Year: 2023

Mobile technologies to support workplace-based assessment for entrustment decisions: Guidelines for programs and educators: AMEE Guide No. 154

Marty, Adrian Philipp ; Linsenmeyer, Machelle ; George, Brian ; Young, John Q ; Breckwoldt, Jan ; ten Cate, Olle

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

AMEE GUIDE

🔓 OPEN ACCESS · Check for updates

Mobile technologies to support workplace-based assessment for entrustment decisions: Guidelines for programs and educators: AMEE Guide No. 154

Adrian Philipp Marty[a] (iD), Machelle Linsenmeyer[b], Brian George[c] (iD), John Q. Young[d] (iD), Jan Breckwoldt[e] (iD) and Olle ten Cate[f] (iD)

[a]Institute of Anesthesia, University Hospital Zurich, Zurich, Switzerland; [b]West Virginia School of Osteopathic Medicine, Lewisburg, WV, United States of America; [c]Surgery and Learning Health Sciences, University of Michigan, Ann Arbor, Michigan, United States of America; [d]Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell &, Zucker Hillside Hospital, NY, United States of America; [e]Institute of Anesthesia at the University Hospital Zurich, Switzerland; [f]Utrecht Center for Research and Development of Health Professions Education at UMC Utrecht, The Netherlands

ABSTRACT

With the rise of competency-based medical education and workplace-based assessment (WBA) since the turn of the century, much has been written about methods of assessment. Direct observation and other sources of information have become standard in many clinical programs. Entrustable professional activities (EPAs) have also become a central focus of assessment in the clinical workplace. Paper and pencil (one of the earliest mobile technologies!) to document observations have become almost obsolete with the advent of digital technology. Typically, clinical supervisors are asked to document assessment ratings using forms on computers. However, accessing these forms can be cumbersome and is not easily integrated into existing clinical workflows. With a call for more frequent documentation, this practice is hardly sustainable, and mobile technology is quickly becoming indispensable. Documentation of learner performance at *the point of care* merges WBA with patient care and WBA increasingly uses smartphone applications for this purpose.

This AMEE Guide was developed to support institutions and programs who wish to use mobile technology to implement EPA-based assessment and, more generally, any type of workplace-based assessment. It covers backgrounds of WBA, EPAs and entrustment decision-making, provides guidance for choosing or developing mobile technology, discusses challenges and describes best practices.

Introduction

Workplace-based assessment has become an increasingly recognized, important component of medical education. When George Miller presented his iconic four-layered Pyramid in 1990, he elaborated on the assessment of knowledge ("Knows"), applied knowledge ('Knows How') and standardized performance (" Shows How") but did not discuss the assessment in the workplace, the "Does" level (Miller 1990). Competency-based medical education (CBME) has propelled the development of Miller's "Does"-level workplace-based assessment methods, and, more recently, so has the movement of programmatic assessment. Now, more than 30 years after Miller's Pyramid, many WBA tools have been developed (Norcini and Burch 2007; Holmboe et al. 2018). These reflect major advances, but also show that their development is not finished (Lurie 2012; Massie and Ali 2016). There is, for example, increasing discussion on whether and how high-stakes decisions about learner progression could be based on multiple lower-stakes assessment data, aggregated to support the validity of such decisions (Van Der Vleuten et al. 2015; Hauer et al. 2018). Another new development is a shift to emphasizing narrative feedback as a rich information source (ten Cate

Practice points

- Mobile technology (MT) is rapidly becoming a standard in workplace-based assessment (WBA).
- MT using smartphones has the potential to align WBA with clinical work in teaching hospitals causing only limited disruption.
- Mobile applications can be bought or created. We provide a checklist of considerations.
- Issues to consider include finances, infrastructure, privacy, ethics, data access and ownership, device sanitation.

and Regehr 2019; Ginsburg et al. 2021). Finally, the use of entrustable professional activities (EPAs), units of professional practice for which learners must become qualified, has turned the attention to entrustment decision-making (ten Cate et al. 2016). This implies a shift in thinking from assessing what and how well a trainee has "done" (in Millers terminology) to estimating a learner's readiness to assume clinical responsibilities in the near future (ten

---

Cate 2016). It has been suggested that this prospective view may align with how clinicians think (Crossley et al. 2011; Weller et al. 2014), and improve the quality of assessments (Williams et al. 2021), while we acknowledge that more research must be done in this area. Entrustment decision making requires direct observation and use of construct-aligned scales, among other features, that may improve the quality of assessments.

One challenging aspect of workplace-based assessment is the administrative effort associated with frequent observations (Cheung et al. 2019; Young et al. 2020a). This is particularly important in the current era of CBME, where frequent assessment is viewed by many as a best practice (Cheung et al. 2021; Williams et al. 2021). Within this framework, trainees are observed and evaluated to enable a determination of their readiness for unsupervised practice. This clearly requires effort (Cheung et al. 2021). At times, faculty may perceive this effort as a burden that distracts from teaching and taking care of patients. Similarly, trainees are often expected to request assessments and can view this work as yet another administrative burden with limited value (Schut et al. 2021).

To minimize burden and optimize value, the best way to collect workplace-based assessment ratings is to use information that is already available from natural encounters with trainees and to minimize the additional effort needed to document that information. Here are where mobile applications on smartphones are most useful. Smartphones have the benefit of being widely available, and make it easy for raters to record their impressions and feedback when it is most valuable—immediately after an observation (Bohnen et al. 2016; Young and McClure 2020; Marty et al. 2022). This value increases when that assessment data is automatically added to a trainee's e-portfolio (Williams et al. 2014).

While reviews have been written about the use of mobile technologies in clinical education in general (Nikou and Economides 2018; Maudsley et al. 2019), this AMEE Guide provides practical guidance for the development and use of mobile technology for the assessment using EPAs. The purpose of the Guide is to provide support for choosing or developing an application, discuss implementation challenges and highlight best practices. We will touch on training, support, logistics, data analysis and reporting, and evaluation for improvement.

Workplace-based assessment to support entrustment decision making

The purpose of using EPAs in medical training is to build individualized curricula for all learners, which guides them from legitimate peripheral participation in healthcare teams to full participation, with appropriate autonomy. The graduation is determined by the EPAs for which a learner is truly qualified as a practitioner, first under supervision, later unsupervised. For each EPA a surrogate "license-to-practice" qualification is needed, sometimes called a "star" or *statement of awarded responsibility* (ten Cate and Scheele 2007). Because of varieties of rotational experiences and personal capabilities, trainees will master EPAs at different moments in time. In addition, within specialties, training programs may differ in the scope of EPAs they can

offer, e.g. smaller programs may offer only core EPAs and larger programs offer also elective EPAs (Kaur and Taylor 2022). To learn the full range of professional practice, trainees usually rotate to different training sites. Each specialty needs to define which level of autonomy for which EPA is sufficient at the end of the training to become a specialist. For undergraduate medical education, EPAs have been defined nationally in some countries.

To support valid summative entrustment decisions to qualify learners for EPAs, adequate workplace-based assessment information is needed (Touchie et al. 2021).

Sources of information to support entrustment decisions

Workplace-based assessment has numerous forms and tools, but for practical purposes we defined four sources of information, relevant to these types of assessments. They include (a) brief, direct observations, (b) longitudinal monitoring, (c) individual discussions and (d) product evaluations. These sources, in which almost all existing WBA methods can be subsumed, have been proposed to support entrustment decisions about EPAs (ten Cate et al. 2015; Chen and ten Cate 2018). Together they provide a rich array of information that can strengthen the validity of summative entrustment decisions, and the recommendation is to include, if possible, information from all sources because they focus on different aspects of competence. Together they may constitute the basis for a program of assessment (van der Vleuten et al. 2020).

Brief, direct observations. include the observation of an activity (e.g. an EPA) typically for 5-15 min, and a subsequent feedback conversation with the trainee. Many examples exist, including the well-known mini-CEX method (Kogan et al. 2017). These activities may be clinical encounters with history and physical examination, or procedures, called DOPS (direct observation of procedural skills) (Norcini and Burch 2007). More recently "field notes" have been used to document these direct observations (Donoff 2009). The result or evaluative conclusion can generally be captured with mobile devices such as smartphones with an appropriate application and voice recognition software (Young et al. 2020). These observations are EPA-specific.

Longitudinal monitoring regards impression formation over time. Repeated direct observations within a longitudinal supervisory relationship are one such example (Young et al. 2020b). Similarly, Multisource feedback (MSF) or 360-degree evaluations, in which various colleagues are asked to report their observations and experiences during a period of time (e.g. from a weekend shift to multiple consecutive weeks) is a suitable approach, applied with a frequency of one or more MSF sequences per year, while approaches vary (Lockyer 2013). MSF can be easily automated (Alofs et al. 2015) and several e-portfolio systems have incorporated it. Longitudinal monitoring is particularly useful to evaluate behavior that is not easily captured in a focused and often planned moment of direct observation, such as a holistic impression of professional behavior and the learner features known to support entrustment decisions (reliability, integrity, humility, agency, capability) (ten Cate and Chen 2020). MSF is best used to inform professional growth.

Individual discussions include case-based discussions (CBD) and chart stimulated recall (CSR) sessions, with or without reference to electronic health record case data. These are sometimes considered mini oral exams, focused on knowledge and clinical reasoning (Holmboe et al. 2018; Yudkowsky et al. 2020). More detailed and often briefer examples are the one-minute-preceptor and SNAPPS methods (Holmboe et al. 2018). One specific approach is the entrustment-based discussion (EBD) technique, designed to support the estimation of risks when making entrustment decisions (ten Cate and Hoff 2017). A mobile device, as with direct observations, can be designed to capture the conclusion of the encounter: is this student ready for more autonomy?

Product evaluation pertains to the outcome of performance. The tangible outcomes of practicing patient care are arguably the most relevant, but also most difficult to measure, as many variables affect the quality of care provided to patients. But some more proximal products may be assessed. Entries and decisional choices in electronic health records can be evaluated, patient satisfaction with care can be solicited, the quality of artifacts such as in dentistry and surgery can be judged. Patient outcomes influenced in a meaningful way by trainees can be abstracted from electronic health records. All those "products" of the trainee that can be assessed without the direct presence of the trainee as subsumed in this category.

The central question for workplace-based assessment using EPAs is how to prepare a decision for a decreased level of supervision, i.e. increased autonomy. Such decisions should be informed by assessment information. For that reason, it is wise to force clinicians to frame their judgment as a recommendation for readiness for autonomy, such as " based on my observations, I judge that this trainee, for this EPA, is ready to act under indirect supervision in similar future cases", or " … act with clinical oversight only … ". To support the judgment, a narrative evaluation can be added, that can also serve as feedback to the learner. In addition, a valuation of relevant competencies can be added, and a note or score about the complexity of the observed case, and the certainty of the judgment. Figure 1 shows one example (of many possibilities) for a scoring format that can be transposed onto a mobile device (ten Cate et al. 2015).

Several entrustment-supervision scales exist to capture a level of supervision or guidance, either as a readiness recommendation for the future (i.e. prospectively) or as an experience, reflecting the support that was actually provided during an activity (i.e. retrospectively) (ten Cate et al. 2020).

Ad hoc and summative entrustment decisions

A core feature of workplace learning is that trainees are regularly entrusted with clinical activities by a supervisor on an ad-hoc basis ('please go see that new patient and let me know what you think'). Autonomy to speak with and examine the patients is crucial for building experience. Those ad hoc moments, directly observed or evaluated afterwards (Landreville et al. 2022), are a rich source of information to support summative entrustment decisions. Summative decisions are made by teams (clinical competency committees or similar entities), based on multiple observations and information bits (Lockyer et al. 2017). A summative decision can result in promotion, graduation, or a STAR (statement of awarded responsibility), comparable to a micro-credential (Norcini 2020), i.e. a formal qualification to act. Much of this process can be automated using e-portfolios and mobile devices. Figure 2 shows a possible flow of data to support the awarding of STARs.

Challenges in workplace-based assessment

Irrespective of the technologies used, workplace-based assessment has well-documented challenges and limitations. Many variables affect scores given based on observations, and psychometric studies have shown that the variance in WBA scores is often determined more by raters, context and error than by trainee differences (Weller et al. 2017). CBME, with its aim to graduate trainees only if they meet standards, has reinforced the need for more valid assessment procedures (Gruppen et al. 2018). To increase the rigor of WBA, several routes have been suggested.

First, supervisors and trainees often are inadequately trained in WBAs, do not understand the purpose, and interact in settings with insufficient time for direct observation and/or feedback conversations (Massie and Ali 2016). Feedback is frequently inadequate, undermining its credibility to the trainee (Holmboe et al. 2011; Telio et al. 2016; Sukhera et al. 2019). Next, many supervisory relationships are brief and change frequently, impeding the development of strong educational alliances including trust (Telio et al. 2016). Finally, trainees often perceive assessment as summative even when intended as formative (Schut et al. 2018, 2021) and adopt behavior just to conform to what they perceive to be desired (Watling et al. 2016). The direct observation can morph into a performance that feels inauthentic, affecting receptivity to feedback (e.g. "*that is not what I normally do anyways*" (LaDonna et al. 2017)). As a result, faculty and residents can develop a negative view of direct observation and structured feedback, as "tick-box" or 'jump through the hoops' exercises, leading to a trivialization of WBA (Young et al. 2020a). This led Young et al.

EPA: *Guiding, informing, and advising patients and families* From my observations, I believe this learner is now ready to:			
• Act under **direct supervision** (supervisor present)			O
• Act under **indirect supervision** (supervisor immediately available if needed)			O
• Act with **distant supervision** (supervisor not immediately available)			O
Case complexity:	low	medium	high
My feedback regards:	Tips	No comment	Tops
Medical expertise	O	O	O
Communication	O	O	O
Collaboration	O	O	O
Health advocacy	O	O	O
Scholar role	O	O	O
Professionalism	O	O	O
Feedback:			

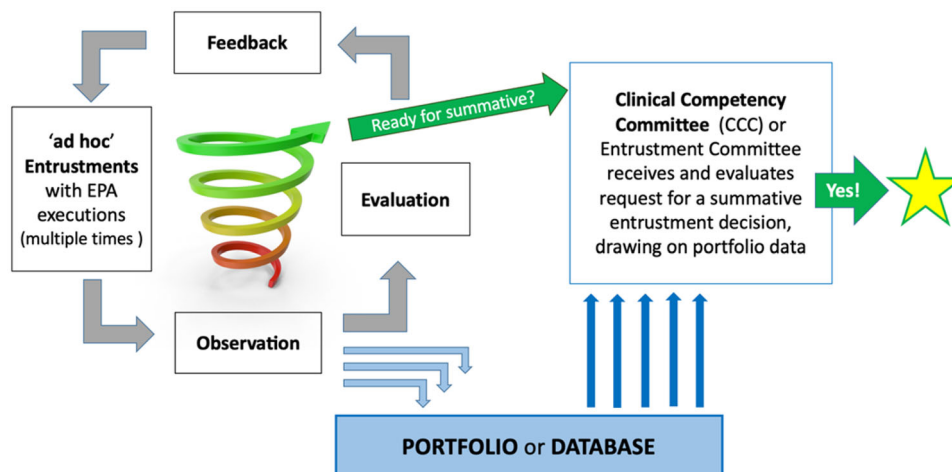Figure 1. Example of a generic scoring form for EPA-based assessments.

Figure 2. General flow of workplace-based assessment data to support summative entrustment decisions making.

Box 1. Key Features of Successful Direct Observation and Structured Feedback Programs.

1. Provide regular, ongoing training in three areas:
   a. Direct observation, including how to support trainee autonomy while observing.
   b. Use of the chosen WBAs, with attention to performance dimensions, frame of reference, and narrative comments.
   c. Feedback as a bi-directional, co-constructed conversation with an emphasis on open ended higher order questions and facilitated listening, encouragement, and agreeing on an action plan.
2. Design clinical rotations to create longitudinal supervisor-trainee relationships.
3. Repeated, frequent observations that become part of the culture.
   Ideally, direct observation is more frequent earlier in training and then tapers off (but does not stop) as a trainee approaches readiness for independent practice.
4. Protected time for faculty to observe and for faculty and trainee to engage in feedback.
5. Utilize, whenever possible, structured observation tools with evidence for validity.
6. Monitor adherence or engagement by faculty and trainees (e.g. number of direct observation assessments completed by each faculty or trainee per rotation or clinic per month).

(Young et al. 2020a; Young et al. 2020) to define enablers and barriers when using a mobile app to capture EPA assessment data, and stressing the need for training of faculty in observation, feedback and the use of technology; the need for rotations long enough to allow for longitudinal relationship building and repeated observations; and educating trainees to engage actively in self-development based on trusted observations and credible feedback (Box 1).

The role of e-portfolios and mobile technology in WBA for EPAs

Mobile technologies cannot overcome all challenges of implementing WBA, but they can serve as a key resource given (a) the strong recommendations from programmatic assessment theorists to collect abundant data, (b) the ubiquitous availability of mobile technology and (c) the opportunity and need to learn from recent education research when applying such innovations.

WBA data collected with mobile technology can be aggregated within a centralized database. Centralized data storage provides one location where learning analytics can be used to make inferences about trainee progress and eligibility for STARs. Assessment data must also be made available to (i) individual trainees, ideally as part of an e-portfolio and (ii) to institutional entities, such as program directors, clinical competency committees, and institutional administrators. These uses of data are most easily supported as part of a dashboard. Data safety, legal and ethical aspects play an important role (see below).

What should or can mobile applications capture?

Figure 1 provided a simplified image of one possible screen design. To generalize, elements that need to be captured and elements that are recommended are shown in Box 2.

The user of the app can be a clinical teacher/assessor, the learner, or both. In the first case, the clinician should have authorization to access the database of relevant learners. In the second case, the learner should have authorization to access their own file. There are multiple options for interactions between the assessor and learner:

- Supervisor assessment. The assessor may complete the assessment at their own time on their own device (authorized by the institution) and submit feedback to the learner to read later.
- Provisional self-assessment. The learner might self-evaluate then send a report to a supervising clinician for confirmation.

Box 2. Elements to be captured during a rating process using a mobile device.

| Required elements	Recommended elements
• Learner name or ID	• Level of complexity of case or situation
• Rater name or ID	• Relevant domains of competence for feedback, if defined for the EPA
• Date/time of the assessment	• Space for written, or, preferably, recorded feedback
• EPA of interest	• Optional explanatory texts (about the EPA, about supervision levels, about criteria)
• Specialty, rotation, clerkship	
• Level of supervision provided and/or recommended of ES-scale	
• Narrative feedback or learning goal	
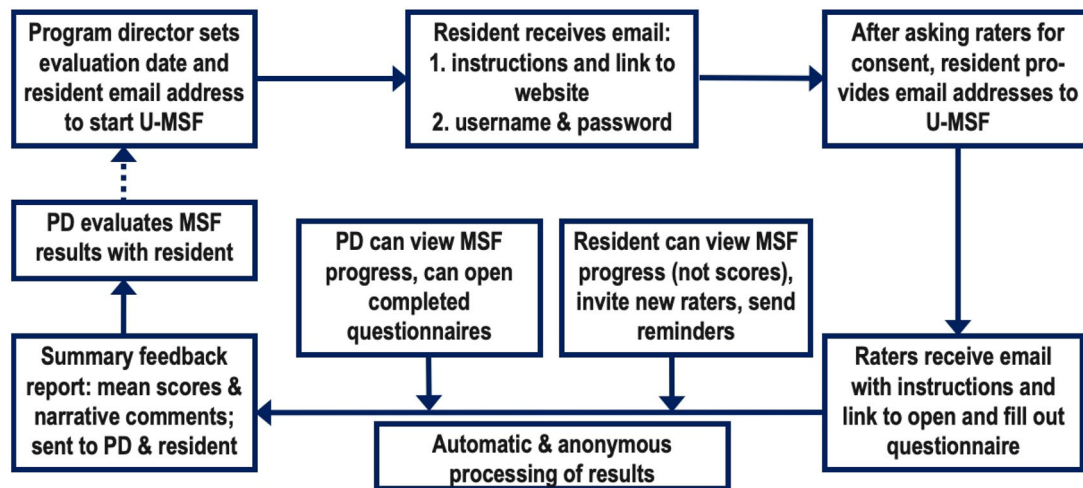
Figure 3. Procedure and data flow in one MSF procedure, annually or semi-annually performed.

- Learner initiated direct access and assessment. The learner may provide a personalized QR code or pull up the assessment directly for EPA assessments on their device, ready for immediate input or scannable for any clinician who then has direct access to the assessment form of interest to complete/submit and discuss with the trainee.
- Simultaneous self- and supervisor assessment. Assessors and learners may both simultaneously complete an assessment and compare immediately, triggering a feedback discussion. Learners may optionally also evaluate a supervisor.

All these modalities have been developed and tested in practice (George et al. 2020; Young and McClure 2020; Marty et al. 2022).

Assessments using mobile apps can be tracked not only by single patient encounters, but also by longitudinal monitoring. Direct observation of multiple patient encounters by the same supervisor over days to weeks to months or years can help identify enduring patterns of strength and struggle, especially when captured electronically and visualized (Young et al. 2020), and generates rich narrative feedback that learners deem credible. Multisource feedback (MSF), or 360-degree evaluation, can be highly automated. As in the case of longitudinal direct observation, raters are asked to observe learners over time (a shift, a week or longer) and then provide structured feedback, often with a focus on more general competencies, rather than specific single activities. Yet, these more general features may be critical for summative entrustment decisions. An example of the data flow of an MSF procedure, developed in Utrecht (U-MSF) and used for many years, is depicted in Figure 3 (Alofs et al. 2015). While fully online, the use of mobile devices for non-direct assessment is less critical than with brief direct observations.

What should or can be included and displayed in an e-portfolio?

A portfolio can be described as "a collection of samples of a person's work, typically intended to convey the quality and breadth of his or her achievement in a particular field"(Oxford English Dictionary, no date). Before the internet era, portfolios in medical education were proposed as an attempt to counteract the limitations of a reductionist approach to assessment to facilitate the assessment of integrated and complex abilities and take account of the level and context of learning. Through documentation, reflection, evaluation, defense, and decision it provides an assessment solution for a curriculum that designs learning towards broad educational and professional outcomes. It personalizes the assessment process while incorporating important educational values. It supports the important principle of "Learning Through Assessment" (David et al. 2001). In workplace-based learning, portfolios reflect the individual's progress and accomplishments but may include additional personal information and reflections that both serve for educational personal development as well as for assessment. In the current day and age, physical portfolios have been replaced by e-portfolios. The data in a learner's e-portfolio can thus be equated with a dashboard of items used for various purposes. The database may be selectively

Portfolio of: trainee Jones	PGY1		PGY2		PGY3		PGY4	
EPA a	1	2	2	2	3	4	4	5
EPA b	1	1	2	2	2	3	3	4
EPA c	2	2	3	4	5	5	5	5
EPA d	2	3	4	4	4	4	5	5

Figure 4. Schematic excerpt from an individualized portfolio, showing expected increase in autonomy for various EPAs across four postgraduate years (ten Cate 2014) (reprinted with permission).

accessible for various parties (learners, teachers, administration, clinical competency committees) and used or optimized as dashboards for different purposes (Oudkerk Pool et al. 2018).

Individual EPA observations, entrustment-based discussions, longitudinal observations have been described from the input side. Learners can benefit from reviewing the feedback of individual assessments from the output side but they seem to be less interested in feedback received longer ago (Young et al. 2020a). However, for summative decision-making about a learner's increased autonomy, overseeing the learners' accomplishments as reported by various observers is crucial and must be aggregated in some way. Clinical competency committees, with limited time for each learner, must avail of maximally rich and visually intuitive data displays. Data may be displayed in different ways – longitudinally and or as spider graphs for individual learners and against aggregated group data or historical reference data. Examples can be found in various publications (Warm et al. 2014, 2016; van der Schaaf et al. 2017; Kamp et al. 2021; Hanson et al. 2022).

Such dashboards may serve different proposes. One is decision-making in clinical competency committees. Another is to provide targeted, aggregated feedback to learners. A third purpose is to display an overview of entrustment decisions which shows the specific levels of supervision for each EPA under which a learner is qualified to act (Figure 4).

One step further is to create digital badges for EPAs, reflecting STARs, to show the scope of practice from which a (usually senior) learner is qualified at which level of supervision (ten Cate 2022), for any relevant stakeholder, such as clinical staff and nursing staff. This reflects a form or micro-credentialling (Norcini 2020).

E-portfolios as the database of learner progress may also serve as reports for accreditation bodies.

Critical issues to consider when reviewing or designing mobile technologies

Choose or design?

Designing an app sounds attractive to individuals with experience in creating apps. There are pros and cons to designing versus picking a product developed by others. Key considerations in these discussions typically fall under three areas: Cost, Quality and Control.

Cost. The cost of building a workplace-based assessment system includes the initial cost of design and development as well as the ongoing cost of maintenance. The cost of design and development depends critically on the scope of features and how integrated the system will be with other existing software systems. Additionally, the cost to maintain the software is often overlooked. Libraries and operating systems get updated in ways that can render software non-functional, security threats evolve and may require software or server hardware updates, and the context within which the software is implemented can require changes to existing features. Therefore, developers of an in-house system will likely need to carefully consider the scope of features they would like to design, build, and maintain. Depending on the context, costs for acquiring smartphones and internet access data packages for individual users may need to be considered.

Quality. The quality of the final software package is a combination of technical features, user experience (how easy the features can be found and used), and support. High-quality user experience requires developers (whether in-house or commercially) to thoroughly understand the needs of the software users requiring close interaction with key stakeholders. The intensity of interaction depends on the complexity and scope of the implementation as well as any specific needs of users. Regular consultation with different stakeholders is highly recommended.

Control. The final element to be considered control. Who should determine which features will and will not be developed in the future? Are there foreseeable circumstances that will necessitate critical customizations of the software? In 2018, essential EU data protection regulations outdated many software packages (NN, 2018). Likewise, a hospital firewall requiring additional authentication might be an important consideration in evaluating whether existing systems could be adapted or if a custom, in-house system would need to be developed. In a similar vein, if new analyses or improvements appear to be useful, an institution should be able to exert control over software development.

Considerations prior to review or design of a tool

In the coming sections we will discuss a variety of specific elements worth considering prior, during and following reviewing or designing a tool. These are compiled as a checklist in Table 1.

Purpose. If the only purpose is to document infrequent formative feedback, an assessment system can be quite simple (Google Forms or other survey tools). On the other hand, if assessment data is captured frequently and is also

Table 1. A Checklist of considerations for review when choosing or designing WBA mobile technology.

Considerations Prior To Review or Design

Purpose	• Will the assessments be used formatively (low-stakes), summatively (high-stakes), or both?
	• Will qualitative comments/feedback be captured? Will growth from feedback be needed/tracked?
	• Will quantitative data be collected only? Will progression toward a specific outcome be tracked?
	• Will data be linked to a dashboard containing data from other sources for an overall snapshot of a trainee?
	• Will there be a need for trainees to reflect on their performance?
	• Will the assessment be top-down only or also bottom-up (evaluation of teaching interaction)?
	• Will digital badging or certification processes be needed?
	• How important is compatibility with other systems for collecting and aggregating assessment data?
Context	• Will the WBA technology be used in undergraduate medical education, graduate medical education, other educational settings, or a combination (e.g. longitudinal assessment strategy)?
	• Will the WBA technology be used at a single institution, multiple institutions, or as part of larger national or international program?
	• Will the assessments be captured predominantly in a busy clinical setting with high patient traffic, slower clinical setting, or varied settings?
	• Will the assessments be captured in person or remotely?
	• Will the assessments be captured real-time in the clinical setting, or will it be asynchronous and postponed?
	• What network infrastructure and capabilities are available?
	• Is offline capture available? Is the upload of offline data manual or automatic?
	• What technical aspects need to be covered: e.g. iOS vs. Android or both operating systems, only mobile or also web-version? What web browser(s) are supported?
	• Will the users use their own device, or will the training program provide devices?
	• What are the data requirements as WIFI may not be free/accessible in all clinical settings?
	• What institutional resources, personnel, and budget is available?
Users	• Who will be the users (e.g. students, residents, preceptor, master assessor, clinical director, other staff as part of the healthcare team, etc.)?
	• Who will be assessors? Only supervisors or will trainees be able to self-assess their performance? Combination of both?
	• Who will be reviewing/accessing the data or reports from the data (e.g. Clinical Competency Committee)?
	• National institutions (e.g. national boards of examiners)

Considerations During the Review or Design

User Interface	• Will the technology be offered as a mobile application only? Will users be able to enter data or access information through a desktop version? Will mobile or desktop versions be different for each stakeholder (e.g. trainees – mobile only and evaluators- desktop or mobile)?
System Adaptability	• Can multiple supervision scales be used (e.g. ability to use ten Cate scale, Chen scale , Ottawa scale, Zwisch scale, etc.(ten Cate et al. 2020))?
	• Does the system support instant additions/modifications (pushed immediately to users) to assessment elements, scales, descriptors, etc.?
	• Can the branding be changed to support institutional branding? Can the system be modified with the look and feel of the institution?
User Experience	• How many total screens will the user have to navigate for main functionality ("number of clicks")? How intuitive is the workflow?
	• Do open comment areas allow dictation and speech-to-text functionality (e.g. through a microphone built into the system or the microphone on the mobile device keyboard)? How much time do users have to spend in documenting WBAs (e.g. 30 s)?
	• Can students share data easily with others (e.g. assessors)? How is the data shared (e.g. email, Quick Response (QR) code, etc.)?
	• Can students easily see data submitted about them? Can assessors share data about a student with another assessor or staff person)? Can stakeholders see this data through the mobile application, browser, or both?
	• Are further definitions of terms or scale attributes readily available through dropdowns or pop-outs?
Security and Validation	• What will be required for authentication? Will integration with an institutional authentication system be needed? Will the authentication be self-contained as part of the system? Will authentication synchronization be available? Will import or bulk uploads be available? Can users be added on the fly?
	• What will be the method of validation (if any) of the assessor? For example, signature, picture of badge, QR code, finger scan, etc.
	• Will the assessor be entering an assessment on the trainee's device, their own device, a device provided by the institution, or other method?
	• Who will own the data? The trainee? The institution? National authorities?
	• Where will the data be stored? On the smartphone? Local server? In the cloud?
Reporting	• What reports will be needed (e.g. total EPAs completed by rotation, average growth curve by EPA, EPA Submissions, etc.)?
	• What filters will be needed for each report (e.g. trainee, assessor, course/rotation, EPA, formative versus summative, supervision scale, case complexity, elements of trustworthiness, start and end dates, etc.)?
	• Will open comments be exported verbatim or will a report with thematic analysis be needed?
	• What formats for data exportation will be needed (e.g. xls, cvs, pdf, doc, etc.)?
	• Will connection to another reporting database/portfolio be needed?
Other considerations	• Can the trainee complete a self-assessment?
	• What other elements can be assessed (e.g. elements of trustworthiness, case complexity, etc.) that might inform entrustment decisions?
	• Does the application need or have other uses (e.g. documentation of case or procedure logs, view of course/rotation schedules, multi-source feedback, other assessments, etc.)?

Considerations Following the Review or Design

Timing and selection of assessments	• How often will assessments be collected?
	• Do users select the elements to be assessed or are the required elements outlined in a syllabus or other formal document?
	• How often is feedback given/communicated? Is there a set meeting time for feedback (e.g. quarterly) or is feedback given on the fly as it is be entered or dictated into the system?
Training	• How will training be conducted? Who will provide the training? Who will need to be trained (e.g. student/assessor/clinical director/ administrator)? What type of model will be used (e.g. direct, scaffolded, train the trainer, etc.)?
	• Will modules need to be included as part of the mobile application?

Support Structure	• What will the support structure look like? Will the support structure be direct, scaffolded, handled by the vendor or a third party, etc.? What will be the mechanism for communicating support (e.g. website, email, link in the mobile application, QR code, combination)? • How is technical support and ongoing updating of the application guaranteed?
Legal Issues	• Are there any legal issues that need to be considered, addressed in policy, discussed for shared understanding? • Are policies in place that outline requirements, expectations, licensing, copyright, etc.?
Evaluation	• How will the implementation of MT be evaluated? (e.g. using frameworks of implementation science)

used for summative entrustment decisions, then a more sophisticated implementation is warranted using mobile technology. At the highest level of complexity, customized software will be needed if it is to interface with other IT-systems, such an institutional learning management system, a university grade or progress registration system, a national e-portfolio system, accreditation software.

Context. If technology is used in undergraduate medical education, it needs to perform in very different settings since the students usually attend courses or clerkships at different locations with many different supervisors. In postgraduate settings, the workflows are more stable. As for platforms, any app should at least support Android and iOS. We recommend that the assessment technology also works on the users' own devices, since many institutions have a bring-your-own-device policy.

Offline capabilities. While a recent report of internet usage indicated there were around five billion internet users around the world, there remains stark differences in internet users according to region (*Number of internet users worldwide as of 2022, by region*, no date; *Countries with the largest digital populations in the world as of January 2022*, no date). For example, Africa and Middle Eastern regions have lower user numbers. Some of the reasons cited for the lower numbers include lack of digital infrastructure, particularly in rural areas, price affordability, barriers for certain sectors (gender-based, urban-rural adoption gaps, people employed in service sectors, etc.) (Ochoa et al. 2021). These issues can manifest themselves in any part of the world. Therefore, ensuring offline capabilities is important to allow users to capture assessment data when internet or network connections are temporarily not available or not reliable.

Users. Digital literacy varies greatly. Therefore, the design and workflow of any assessment app should be intuitive and adaptable enough for all user groups (e.g. font size). It should be possible to assign different roles to the users (e.g. assessor, trainee, administrator, etc.).

Considerations during review or design of a tool

User Interface. The idea of mobile technology is to bring documentation of assessment situations to the point of care. For data capturing, mobile technology is best, while web-based versions should be a fallback option only. For data analytics and dashboards, larger screens are needed (e.g. desktop computers, tablets, or meeting room displays). Programs should select or adapt WBAs, when possible, with evidence for validity. WBAs developed for paper-based or desktop-based systems will need to be adapted to ensure an interface appropriate to the smaller screens of mobile devices (e.g. minimize scrolling, number of screens, required taps).

System Adaptability. Non-technical changes, e.g. in EPA lists, should be manageable in the backend by local developers, program directors or administrators without needing a new app release. Different assessment elements should be developed even within the EPA framework: e.g. the technology should be flexible enough to adopt different scales.

User Experience. Usability is key for adoption. The higher the usability the less effort needs to go into user education or faculty development. Engage stakeholders in the early design and pilot testing to optimize the user-experience. This might include recruiting a few faculty, learners, and administrators to pilot the mobile app and give feedback that improves ease-of-use. This will help facilitate engagement when the mobile app is implemented more broadly.

Security and Validation. Any system must adhere to strict data safety rules (e.g. GDPR, HIPAA). Furthermore, some way to validate assessors is key for data quality. One approach is to require user authentication.

Data Management. In general, data should be managed in a way that promotes education and, ultimately, high-quality patient care. This implies that data should be available to programs, institutions, and, in some cases, larger regulatory bodies. Researchers might also want to access it, to improve the educational system for all learners. However, these interests must be balanced against the interests of the individual trainee, who should have agency in how the data is used. How best to reconcile these competing interests remains an active area of exploration, and in some cases, research ethics committees may need to be involved.

Reporting. Implementing a new mobile assessment technology requires careful consideration regarding which data should be reported or made visible and for whom (trainee, program director, CCC, university, hospital administration, etc.), how long, and how it should be presented. There is no gold standard.

Other considerations. Having the option to also use the mobile application for other functions like direct observation, course confirmations and multisource feedback is recommended.

Considerations following review or design of a tool

Timing and selection of assessments. In general, WBAs should be performed as often as possible, ideally several times per day, and a subset should occur within longitudinal supervisory relationships and clinical attachments. Instead of the required numbers of performed procedures, every training program or national specialty society should define how many *satisfactory* WBAs are required before

entrustment with a patient care EPA is warranted. We recommend being as open as possible, such that every assessment is discussed with the trainee.

Training. Initial training of local champions should be provided by the group or company providing the app. We recommend identifying trainees and supervisors as local champions to act as multipliers for training content and ambassadors for the change process. Suitable resources should be made available: white papers, tutorials, webinars, etc. In addition, training will need to be ongoing to support adoption and to ensure that supervisors and learners apply the tool (e.g. the entrustment scale) in the intended way.

Support Structure. This is the most crucial aspect in regard to the sustainability of a system. Of course, there needs to be a user support facility for practical issues while using the app. It needs to be very responsive to questions and deliver support as quickly as possible. Even more important is ongoing technical support by app developers. On a fundamental level, there should be a constant effort to keep the app updated to newer versions of the operating systems of smartphones (android, iOS). To keep an app attractive and user-friendly, new features or workflows may regularly be added.

Legal, ethical and privacy issues. The collection of data from individuals is not trivial. Complex regulations often hold for research subject data, overseen by ethical review committees; but for education, such rules were, until recently, less strict. That is changing and anyone considering to apply or create software to collect and store personal data should be aware of restrictions (ten Cate et al. 2020). Since 2018, Europe has had strict rules, and high fines for breaches, to protect the privacy of individual citizens. European initiatives to collect data must abide by these general data protection regulations (GDPR) (NN, 2018), but even companies outside Europe must follow these rules too if EU citizens' data are involved. The rules are summarized in Box 3.

Disinfecting/Sanitizing. There may be concerns regarding the sanitation of mobile devices in clinical settings. Institutions should be clear with rules and guidance on cleaning and disinfecting electronics. The US Centers for Disease Control has guidelines specific to cleaning electronics (*Cleaning and Disinfecting Your Facility* 2021) while most popular mobile phone manufacturers also provide instructions for cleaning and disinfecting devices. Institutions are encouraged to explore and adopt guidance that will address concerns.

Ongoing Improvement. As with any intervention, ongoing evaluation and improvement should be part of the implementation. Those efforts should focus on "what works where and why?" and how the system should be adapted for the future.

Discussion and conclusion

Clearly, the implementation of CBME is a complex and challenging change process. At the core, it involves a culture change towards valuing frequent workplace-based assessment and feedback conversations. Mobile technology can be viewed as an important catalyst for this process by reducing many of the perceived barriers.

Having said that, workplace-based assessment data collection can feel as a burden for clinicians (Cheung et al. 2021) and trainees (Ott et al. 2022). Data must only be collected if it adds to relevant decisions about learners. It will be useful to design studies that focus on the optimum of data quantity and show when more data, and which data, are redundant. Data gathered may also be deleted as soon as storage is no longer relevant. Learners should be able to make mistakes and perform sub-optimally. Documented failures however should not haunt learners any longer than necessary (ten Cate et al. 2020).

One other aspect of collecting assessment data on a large scale is the fear of building or at least adding to a surveillance society. As with any process or system, it is important to remember its purpose. In the realm of medical education, the only purpose of any system of assessment is to support trainees to become competent professionals. When data collection leads to unintended usages, breaches of ethical principles are at stake. As said earlier in this Guide, the ethics or scientific research using human beings is well covered with appropriate rules; the ethics of education has no universal dedicated rules. But data for education can be misused too. It is wise for institutions to create such ethical rules, and perhaps even on a national level; not as a bureaucratic layer of administration but to protect the user who demands fair and careful data management.

Mobile technology should make life easier, more enjoyable, comfortable and efficient, and provide valid information to decide upon readiness for autonomy in patient care, and hence for progression in a training trajectory.

The authors acknowledge that in many contexts, the use of mobile technology for assessment, and in fact often any workplace-based assessment in general, is completely new. Hospitals and programs have done without for decades, and such innovation requires more than instruction with new tools. For many educators, competency-based education and workplace-based assessment may feel as a paradigm shift, that would require transformative learning (Mezirow 2009). It is good to realize that all we have

---

Box 3. The 7 principles of the European Union General Data Protection Regulations (NN, 2018).

1. Lawfulness, fairness, and transparency	Processing must be lawful, fair, and transparent to the data subject
2. Purpose limitation	You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
3. Data minimization	You should collect and process only as much data as absolutely necessary for the purposes specified
4. Accuracy	You must keep personal data accurate and up to date
5. Storage limitation	You may only store personally identifying data for as long as necessary for the specified purpose
6. Integrity and confidentiality	Processing must ensure appropriate security, integrity, and confidentiality (e.g. by using encryption)
7. Accountability	The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles

discussed must be embedded in a culture that is ready for this change.

Disclosure statement

- Adrian Philipp Marty is president of the board and chief visionary officer of precision ED Ltd (Wollerau, Switzerland), a company founded to sustainably improve and distribute *preparedEPA*.
- Brian George is the Executive Director of the Society for Improving Medical Professional Learning (SIMPL), a non-profit collaborative network of training programs.

Notes on contributors

Adrian Philipp Marty, MD MME, is a senior attending physician at the Institute of Anesthesia at the University Hospital Zurich, Switzerland.

Machelle Linsenmeyer, Ed.D, is Associate Dean for Assessment and Educational Development and Professor of Clinical Sciences at West Virginia School of Osteopathic Medicine, Lewisburg, WV, United States of America.

Brian George, MD MAEd is an Associate Professor of Surgery and Learning Health Sciences at the University of Michigan, Ann Arbor, Michigan, United States of America.

John Q. Young, MD, MPP, PhD, is Professor and Chair of the Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell & Zucker Hillside Hospital, NY, United States of America.

Jan Breckwoldt, MD, MME, FERC, is a senior attending physician at the Institute of Anesthesia at the University Hospital Zurich, Switzerland.

Olle ten Cate, PhD, is a professor of medical education and senior scientist at the Utrecht Center for Research and Development of Health Professions Education at UMC Utrecht, The Netherlands..

ORCID

Adrian Philipp Marty http://orcid.org/0000-0003-3452-9730
Brian George http://orcid.org/0000-0002-9404-5255
John Q. Young http://orcid.org/0000-0003-2219-5657
Jan Breckwoldt http://orcid.org/0000-0003-1716-1970
Olle ten Cate http://orcid.org/0000-0002-6379-8780

References

Alofs L, Huiskes J, Heineman MJ, Buis C, Horsman M, van der Plank L, ten Cate O. 2015. User reception of a simple online multisource feedback tool for residents. Perspect Med Educ. 4(2):57–65.

Bohnen JD, George BC, Williams RG, Schuller MC, DaRosa DA, Torbeck L, Mullen JT, Meyerson SL, Auyang ED, Chipman JG, et al. 2016. The feasibility of real-time intraoperative performance assessment with simpl (system for improving and measuring procedural learning): early experience from a multi-institutional trial. J Surg Educ. 73(6):e118–e130.

Chen HC, ten Cate O. 2018. Assessment through entrustable professional activities. In: Delany, C., Molloy E., editors. Learning &

Teaching in Clinical Contexts: A practical guide. Chatswood: Elsevier Australia, p. 286–304.

Cheung K, Rogoza C, Chung AD, Kwan BYM. 2021. Analyzing the administrative burden of competency-based medical education. Can Assoc Radiol J. 73(2):299–304.

Cheung WJ, Patey AM, Frank JR, Mackay M, Boet S. 2019. Barriers and enablers to direct observation of trainees' clinical performance: a qualitative study using the theoretical domains framework. Acad Med. 94(1):101–114.

Cleaning and Disinfecting Your Facility. 2021. Centers for Disease Control. https://www.cdc.gov/coronavirus/2019-ncov/community/disinfecting-building-facility.html [accessed 2022 Oct 25].

Countries with the largest digital populations in the world as of January 2022. no date. Statista. [accessed 2022 Oct 25]. https://www.statista.com/statistics/262966/number-of-internet-users-in-selected-countries/

Crossley J, Johnson G, Booth J, Wade W. 2011. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. Med Educ. 45(6):560–569.

David MFB, Davis MH, Harden RM, Howie PW, Ker J, Pippard MJ. 2001. AMEE medical education Guide No. 24: portfolios as a method of student assessment. Medical Teacher. 23(6):535–551.

Donoff MG. 2009. Field notes. Can Fam Physician. 55(12):1260–1262.

George BC, Bohnen JD, Schuller MC, Fryer JP. 2020. Using smartphones for trainee performance assessment: a SIMPL case study. Surgery. 167(6):903–906.

Ginsburg S, Watling CJ, Schumacher DJ, Gingerich A, Hatala R. 2021. Numbers encapsulate, words elaborate: toward the best use of comments for assessment and feedback on entrustment ratings. Acad Med. 96(7S):S81–S86.

Gruppen LD, Ten Cate O, Lingard LA, Teunissen PW, Kogan JR. 2018. Enhanced requirements for assessment in a competency-based, time-variable medical education system. Acad Med. 93(3S Competency-Based, Time-Variable Education in the Health Professions):S17–S21.

Hanson MN, Pryor AD, Jeyarajah DR, Minter RM, Mattar SG, Scott DJ, Brunt LM, Cummings M, Vassiliou M, Feldman LS. 2022. Implementation of entrustable professional activities into fellowship council accredited programs: a pilot project. Surgical Endoscopy. DOI:10.1007/s00464-022-09502-5.

Hauer KE, O'Sullivan PS, Fitzhenry K, Boscardin C. 2018. Translating theory into practice: implementing a program of assessment. Acad Med. 93(3):444–450.

Holmboe E, Ginsburg S, Bernabeo E. 2011. The rotational approach to medical education: time to confront our assumptions? Med Educ. 45(1):69–80.

Holmboe ES, Durning SJ, Hawkins RE., editors. 2018. A practical guide to the evaluation of clinical competence. Philadelphia (PA, USA): Elsevier.

Kamp MA, Malzkorn B, von Sass C, DiMeco F, Hadjipanayis CG, Senft C, Rapp M, Gepfner-Tuma I, Fountas K, Krieg SM, et al. 2021. Proposed definition of competencies for surgical neuro-oncology training. J Neurooncol. 153(1):121–131.

Kaur B, Taylor E. 2023. Development of a pediatric anesthesia fellowship curriculum in Australasia by the society for pediatric anesthesia of New Zealand and Australia (SPANZA) education sub-committee. Pediatr Anaesth. 33(2):100-106.

Kogan JR, Hatala R, Hauer KE, Holmboe E. 2017. Guidelines: the do's, don'ts and don't knows of direct observation of clinical skills in medical education. Perspect Med Educ. 6(5):286–305.

LaDonna KA, Hatala R, Lingard L, Voyer S, Watling C. 2017. Staging a performance: learners' perceptions about direct observation during residency. Med Educ. 51(5):498–510.

Landreville JM, Wood TJ, Frank JR, Cheung WJ. 2022. Does direct observation influence the quality of workplace-based assessment documentation? AEM Educ Train. 6(4):e10781.

Lockyer J, Carraccio C, Chan M-K, Hart D, Smee S, Touchie C, Holmboe ES, Frank JR, 2017. Core principles of assessment in competency-based medical education. Med Teach. 39(6):609–616.

Lockyer J. 2013. Multisource feedback: can it meet criteria for good assessment? J Contin Educ Health Prof. 33(2):89–98.

Lurie SJ. 2012. History and practice of competency-based assessment. Med Educ. 46(1):49–57.

Marty AP, Braun J, Schick C, Zalunardo MP, Spahn DR, Breckwoldt J. 2022. A mobile application to facilitate implementation of programmatic assessment in anaesthesia training. Br J Anaesth. 128(6):990–996.

Massie J, Ali JM. 2016. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. Adv Health Sci Educ Theory Pract. 21(2):455–473.

Maudsley G, Taylor D, Allam O, Garner J, Calinici T, Linkman K. 2019. A best evidence medical education (BEME) systematic review of: what works best for health professions students using mobile (hand-held) devices for educational support on clinical placements? BEME Guide No. 52. Med Teach. 41(2):125–140.

Mezirow J. 2009. An overview on transformative learning. In: Illeris K, editor. Contemporary Theories of Learning. Oxon: Routledge, p. 90–_105.

Miller GE. 1990. The assessment of clinical skills/competence/performance. Acad Med. 65(9 Suppl):S63–S67.

Nikou SA, Economides AA. 2018. Mobile-based assessment: a literature review of publications in major referred journals from 2009 to 2018. Computers and Education. Elsevier, 125:101–119.

NN 2018. What is GDPR, the EU's new data protection law? General Data Protection Regulation. [accessed 2022 Oct 25]. https://gdpr.eu/what-is-gdpr/.

Norcini J, Burch V. 2007. Workplace-based assessment as an educational tool: AMEE Guide No. Med Teach. 29(9):855–871.

Norcini J. 2020. Is it time for a new model of education in the health professions? Med Educ. 54(8):687–690.

Number of internet users worldwide as of 2022, by region no date. Statista. [accessed 2022 Oct 25]. https://www.statista.com/statistics/249562/number-of-worldwide-internet-users-by-region/.

Ochoa R, Lach S, Masaki T, Rodríguez-Castelán C. 2021. Why aren't more people using mobile internet in West Africa?, World Bank Blogs. [accessed 2022 Oct 25]. https://blogs.worldbank.org/digital-development/why-arent-more-people-using-mobile-internet-west-africa

Ott MC, Pack R, Cristancho S, Chin M, Van Koughnett JA, Ott M. 2022. 'The most crushing thing': understanding resident assessment burden in a competency-based curriculum. J Grad Med Educ. 14(5):583–592.

Oudkerk Pool A, Govaerts MJB, Jaarsma DADC, Driessen EW. 2018. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. Adv Health Sci Educ Theory Pract. 23(2):275–287.

Oxford English Dictionary. no date. Oxford University Press. https://www.oed.com.

Schut S, Driessen E, van Tartwijk J, van der Vleuten C, Heeneman S. 2018. Stakes in the eye of the beholder: an international study of learners' perceptions within programmatic assessment. Med Educ. 52(6):654–663.

Schut S, Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten C, Driessen E. 2021. Where the rubber meets the road—An integrative review of programmatic assessment in health care professions education. Perspect Med Educ. 10(1):6–13.

Sukhera J, Wodzinski M, Milne A, Teunissen PW, Lingard L, Watling C. 2019. Implicit bias and the feedback paradox: exploring how health professionals engage with feedback while questioning its credibility. Acad Med. 94(8):1204–1210.

Telio S, Regehr G, Ajjawi R. 2016. Feedback and the educational alliance: examining credibility judgements and their consequences. Med Educ. 50(9):933–942.

ten Cate O, Chen HC. 2020. The ingredients of a rich entrustment decision. Med Teach. 42(12):1413–1420.

ten Cate O, Hoff RG. 2017. From case-based to entrustment-based discussions. Clin Teach. 14(6):385–389.

ten Cate O, Scheele F. 2007. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? Acad Med. 82(6):542–547.

ten Cate O, Schwartz A, Chen HC. 2020. Assessing trainees and making entrustment decisions: on the nature and use of entrustment-supervision scales. Acad Med. 95(11):1662–1669.

ten Cate O. 2016. Entrustment as assessment: recognizing the ability, the right, and the duty to act. J Grad Med Educ. 8(2):261–262.

ten Cate O. 2022. How can entrustable professional activities serve the quality of health care provision through licensing and certification? Can Med Educ J. 13(4):8–14.

ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. 2015. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE Guide No. 99. Med Teach. 37(11):983–1002.

ten Cate O, Dahdal S, Lambert T, Neubauer F, Pless A, Pohlmann PF, van Rijen H, Gurtner C. 2020. Ten caveats of learning analytics in health professions education: a consumer's perspective. Med Teach. 42(6):673–678.

ten Cate O, Hart D, Ankel F, Busari J, Englander R, Glasgow N, Holmboe E, Iobst W, Lovell E, Snell LS, et al. 2016. Entrustment decision making in clinical training. Acad Med. 91(2):191–198.

ten Cate O, Regehr G. 2019. The power of subjectivity in the assessment of medical trainees. Acad Med. 94(3):333–337.

ten Cate O. 2014. AM last page: what entrustable professional activities add to a competency-based curriculum. Acad Med. 89(4):691.

Touchie C, Kinnear B, Schumacher D, Caretta-Weyer H, Hamstra SJ, Hart D, Gruppen L, Ross S, Warm E, Ten Cate O, I, et al. 2021. On the validity of summative entrustment decisions. Med Teach. 43(7):780–787.

van der Schaaf M, et al. 2017. Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. Educ Technol Res Dev. 65(2):359–380.

van der Vleuten C, Heeneman S, Schut S. 2020. Programmatic assessment: an avenue to a different assessment culture. In: Yudkowsky R, Park YS, Downing SM, editors. Assessent in Health Professions Education. 2nd edn. New York(NY): Routledge, p. 245–256.

Van Der Vleuten CPM, Schuwirth LWT, Driessen EW, Govaerts MJB, Heeneman S. 2015. Twelve tips for programmatic assessment. Med Teach. 37(7):641–646.

Warm EJ, Held JD, Hellmann M, Kelleher M, Kinnear B, Lee C, O'Toole JK, Mathis B, Mueller C, Sall D, et al. 2016. Entrusting observable practice activities and milestones over the 36 months of an internal medicine residency. Acad Med. 91(10):1398–1405.

Warm EJ, Mathis BR, Held JD, Pai S, Tolentino J, Ashbrook L, Lee CK, Lee D, Wood S, Fichtenbaum CJ, et al. 2014. Entrustment and mapping of observable practice activities for resident assessment. J Gen Intern Med. 29(8):1177–1182.

Watling C, LaDonna KA, Lingard L, Voyer S, Hatala R. 2016. "Sometimes the work just needs to be done": socio-cultural influences on direct observation in medical training. Med Educ. 50(10):1054–1064.

Weller JM, Castanelli DJ, Chen Y, Jolly B. 2017. 'Making robust assessments of specialist trainees' workplace performance. Br J Anaesth. 118(2):207–214.

Weller JM, et al. 2014. Can i leave the theatre? A key to more reliable workplace-based assessment. British Journal of Anaesthesia. 112(6):1083–1091.

Williams RG, Chen XP, Sanfey H, Markwell SJ, Mellinger JD, Dunnington GL. 2014. The measured effect of delay in completing operative performance ratings on clarity and detail of ratings assigned. J Surg Educ. 71(6):e132–e138.

Williams RG, George BC, Bohnen JD, Dunnington GL, Fryer JP, Klamen DL, Meyerson SL, Swanson DB, Mellinger JD. 2021. A proposed blueprint for operative performance training, assessment, and certification. Ann Surg. 273(4):701–708.

Young JQ, McClure M. 2020. Fast, easy, and good: assessing entrustable professional activities in psychiatry residents with a mobile app. Acad Med. 95(10):1546–1549.

Young JQ, Sugarman R, Schwartz J, McClure M, O'Sullivan PS. 2020. A mobile app to capture EPA assessment data: utilizing the consolidated framework for implementation research to identify enablers and barriers to engagement. Perspect Med Educ. 9(4):210–219.

Young JQ, Sugarman R, Schwartz J, O'Sullivan PS. 2020a. Faculty and resident engagement with a workplace-based assessment tool: use of implementation science to explore enablers and barriers. Acad Med. 95(12):1937–1944.

Young JQ, Sugarman R, Schwartz J, O'Sullivan PS. 2020b. Overcoming the challenges of direct observation and feedback programs: a qualitative exploration of resident and faculty experiences. Teach Learn Med. 32(5):541–551.

Yudkowsky R, Park YS, Downing SM, editors. 2020. Assessment in Health Professions Education. 2nd ed. New York: taylor & Francis.