302:1330–1.
5. Ajjawi R, Regehr G. When I say ... feedback. *Med Educ.*
6. McGill DA, van der Vleuten CPM, Clarke MJ. A critical evaluation of the validity and the reliability of global competency constructs for supervisor assessment of junior medical trainees. *Adv Heal Sci Educ. Advances in Health Sciences Education;* 2013 Oct 1;18(4):701–25.
7. Miller A, Archer J. Impact of workplace based assessment on doctors' education and performance: a systematic review. *BMJ.* 2010;341:c5064.
8. Weller JM, Jolly B, Misur MP, Merry AF, Jones A, Crossley JGM, et al. Mini-clinical evaluation exercise in anaesthesia training. *Br J Anaesth.* 2009;102(5):633–41.
9. Bing-You R, Hayes V, Varaklis K, Trowbridge R, Kemp H, McKelvy D. Feedback for Learners in Medical Education: What is known? A scoping review. *Acad Med.* 2017;92(9):1.
10. Massie J, Ali JM. Workplace-Based Assessment: A Review of User Perceptions and Strategies to Address the Identified Shortcomings. *Adv Heal Sci Educ.*

Advances in Health Sciences Education; 2016 May 1;21(2):455–73.

11. Collins English Dictionary [Internet]. Collins English Dictionary. [cited 2018 Aug 16]. Available from:
<https://www.collinsdictionary.com/dictionary/english/perspective>
12. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS Med*. 2013;10(4).
13. Reed DA, Cook DA, Beckman TTJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA*. 2007;298(9):1002–9.
14. Walsh D, Downe S. Appraising the quality of qualitative research. *Midwifery*. 2006;22(2):108–19.
15. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77–101.
16. Barrett A, Galvin R, Scherpbier AJJA, Teunissen PW, O’Shaughnessy A, Horgan M. Is the learning value of workplace-based assessment being realised? A qualitative study of trainer and trainee perceptions and experiences. *Postgrad Med J*. 2016;postgradmedj-2015-133917.
17. Bindal T, Wall D, Goodyear HM. Trainee doctors’ views on workplace-based assessments: Are they just a tick box exercise? *Med Teach*. 2011;33(January):919–27.
18. Castanelli DJ, Jowsey T, Chen Y, Weller JM. Perceptions of purpose, value, and process of the mini-Clinical Evaluation Exercise in anesthesia training. *Can J Anesth*. 2016;63(12):1345–56.
19. Patel R, Drover A, Chafe R. Pediatric faculty and residents perspectives on in-training evaluation reports (ITERS). *Can Med Educ J*. 2015;6(2):41–53.
20. Sabey A, Harris M. Training in hospitals: What do GP specialist trainees think of workplacebased assessments? *Educ Prim Care*. 2011;22(2):90–9.

21. Watling CJ, Kenyon CF, Zibrowski EM, Schulz V, Goldszmidt MA, Singh I, et al. Rules of engagement: residents' perceptions of the in-training evaluation process. *Acad Med*. 2008 Oct;83(10 Suppl):S97-100.
22. Gaunt A, Patel A, Rusius V, Royle TJ, Markham DH, Pawlikowska T. 'Playing the game': How do surgical trainees seek feedback using workplace-based assessment? *Med Educ*. 2017;51(9):953–62.
23. Shalhoub J, Marshall DC, Ippolito K. Perspectives on procedure-based assessments: A thematic analysis of semistructured interviews with 10 UK surgical trainees. *BMJ Open*. 2017;7(3):1–8.
24. Bearman M, Dawson P. Qualitative synthesis and systematic review in health professions education. *Med Educ*. 2013;47(3):252–60.
25. Mehta F, Brown J, Shaw NJ. Do trainees value feedback in case-based discussion assessments? *Med Teach*. 2013;35:e1166-72.
26. Goel A, Singh T. The usefulness of Mini Clinical Evaluation Exercise as a learning tool in different pediatric clinical settings. *Int J Appl Basic Med Res*. 2015;5:S32-4.
27. Grennan S, Crowley S, Quidwai S, Barrett O, Kooblall M. Is the current BST ePortfolio fulfilling its role in the training of clinical medicine SHO's? *Ir Med J*. 2016;109(1):343–6.
28. Mohanaruban A, Flanders L, Rees H. Case-based discussion: perceptions of feedback. *Clin Teach*. 2018;15(2):126–31.
29. Babu KS, Htike MM, Cleak VE. Workplace-based assessments in Wessex: the first 6 months. *Psychiatr Bull*. 2009;33(12):474–8.
30. Bodgener S, Tavabie A. Is there a value to case-based discussion? *Educ Prim Care*. 2011;22(4):223–8.
31. Malhotra S, Hatala R, Courneya C-A. Internal medicine residents' perceptions of the Mini-Clinical Evaluation Exercise. *Med Teach*. 2008;30(4):414–9.

32. Alves de Lima A, Henquin R, Thierer J, Paulin J, Lamari S, Belcastro F, et al. A qualitative study of the impact on learning of the mini clinical evaluation exercise in postgraduate training. *Med Teach*. 2005;27(1):46–52.
33. Ringsted C, Henriksen AH, Skaarup AM, Van der Vleuten CPM. Educational impact of in-training assessment (ITA) in postgraduate medical education: A qualitative study of an ITA programme in actual practice. *Med Educ*. 2004 Jul;38(7):767–77.
34. Alofs L, Huiskes J, Heineman MJ, Buis C, Horsman M, van der Plank L, et al. User reception of a simple online multisource feedback tool for residents. *Perspect Med Educ*. 2015;4(2):57–65.
35. Hunter AR, Baird EJ, Reed MR. Procedure-based assessments in trauma and orthopaedic training-The trainees' perspective. *Med Teach*. 2015;37(5):444–9.
36. Cohen SN, Farrant PBJ, Taibjee SM. Assessing the assessments: U.K. dermatology trainees' views of the workplace assessment tools. *Br J Dermatol*. 2009;161(1):34–9.
37. Simmons M. How are workplace-based assessments viewed by UK psychiatry trainees? *Psychiatr Danub*. 2013;25(SUPPL.2):182–4.
38. Taylor A, Dubrey S, Das S. Opinions of the ePortfolio and workplace-based assessments: a survey of core medical trainees and their supervisors. *Clin Med*. 2014;14(5):510–6.
39. Phillips A, Lim J, Madhavan A, Macafee D. Case-based discussions: UK surgical trainee perceptions. *Clin Teach*. 2016;13(3):207–12.
40. Marriott J, Purdie H, Crossley J, Beard JD. Evaluation of procedure-based assessment for assessing trainees skills in the operating theatre. *Br J Surg*. 2011;98(3):450–7.
41. Tsagakataki M, Choudhary A. Mersey deanery ophthalmology trainees' views of the objective assessment of surgical and technical skills (OSATS) workplace-based assessment tool. *Perspect Med Educ*. 2013;2(1):21–7.

42. Findlay N. General practitioner registrars' experiences of multisource feedback: a qualitative study. *Educ Prim Care*. 2012;23(5):323–9.
43. Menon S, Winston M, Sullivan G. Workplace-based assessment: survey of psychiatric trainees in Wales. *Psychiatr Bull*. 2009;33(12):468–74.
44. Pearce I, Royle J, O'Flynn K, Payne S. The record of in-training assessments (RITAs) in urology: An evaluation of trainee perceptions. *Ann R Coll Surg Engl*. 2003;85(5):351–4.
45. Mukhtar O, Griffin J, Naheed S, Bryant C. The Multiple Consultant Report: new assessment; same old problems. *Clin Teach*. 2018;15(1):34–7.
46. Rosen S, Tesser A. On reluctance to communicate undesirable information: The MUM effect. *Sociometry*. 1970;33(3):253–63.
47. Dijksterhuis MGK, Schuwirth LWT, Braat DDM, Teunissen PW, Scheele F. A qualitative study on trainees' and supervisors' perceptions of assessment for learning in postgraduate medical education. *Med Teach*. 2013;35(8):e1396-402.
48. Powell H, DiMarco A, Saeed S. Trainee and trainer opinion on increasing numbers of workplace-based assessments. *Bull R Coll Surg Engl*. 2014;96(5):160–2.
49. Bing-You R, Varaklis K, Hayes V, Trowbridge R, Kemp H, McKelvy D. The feedback tango: An integrative review and analysis of the content of the teacher-learner feedback exchange. *Acad Med*. 2018;93(4):657–63.
50. Jackson D, Wall D. An evaluation of the use of the mini-CEX in the foundation programme. *Br J Hosp Med*. 2010;71(10):584–8.
51. Plymale MA, Donnelly MB, Lawton J, Pulito AR, Mentzer RM. Faculty evaluation of surgery clerkship students: important components of written comments. *Acad Med*. 2002;77(10 Suppl):S45-7.
52. Sokol-Hessner L, Shea JA, Kogan JR. The Open-Ended Comment Space for Action Plans on Core Clerkship Students' Encounter Cards: What Gets Written? *Acad Med*. 2010;85(10):S110–4.

53. Devlin AP, Fazackerley J, Penfold N. Improving workplace-based assessments. *Br J Anaesth.* 2016;116(5):718–9.
54. Deepak S, Ponnampalam A, Brown C. An innovative and effective model of completing WPBAs - CbD (case based discussion) clinics. *Arch Dis Child.* 2017;102(5):A83.

Author Manuscript

Table and figure legend

Table 1: Search terms used for database searches.

Table 2: Source of articles for systematic review.

Table 3: Inclusion and exclusion criteria for systematic review.

Figure 1: PRISMA diagram for systematic review question – “What are vocational medical trainees’ views of the assessment messages they receive in clinical performance assessments?”

Appendices

Author Manuscript

Author Manuscript

Appendix 1: Data extracted from included articles

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
Alofs, Huiskes, Jan Heineman, Buis, Horsman, Van der Plank & ten Cate, 2015 (1)	MSF Netherlands	An evaluation to collect opinions on the use of the MSF and its effect (on Kirkpatrick levels) to enable improvements if necessary	Quantitative Survey – online (anonymous)	108/404 residents in anaesthesiology, OBGYN, internal medicine, paediatrics, ENT, etc. up to 6 years (22/59 programme directors)	MSF is a useful feedback tool in relation to the broad anonymous nature and comments. Risk of bias not applicable to all specialties
Alves de Lima, Henquin, Thierer, Paulin, Lamari, Belcastro & van der Vleuten, 2005 (2)	mini-CEX Argentina	To focus on how residents perceive an assessment tool and its influences on their approach to learning	Qualitative Open and semi-structured interview Buenos Aires province, Argentina	16/16 cardiology residents, all 4 years of training	Mini-CEX is a valuable tool for exercise, but anxiety can interfere with the results
Babu, Htike, & Cleak, 2009 (3)	WBA (8 types)	To determine the experience of supervisors and trainees with the first 6 months of using WBAs	Quantitative Survey (anonymous) Wessex	39/62 psychiatry trainees, all training years (1-4) (40/65 supervisors)	WBAs found very useful by trainees and barriers identified. Useful to identify strengths and weaknesses though not their ability to do so.
Bache, Brown, & Graham, 2002 (4)	RITA UK	To ascertain whether the broadened review process in the Mersey Deanery was thought beneficial	Quantitative Survey – paper Mersey Deanery	605/744 specialty trainees (309/349 trainers)	The RITA provides a structured training period and tool and is able to provide feedback to determine the educational needs of the post.
Barrett, Galvin,	WBA (mini-CEX, DOPS,	What are trainer's and trainee's perceptions of	Qualitative Semi-structured	8 medicine trainees, all in first three years	Value of WBA in assessing and supporting the force trainer for feedback

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
Scherpbier, Teunissen, O'Shaughnessy, Horgan, 2016 (5)	CbD) Ireland	the learning value of WBAs? How have trainer and trainee WBA experiences shaped these perceptions?	interviews Royal College of Physicians, Ireland	of higher specialist training (9 trainers)	learning exercise, but especially engagement are not being reified assessments they were
Bindal, Wall, & Goodyear, 2011 (6)	WBA - (mini-CEX, DOPS, CbD) UK	To explore whether WBAs had been successfully integrated into run-through paediatric specialty training programmes	Quantitative 20 item questionnaire West Midlands Deanery	130/150 regional paediatric trainees	Trainees are ambivalent of the WBA to reflect and opinions on their widely. Issues of variation and poor practices by Concerns about tick completion.
Bodgener & Tavabie, 2011 (7)	CbD UK	To explore the attitudes of trainees and trainers towards the CbD, its educational value and its ability to assess trainee performance	Mixed Questionnaire and semi-structured interviews (different participants) Kent, Surrey and Sussex GP Deanery	Questionnaire - 45/45 final year GP trainees (30 or 31/40 GP trainers) Interviews - 5 trainees (5 trainers)	Feedback is overwhelming development and to show progress, though not assessed performance to the view of the assessor centred. The trainer's skills are noted to be
Castanelli, Jowsey, Chen, & Weller, 2016 (8)	mini-CEX Australia, New Zealand	To explore trainees' and supervisors' understanding of the mini-CEX, their experience with the assessment and their perceptions of its influence on learning and supervision.	Qualitative Semi-structured telephone interviews Australia and New Zealand	17 trainees across all years of training and diverse practice locations anaesthesia (18 supervisors)	Scepticism expressed relation to leniency and poor practice by trainees feedback is appreciated feedback valued over conflict as to purpose formative or summative assessment behaviour
Cohen, Farrant, & Taibjee, 2009 (9)	WBA (Mini-CEX, DOPS, MSF) UK	To gather the views of trainees on the WBA to assess whether or not the newly introduced assessments were working as planned and to identify any	Quantitative Questionnaire – paper and electronic All UK dermatology trainees	138/269 dermatology trainees all in UK	Feedback is appreciated timely, though not all receiving useful feedback is valued, but frustration results (in DOPS and training noted. Some to be stressful and ar

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
		problems with them.			
Findlay, 2012 (10)	MSF UK	To explore the experiences of GP specialty training registrars, of the ways in which MSF impacts upon their self-perceptions and professional practice	Qualitative Semi-structured interviews Northumbria GP training programme	5 GP trainees	The educational supervisor role in helping the trainee develop from feedback. MSF is useful but often hampered by trainee reluctance of raters to give constructive feedback which can hamper use.
Gaunt, Patel, Rusius, Royle, Markham and Pawlikowska, 2017 (11)	WBA UK	To explore how WBAs influence the ways surgical trainees seek feedback and feedback interactions	Qualitative Focus groups West Midlands, London, Wales and Yorkshire	42 surgical trainees All levels of training	Trainees' perceptions of WBA and their relationship with assessors impacts on learning and what they can achieve. Distinct trainee types who view assessment as a way to play the game; those who seek learning seek to develop despite negative feedback. The timing of the WBA affected the effectiveness of informing the annual review of performance. Timing is important.
Goel & Singh, 2015 (12)	mini-CEX India	To assess the acceptability and feasibility of mini-CEX in different paediatric settings and to know if this exercise helped the residents by improved learning	Quantitative Questionnaire Ludhiana	23 paediatric residents (11 teachers)	Mini-CEX provides a structured way to help learning. Some level is of relevance and can play a role.
Goodyear, Wall, & Bindal, 2013 (13)	Annual review of competence progression (ARCP) UK	To explore specialty trainees' views on ARCPs including the usefulness of the annual review and how well prepared they felt	Mixed Questionnaire, telephone semi-structured interviews West midlands	Questionnaire: 63/70 (years 1-4) Interview: 21 (years 2-6) Paediatric trainees	The majority agreed that the ARCP was helpful, though did not always find it and were dissatisfied with it. Changes in the process were relevant. Some feedback on reflection and misinterpretation.

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
		for it.	deanery		assessment messages
Grennan, Crowley, Quidwai, Barrett, & Kooblall, 2016 (14)	ePortfolio and WPBAs Ireland	To determine if trainees and supervisors see the ePortfolio as an important learning exercise or as a time-consuming obligation	Quantitative Online survey 2 hospitals in Dublin	27/149 trainees, SHO's completing BST in clinical medicine (24/307 consultants)	There is significant difference among trainees in relation to ePortfolio (of which Engagement is key. 70% agreed the feedback is useful. Feedback is only given in 10% of cases (never in a significant way). Some trainees manipulate feedback to avoid negative feedback.
Hunter, Baird, & Reed, 2015 (15)	PBA UK	To gain an understanding of the attitudes of trauma and orthopaedic trainees regarding procedure-based assessments and identify factors influencing any perceived educational benefit.	Quantitative Internet-based questionnaire All UK trainees	616/1144 Trauma & Orthopaedic trainees Year 3 to post-CCT trainees	Formative role valued more than summative one. Over half the trainees found feedback a useful learning tool. A minority felt they improved their practice due to this. Following completion of the PBA, there was a significant increase in enthusiasm of the assessors. There was a significant effect on perceived satisfaction and outcomes. Half perceived the experience as fulfilling and different. 70% of assessors noted.
Kolli, de Waal, Shivashankar, & Rubinsztein, 2010 (16)	WBA UK	Nil stated	Quantitative Questionnaire	Psychiatry trainees (number not known)	Significant majority of trainees were satisfied with quality of feedback received as at least 3.0. 70% report that the ARCP process is fair.
Malhotra, Hatala, & Courneya, 2008 (17)	mini-CEX Canada	To assess residents' perceptions of the mini-CEX using qualitative methods	Qualitative Semi-structured focus groups (2) UBC Vancouver	12 internal medicine residents – year 1 and 2	The use of the mini-CEX assessment tool placed a high perceived value as an educational tool. Anxiety influences perceived value.

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
					decreases with experience
Marriott, Purdie, Crossley, & Beard, 2011 (18)	PBA UK	To evaluate the validity, reliability and acceptability of procedure based assessments in surgical training programmes.	Quantitative Questionnaire at end of study period Three hospitals in Sheffield	81 trainees Surgery – all types and all levels of trainee (focus on years 3-7)	Trainees report confidence in the purposes of the tool for assessment. Most are positive about feedback received, though some received no feedback. Trainees who received feedback reported it to be very useful.
McGraw, 2003 (19)	RITA UK	Review of the trainees' perception of the annual record of in-training assessments.	Quantitative Questionnaire Southwest region	34/50 trainees Paediatrics	The majority of trainees received positive and constructive comments on their performance. A detailed discussion of the needs.
Mehta, Brown, & Shaw, 2013 (20)	CbD UK	To explore trainee's perceptions of the educational value of case-based discussions specifically focusing on feedback.	Mixed Questionnaire (online), interviews Mersey Deanery	Questionnaire: 26/32 Interviews: 9 interviewed from 12 volunteers Paediatric trainees, years 1 and 2	Opinions on usefulness of the tool but many found the tool useful for learning, is fair and provides a reasonable picture of competence but can be subjective. Trainees value narrative feedback ratings. Trainees noted that the tool such as poor practice and assessors and environment constraints. Limitations of the tool for learning points and areas noted. Conflict between the tool and learning tool noted.
Menon, Winston, & Sullivan, 2009 (21)	WBA UK, Wales	To explore attitudes and perceptions regarding workplace-based assessment (WPBA) among psychiatric trainees in Wales, identify current problems and recommend measures	Quantitative Questionnaire (anonymous, cross-sectional, online) Wales	81/88 psychiatric speciality trainees (minus one who was author) All levels of training (years 1-6 plus "other")	There were widespread concerns about the WBAs. Most trainees have no real educational value of dubious validity. Concerns that assessors' background and training including background and impact of the personal factors result in an inaccurate outcome. Concerns that

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
		to successfully implement WPBA.			potential for the tools exercises. Both trainees need training in education to avoid unintended outcomes
Mohanaruban, Flanders, Rees, 2018 (22)	CbD UK	To explore how medical trainees perceive the feedback they receive during CbDs and how these assessments are being put into practice in the workplace.	Mixed Questionnaire (paper-based, distributed to whole trainee population) Two focus groups District general hospital, London, UK	78/91 medical trainees completed questionnaire, 12 volunteered to participate in focus groups All levels of training	Feedback occurs in most cases, but one trainee receiving feedback and report it is useful. Most address learning points feedback though things specific enough and convenient and afraid of Assessor engagement importance.
Mukhtar, Griffin, Naheed, Bryant, 2017 (23)	MCR UK	To explore opinions of medical trainees on the MCR	Qualitative Questionnaire (open questions and free text) and focus group A London teaching hospital	22/42 medical trainees completed questionnaire, 6 volunteered to participate in focus group Junior trainees	A lack of understanding of the assessment leads like a tick-box exercise feedback limited the the assessment. The the form may not be one to assess the trainee
Patel, Drover, & Chafe, 2015 (24)	ITER Canada	To explore the perspectives of faculty and residents on structure and process of ITER as a means to improving the use of the ITER as an evaluation tool.	Qualitative Two focus groups – trainees and assessors Memorial University of Newfoundland paediatric training program	8/23 invited paediatric resident attended All training years	Trainees unanimously is lack of understanding Written comments were most useful aspect of Residents greatly appreciate that assessors go to in The ITER is useful because designed and attention discussed could improve feedback.

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
Pearce, Royle, O'Flynn, & Payne, 2003 (25)	RITA UK	To evaluate the effect of the Calman reforms on trainees in urology with respect to their educational goals, their experience of the RITA process and its value in preparing them for consultant careers.	Quantitative Questionnaire (anonymous, paper) All UK urology trainees holding national training numbers who had completed at least one RITA review	100 replies Urology trainees All years	Trainees described the RITA process. L educational goals and felt the process would failing trainee. The n failing assessment re the training establish educational requirem
Phillips, Lim, Madhavan, & Macafee, 2016 (26)	CbD UK	To review the use of CbDs by surgical trainees to determine if and when they are useful, and whether they are perceived as being used correctly.	Qualitative Interviews Surgical trainees in two deaneries	42 surgical trainees – 21 higher surgical trainees, 21 core surgical trainees, in all subspecialties	Not all CbDs are equ useful to improve per factors impact, includ training (being most levels) and quality of vary from CbD being complete only, throu learning experience. lack of engagement i trainees and assessor
Ringsted, Henriksen, Skaarup, & Van Der Vleuten, 2004 (27)	ITA Netherlands	To investigate the experiences and opinions of programme directors, assessors and trainees about a recently introduced ITA programme	Qualitative Semi-structured interviews 3 departments of anaesthesiology at different hospitals	14 first year trainees in anaesthesiology (3 programme directors, 9 supervisors)	Viewing ITA as an e rather than just docum performance dictated Trainees reported val experience even whe an assessment. Succ assessments viewed confidence. The imp immediate feedback some assessors were
Sabey & Harris, 2011 (28)	WBA UK	To establish how the new system of WBA is working in day-to-day practice in hospital posts.	Mixed Questionnaire – Focus groups and semi-structured interviews 2 training locations in Severn Deanery	Questionnaire – 52/78 GP trainees, Focus groups and semi-structured interviews – 20/31 trainees	Trainees cited narrati useful more than sco that the quality of fee practice by assessors including bias and la judgements seen to li learning, validity and

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
					risks degeneration to Views about the usefulness of various assessments
Shalhoub, Marshall, Ippolito, 2017 (29)	PBA UK	To explore the opinions of trainees on their PBA experiences – how they perceive and make sense of them.	Qualitative Semi-structured interviews Surgical trainees across UK who had used the Intercollegiate Surgical Curriculum Programme for >1 year	10 surgical trainees all levels of training	Lack of clarity of the tool leads to problems. The tool is not intended and a quota system encourages tick-box practices are noted and seen as more important than assigned, though feedback is not obtain. The tool stands
Simmons, 2013 (30)	WBA UK	To explore trainees' perceptions of opportunities, barriers and suggestions for development of WBAs as formative learning tools	Qualitative Two focus group interviews Cambridge	6 psychiatry trainees first and fourth year trainees	Trainees note the opportunity for personalised feedback for trainees in difficulty. Trainees noted (e.g. clinical) based on their lenient WBAs are not appropriate and that there is little feedback on forms. The training of assessors is essential to enhance outcomes with
Taylor, Dubrey, & Das, 2014 (31)	WBA portfolio UK	To assess core medical trainee and trainer opinions of the ePortfolio and WPBAs.	Quantitative Surveys (paper-based) Five NHS trusts in north-west London	60/158 core medical trainees (30/62 trainers)	The vast majority view the "means to passing the" one third viewing the exercise. Poor practice by assessors reported. Direct feedback reported only 20% felt their practice. Direct feedback felt to be essential.
Tsagkatakaki & Choudhary, 2013 (32)	Objective assessment of surgical and technical skills (OSATS)	To gather the views of a sample of UK ophthalmology trainees on the OSATS WPBA for cataract surgery for	Quantitative Questionnaire – paper and online Mersey Deanery	28/34 ophthalmology trainees	Strong opinions for a Some trainees valued be observed and the formative feedback a reflective practice. S

Author, year	Assessment type, Country	Research aim / objective	Data type, data collection method Location	Participants – respondents Trainee type – year, specialty	Key findings relevant to research question
	UK	its acceptability and to identify potential problems			was unhelpful. Many feeling stressed and reported insufficient. Concerns noted that strain relationships and box behaviours.
Watling, Kenyon, Zibrowski, Schulz, Goldszmidt, Singh, Maddocks, Lingard, 2008 (33)	ITER Canada	To describe residents' experiences with and perceptions of the ITER process to gain insight into why the process succeeds or fails.	Qualitative Semi-structured interviews London, Ontario	20 residents – medical, pediatric and surgical Year 1-5	Engagement of both trainee is key to the success whether it is perceived. Features which promote each discussed.

¹ Quality assessment was performed on quantitative aspects of the studies with the MERSQI tool (34), where “higher” was defined as a score of 7 or more, and on qualitative aspects of the studies with the Walsh and Downe tool (W&D)(35), where “higher” was defined as 10 or more ‘yes’ responses to questions.

Appendix 2 - Quality assessment description

Quality assessment was performed using the MERSQI tool (12) for quantitative studies and the Walsh and Downe tool (13) for qualitative. The BEME quality assessment tool (46) was also examined but was not used for this review as a separate tool was desired for each type of study (qualitative and quantitative), which could provide a greater depth of information as compared with a tool that covered both in one. Cook and Reed have shown that the MERSQI and Newcastle-Ottawa Scale – Education (NOS-E) are reasonably well correlated (47) and the former was chosen given its increasing use (48). It is acknowledged that quality assessment of studies is a dynamic and evolving area (49). While definitions of scores which correlate with high-quality studies have been defined by some authors (for example (50), Cook and Reed advocate the use of scores for “relative rather than absolute judgements” as the scores generated are highly dependent on the area of study and the inclusion criteria (12). Walsh and Downe echo this recommendation for qualitative study assessment, suggesting that while they would expect their overall criteria to be addressed sufficiently in studies they would include in a primary meta-synthesis, they offer their tool to be used “imaginatively rather than prescriptively” (13).

For qualitative study assessment, the original article (13) states that it would be expected that higher quality studies would have ‘yes’ to the criteria in column 2 and some purely qualitative studies demonstrate this (16). Many other qualitative studies however, showed weaknesses in the audit trail and reflexivity categories, so in the current context, a study has been considered higher quality if it met 10 or more criteria.

Many of the other studies included in this study are primarily quantitative or mixed, and the qualitative criteria have been applied to the qualitative aspects only. In this group, other aspects of the studies were also noted to be weak for example ‘Demonstration of clear audit trail’ (explaining how they obtained the information – free text comments etc.) and ‘Data used to support interpretation’, as there was often not extensive use of the quotes in discussion of findings. Rather, these data were used to support the quantitative data.

In regards to quantitative studies, a higher quality study in this context was considered as one with a MERSQI score of ≥ 7 . This score was determined because the research question for the review itself precluded obtaining a high score. In explanation, given the research question asked for trainee views and perspectives, the studies included were therefore mainly observational ones, with assessment by study participant and with an

outcome of perception / attitude etc. only. This limited the potential score that could be achieved. As such the maximum score for most studies was considered to be 12.

Author Manuscript

Table 1: Search terms used for database searches

Participant		Interest
specialist trainee* specialist registrar* intern SpR not (surface plasmon resonance or spastic or secondary) registrar* interns learner education, professional/ trainee learners education, continuing/ trainees medical staff, hospital/ schools, medical/ medical trainee internship and residency/ internship health profession* student* adj3 (medical or medicine) residency resident residents exp education, medical/ health personnel/	AND	clinical rotation assessment* end of clinical rotation assessment* clinical performance report* DOPS tool patients' assessment questionnaire* t-mex work place based assessment* mini-peer assessment tool in training report* WPBA procedur* based assessment* eportfolio* work based assessment* direct observation of procedural skills in training assessment* multi source feedback in training evaluation* mini-clinical evaluation mini-CEX multisource feedback case based discussion* clinical assessment tool workplace based assessment* WBA peer assessment DOPS MSF clinical supervision assessment form* performance assessment* CBD not (central business district)

Author Manuscript

		evaluation not quality
--	--	------------------------

MeSH terms are denoted by '/'

Table 2: Inclusion and exclusion criteria for systematic review.

	Inclusion criteria	Exclusion criteria
Participant	Medical specialist trainee (vocational / specialist / postgraduate).	Medical student. Consultant doctor (i.e. Fellow), doctor not in training programme to become specialist. Other health professional groups. School students, US College students.
Interest	Trainee perspectives, views and opinions on assessment messages received from clinical performance-based assessment formats in actual use for assessment e.g. WBA, ITA. Satisfaction only insufficient. Minimum of approx. half of focus of paper is about trainees.	Studies concerning other assessments (e.g. assignments, written exams, standardised oral exams including OSCEs, short and long cases). Assessor opinions on clinical performance-based assessment formats. Trainee opinion is less than half of the focus of the study. Pilot assessment.
Trial design	Data is reported.	Literature reviews, commentaries, letters to editors, non-peer-reviewed works. Non-English.
Publications	Peer-reviewed journals.	Other literature.

Figure 1: PRISMA diagram for systematic review question – “What are medical specialist trainees’ perspectives of the assessment messages they receive in clinical performance assessments?”

