

CHAPTER 17

Workplace-based assessment to support entrustment decision-making: four sources of information

Olle ten Cate, Gersten Jonker, Yoon Soo Park,
Eric S. Holmboe, Vanessa C. Burch

Abstract

A program of assessment that enables summative decisions about the trainee's readiness, such as a clinical competency committee, requires a synthesis of various sources of information and sufficient data points. Besides information about knowledge and skills, assessed outside the workplace (written examinations and standardized skills tests), workplace-based assessment can be categorized into four sources of information about a trainee: direct observation, conversation, longitudinal observation, and evaluation of 'products' of patient care. Direct observation of a trainee happens during a natural patient care activity in an authentic clinical setting, usually 10 to 20 minutes, followed by a few minutes of focused feedback. Conversations are a five- to 20-minute one-on-one discussion with a trainee to probe knowledge, understanding, reasoning, and/or decision-making. Longitudinal observation or monitoring checks the natural, unplanned observation of a trainee over time by collaborators and others (including patients) who have natural encounters with the trainee, often in the form of multisource feedback. Product evaluation pertains to the assessment of trainees through their output of patient care that does not require their direct presence during the assessment.

How to cite this book chapter:

ten Cate O, Jonker G, Park YS, Holmboe ES, Burch VC. Workplace-based assessment to support entrustment decision-making: four sources of information. In: ten Cate O, Burch VC, Chen HC, Chou FC, Hennis MP. (Eds). *Entrustable Professional Activities and Entrustment Decision-Making in Health Professions Education*, Chapter 17, pp. 197–211. [2024] London: Ubiquity Press. DOI: <https://doi.org/10.5334/bdc.q>

This chapter uses cross-references to other chapters of the same book. For those who read this chapter as a standalone publication: all cross-references can be found at: <https://doi.org/10.5334/bdc>

All four sources of information are discussed with examples and literature references. We end with notes on documentation of information, and feedback processes as an intrinsic component of workplace-based assessment.

Authors

- Olle ten Cate, PhD. University Medical Center Utrecht, Utrecht, the Netherlands and University of California, San Francisco, USA. Correspondence: t.j.tencate@umcutrecht.nl
- Gersten Jonker, MD, PhD. Department of Anesthesiology, University Medical Center Utrecht, Utrecht, the Netherlands.
- Yoon Soo Park, PhD. Department of Medical Education, University of Illinois College of Medicine, Illinois, USA.
- Eric Holmboe, MD. Intealth, Philadelphia, Pennsylvania, USA.
- Vanessa C. Burch, MD, PhD. Colleges of Medicine of South Africa, South Africa.

Workplace-based assessment to support feedback and summative entrustment decisions

In 1990, George Miller, a senior scholar in medical education, presented a famous model of assessment, arguing that significant advances had been made across his lifetime to increase the psychometric quality of the assessment of knowledge, applied knowledge and reasoning, and integrated knowledge and skills in a standardized environment.

However, the ‘does’ level in the pyramid that was named after him was left for a necessary future advance in medical education.¹ Workplace-based assessment reflects the assessment at this fourth level of the pyramid but it was not well developed at the time.

Since Miller’s publication, workplace-based assessment (WBA) has been developed in many ways^{2,3} to include tools, forms, procedures, and approaches. Assessment of trainees in the health professions, to determine their readiness to practice, has evolved from limited static moments of high-stakes assessment into a more continuous and developmental interaction with trainees to support learning, combined with evaluating their readiness for practice requiring varying levels of supervision and support.⁴ Readiness for practice is increasingly translated to entrustment: the decision to transfer patient care responsibilities to a trainee who has shown to possess the required competencies.

Entrustment decisions require solid information about the trainee. In the clinical workplace, a multitude of sources of information is available to inform their development and their readiness for practice.

WBA is assessment in the variable context of patient care. While trainees have acquired generalized, context-independent knowledge and skill (this can be called a canonical layer of competence), working in practice requires the ability to apply generalized knowledge and skills in circumstances that vary. This ‘contextual’ layer of competence must be evaluated to justify summative entrustment decisions. Students must meet standards of contextual performance derived from observations in actual practice before permission to practice should be granted. In contrast with assessment under standardized conditions, WBA requires case-by-case judgments by individuals, because the conditions in the workplace (patient cases, assessors’ experience and expertise, team composition, time of day, and day of week) cannot be easily standardized. Assessors are

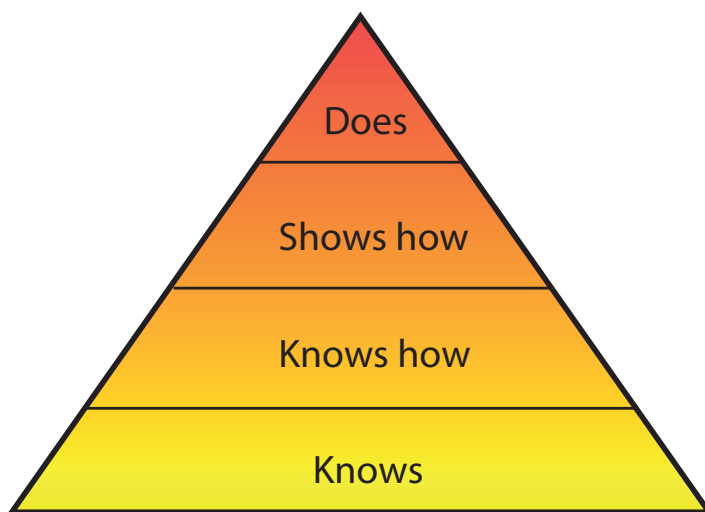


Figure 17.1: Miller’s pyramid.

often clinician-preceptors who happen to be available in the workplace but they can also include others, such as interprofessional personnel. Traditional psychometric criteria that work well with standardized assessment (e.g., written tests) are seldom met in workplace-based assessment.^{5,6} A different approach is required, including aggregation of data from multiple sources.⁷ Crossley et al. have argued that good, intuitive, questions when assessing trainees will lead to more reliable answers.⁸ ‘Would you entrust this learner with care for the next patient?’ is a more intuitive question than ‘Please rate trainee for professionalism on a nine-point scale.’ The concept of entrustment decisions is essentially new as a way of assessing trainees, although it has been informally practiced for as long as anyone can remember. Entrustment for unsupervised future practice may be conceptualized as a fifth level in Miller’s pyramid, beyond observing how well a trainee has ‘done’ in the past so far.⁹

While the assessment of knowledge and reasoning, as well as standardized skills assessments, are very important in health professions education, the focus in this chapter is on assessment in the workplace, to support decisions of entrustment with tasks in health care.

EPAs and entrustment decision-making within workplace-based assessment

Assessment usually concerns a conclusion regarding work done and thus draws a *retrospective* summative conclusion. A *prospective* view implies an estimation of readiness for future practice, which requires more than retrospection of completed exams and assignments. A trainee, for example, with high scores on exams, and often doing well while observed in the clinical workplace, but who is not inclined to ask for help if needed, or does not collaborate well, may be less trustworthy for future tasks than someone who maybe scored less well on knowledge exams but who knows when to ask for help and how to mobilize resources when needed to bring a clinical problem to a satisfactory end,¹⁰ or ‘knows what to do when you don’t know what to do.’^a

Permission to practice patient care can be fully unsupervised, but in education it usually refers to permission under a specified level of supervision. Translated to a commonly used entrustment–supervision (ES) scale,¹² *observation only* (level 1) permits no trainee contribution to care, *direct supervision* (level 2) signifies a supervisor being present to observe and take over the activity if needed, and *indirect supervision* (level 3) means that a supervisor is not present but on site and available to assist if needed or if so requested. Off-site supervision (level 4), e.g., by phone, relies heavily on a trainee’s readiness to work with virtually no supervision, and their willingness to seek help if really needed.

Ad hoc entrustment decisions are those that occur ‘on the fly’ in the clinical moment (‘now you go ahead with this patient; call me if you need me’). This reflects individual supervisor decisions, in contrast to *summative* entrustment decisions, which have the nature of a formalized qualification, supported by a team decision.

Entrustable professional activities (EPAs), being units of practice to be entrusted to trainees once they demonstrate the required competence, can be regarded as tasks (or bundles of tasks) that can be well observed, overseen, and monitored. A summative entrustment decision should be made only when a trainee has demonstrated readiness for more autonomy. That requires thoughtful decision-making, based on various data sources, by a team of knowledgeable educators, often called a (clinical) *competency committee* or entrustment committee.¹⁰

^a Derived from a child-pedagogy book¹¹ and cited in Dr. Ian Bates’s 2017 keynote talk at a Dutch HPE conference.

Multiple sources of information in the workplace are available to estimate competence and add to the grounding of trust.¹³ Clearly, a valid, holistic picture of the trainee cannot be based on a single observation. Following the premise of *programmatic assessment*,¹⁴ many and varied observations together must build an adequate picture of the trainee's readiness to practice. This includes not just knowledge and skills but also professional attributes like humility, agency, integrity, and reliability.¹⁵

In a program that is serious about a stepwise increase in the progressive autonomy and responsibility of trainees, clinical education should include formalized moments of summative entrustment for EPAs. EPAs can be narrow or broad tasks, depending on what the activity includes, but an educational program should organize their EPAs such that summative entrustment decisions to qualify trainees for unsupervised practice (level 4; in undergraduate medical education for indirect supervision: level 3) are made not more than a few times per year, reflecting significant steps toward more autonomy. Early in education, EPAs are usually small in scope (e.g., taking general neurological histories; placing IUDs; removing a small benign lump); later they can be broad (e.g., running the internal medicine ward day service; running the Friday anesthesia night shift).

Four sources of assessment information to support summative entrustment decisions

The emphasis in WBA is on contextual competence. The presumption is that trainees possess a foundation of canonical knowledge and skills; they are assessed with the purpose of entrustment with clinical tasks. To arrive at summative entrustment decisions, a clinical competency committee must rely on data from various sources. These sources can be categorized into four groups: direct observations, conversations, longitudinal observations, and evaluations of 'products' of the trainee's care provision.^{16,17} In a *program of assessment*, sources of information must be identified that align best with a particular EPA and inform a summative entrustment decision. The recommendation is to draw from all four sources of information.

1. Direct observation

A direct, focused observation of a trainee during a natural patient care activity in an authentic clinical setting usually takes 10 to 20 minutes, followed by a few minutes of rating and feedback. Trainee and supervisor should have a shared understanding of its purpose, i.e., both to assess and to support the trainee with performance feedback. The most common types of direct observation focus on a clinical encounter or a procedure.

Forms and scales. Traditional rating forms have been criticized because they do not align well with the thinking of a clinical assessor.⁸ Numerical scales (e.g., 1–9, or unsatisfactory–satisfactory–superior; or below–meets–exceeds expectations) are increasingly replaced by ES scales, either retrospective ('how much guidance was needed?')^{28,29} or prospective ('based on my observations so far, I recommend that this trainee is ready for supervision level [X] for this EPA').¹² Ample space for narrative, descriptive assessment data is also increasingly recommended, albeit that more research is needed to optimize its quality and use.^{30,31}

How many observations? There is no simple answer, as this depends on the task, the trainee, and their stage of training or development. Complex tasks may require much practice under direct supervision; simple, low-risk tasks require less. An advanced or skillful trainee may

Table 17.1: Four sources of workplace information to support summative entrustment decisions.

	Defined	Examples and references
Direct observation	Focused observation of a trainee during a natural patient care activity in an authentic clinical setting, usually 10–20 minutes, followed by a few minutes of focused feedback	Among the dozens of tools ¹⁸ and guidelines, ¹⁹ observation of history and physical examination in clinical encounters using mini-CEX and observation of procedural skills using DOPS are frequently used. ³ ES scales are used to support entrustment decisions made ¹²
Conversation	A five- to 20-minute one-on-one discussion with a trainee to probe knowledge, understanding, reasoning and/or decision-making	Case-based discussions (CBDs) and chart-stimulated recall (CSR) ^{2,3,20} focus on knowledge and reasoning, and performed actions; entrustment-based discussions (EBD) ²¹ focus on anticipated action and risk assessment. Brief teaching conversations ²² such as SNAPPS* and OMP* and pre- and debriefings of encounters contribute to trainee evaluation
Longitudinal observation	The natural, unplanned observation of a trainee over time by collaborators and others (including patients) who have natural encounters with the trainee	Multisource feedback procedures ²³ are typically longitudinal. Automated procedures can be done. ^{24,25} The Mini-PAT is a well-known MSF form ² and the A RICH framework ¹⁵ may inform narrative comments
Product evaluation	Assessment of trainees through their outputs of patient care that do not require their direct presence during the assessment	Examples are entries into health records (discharge letters), physical products (dentistry, plastic surgery, orthopedics, etc.), patient experiences, clinimetrics, complication rates, and readmissions ^{26,27}
* See text		

require less than a beginner, and patient availability and clinical workload³² can be a limiting factor for observations.

Important, however, is that only requiring a fixed number of observations neglects the importance of individual differences. There is nothing wrong with a threshold number ('at least X'), as long as 'completing the number' is not used to conclude that the trainee is competent.³³ It is more important to consider the number of *recent satisfactory* performances, preferably under different circumstances. Up to eight satisfactory observations have been suggested for undergraduate medical EPAs as thresholds for entrustment,³⁴ but for some EPAs fewer may be enough. Obviously, older observations, when a trainee is still inexperienced, should not weigh as much as recent ones. Sampling should reflect a thoughtful representation, adequate for the entrustment decision, but that does not mean that every trainee should be observed as much; underperforming trainees, for instance, generally require more observations.

Who initiates the observations? In many programs, it is the trainee who initiates direct observations by inviting clinicians to observe and report. There is benefit of trainee ownership and control of their curriculum and development, which alleviates staff of the burden of planning observations.³⁵ A downside is that trainee-initiated observations are not always a random 'snapshot' or adequate sample of practice and may stimulate trainees to document their best performances, including requests for hindsight documentation when an activity was done well in the presence

Box 17.1: Direct observation.

A clinical encounter with a patient. This can include history taking, physical examination of the patient, a conversation with the patient about a clinical decision, asking for informed consent, breaking bad news, etc. Observations are documented on a structured form. This invites the observer to evaluate interviewing skills, physical examination skills, humanistic qualities, clinical judgment, counseling skills, organization/efficiency, and an overall score, plus free comments.

A procedure involving a patient, requiring technical skill. Procedures can range from inserting an IV, vaccination, venous or arterial blood withdrawal, various punctures, a surgical procedure, administering a cognitive screening test, etc. Usually it is not only technical skill that is evaluated but also the interaction.

of a clinician. That is not always problematic, but it can create an imbalanced picture. Showing one's 'best practices' is not a problem in itself, as long as it does not mask critical deficiencies. In contrast, supervisor-selected convenient observation moments can contribute to bias, including leniency bias.³⁶ Cocreation of individualized WBA among trainees and multiple supervisors is recommended to arrive at a balanced selection of observations.

Faculty and trainee experiences with direct observation. Direct observation requires faculty effort that cannot be spent on efficient patient care. Resistance against implementing direct observation is frequently heard,³⁷ but establishing an observation-based training culture does not necessarily imply a huge time commitment.³⁸ Trainees may feel awkward or uncomfortable asking attending faculty to observe them.^{39,40} Also, when trainees experience every observation as a critical, high-stakes assessment moment, rather than a welcomed, low-stakes learning moment,⁴¹ a culture change is needed, to reframe a teaching hospital as more than just a health care institution. When formative assessment and feedback become routine, and not just incidental, observation can be experienced as a natural component of the training culture, rather than a continuous examination.

2. Conversation

Case-based discussions (CBDs) with trainees—if focused on a trainee's case that has been documented in a health record, also called chart-stimulated recall (CSR)—is a well-established method to capture trainee knowledge and understanding needed to perform patient care. Unlike standardized reasoning tests, these conversations generally relate to cases in the local context, often the trainee's own patients, and test what some authors call situated cognition,⁴² rather than decontextualized cognition. CBDs and CSR particularly focus on what a trainee's reasoning and rationales are (or were) as related to the case. These conversations are 10 to 20 minutes including feedback and documentation. The procedures are well described elsewhere.^{2,3,20}

Some conversational approaches have been designed as individual teaching moments rather than assessments, but they do add to a preceptor's picture of the trainee. Well known are SNAPPS (Summarize history and findings–Narrow the differential–Analyze–Probe the Preceptor about uncertainties–Select issues for further study), which is trainee-driven, and the One-Minute Preceptor (Get a commitment–Probe for underlying reasoning–Teach a general rule–Reinforce what was done well–Correct mistakes), which is more supervisor-driven.⁴³

One approach that is particularly suitable for entrustment decision-making is the ‘entrustment-based discussion.’²¹ This 10- to 20-minute conversation has a focus on risk assessment when anticipating an entrustment decision, either directly after an EPA has been performed or before an upcoming activity (supporting ad hoc entrustment). Box 17.2 explains the procedure.

Box 17.2: Four steps in the entrustment-based discussion.²¹

1. Explain the activity

What are procedural steps? Have you done this before? What are critical parts?

2. Why is/was that needed?

Relevant anatomy/physiology, indications, rationale in this patient’s management

3. What are/were associated risks?

What are consequences of inadequate performance, what could happen?

4. What would you do if...?

Give scenario (unusual or uncooperative patient, different history and physical findings, different test results, comorbidity) and probe for actions.

All these encounters are meant to be brief, regular, and formative, rather than long, infrequent, and merely assessment-focused. Long oral exams, such as a ‘long case’ (an hour or more with a patient and a subsequent interrogation) at the end of rotations are no longer considered adequate WBA methods.⁴⁴

3. Longitudinal observation

Not all important behaviors, relevant for summative entrustment decisions, can be captured in brief direct observations and discussions. During direct observation, the trainee is usually aware of being observed and may not behave completely naturally⁴⁵; longitudinal observation does not happen in an arranged setting and is better suited to capture qualities of professionalism, relevant to entrustment decisions (arriving at meetings on time, doing what was promised, setting priorities, communicating adequately interprofessionally, etc.). Longitudinal observation is the natural, unplanned observation of a trainee over time by collaborators or others (including patients) who have natural encounters with the trainee. Features that are suitable for longitudinal observation include the behavioral characteristics that are important for entrustment. These were recently summarized and grouped into five categories as the ‘A RICH’ entrustment decision (Box 17.3).¹⁵

Box 17.3: Trainee features relevant for A RICH entrustment decision.¹⁵

Agency:	Self-confident, proactive toward work, team, safety, and own development
Reliability:	Conscientious; predictable; accountable; responsible
Integrity:	Truthful; having good, benevolent, intentions; being patient-centered
Capability:	Having focused knowledge and skills, experience, adaptive expertise
Humility:	Observing personal limits; willing to ask help; receptive to feedback

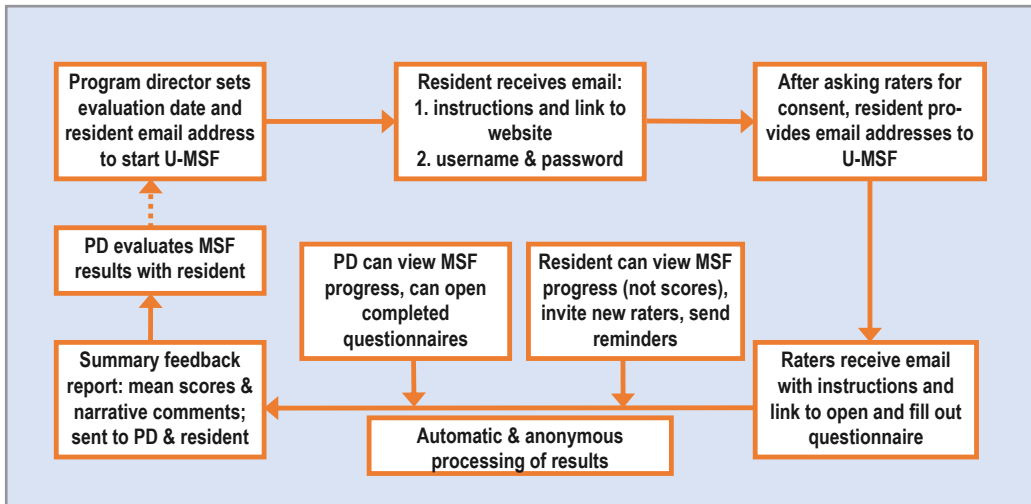


Figure 17.2: Example of an automated MSF procedure.

Opinions of supervisors and others may accumulate over time to build a picture of a trainee in colleagues with whom the trainee works. Such informal opinions are not sufficiently grounded for important decision-making; more formalized, structured approaches are needed. One such procedure is multisource feedback (MSF) or 360-degree evaluation, in which clinical staff members, peers and other trainees, interprofessional colleagues, and patients and their families may be included. For longitudinal observation, assessors must be asked in advance to observe over a period of time (a week, a weekend shift, a rotation). While MSF may require quite some logistical effort,⁴⁶ web-based methods can alleviate most of the administrative burden. Figure 17.2 shows a MSF procedure that has been successfully applied in many residency programs in the Netherlands as a fully automated procedure with little administrative effort.^{24,25}

Trainees in this example select their own observers (six medical colleagues, six nonmedical colleagues, and 10 patients), who then respond to a brief email survey, combining a rating scale with adequate space for narrative comments. The automated, anonymous reports (two to three pages), generated by the system, have proved to be highly useful, in particular the narrative component.^{24,25} Another regularly used example of an MSF form is the mini-Peer Assessment Tool (mini-PAT).² While the A RICH components are less suitable for rating scales, they may inform MSF raters' frame of reference for narrative comments, over longitudinal periods of time. The trainee's reflection on an MFS report is important, and can be supported by a coach or mentor.²⁵

Longitudinal monitoring provides critical information for clinical competency committees. MSF reports can show trends, if various sources point in a similar direction. This analysis should inform CCCs; even less structured longitudinal information can be useful but needs to be corroborated by views from different observers, akin to quality journalism that relies on multiple independent sources.

4. Product evaluation

One of the quality features of entrustable professionals activities is that the EPA is 'measurable in its outcome,' enabling a judgment of 'well done' or not 'well done'.^{47,48} This looks at the result or product of an activity. By product evaluation we mean the assessment of trainees through their output of patient care, while not requiring their direct presence during the assessment. Those familiar

with Kirkpatrick's four hierarchical levels of educational effectiveness (satisfaction → learning → behavior change → results)⁴⁹ will agree that most WBA focuses on 'learning' (as established in tests and performances) and on some behavior change but less on 'results,' i.e., on effective health care provided. In other sectors of society, professional services and their providers are frequently evaluated by products (artifacts such as reports, designs, manufactured and engineered products), but in health care product evaluation happens less frequently. The effect of learning, training, and competence on patient care, including benefits to patients, is not easy to demonstrate or measure but it is worth pursuing.⁵⁰ Products of individuals in health care can include entries in an electronic health record, discharge summaries, physical products from procedures performed, quality improvement and patient safety (QIPS) measures, patient experiences, clinimetrics,²⁶ portfolio-logged patient encounters, and other, specialty-specific products. Research output, reports created by a trainee, and presentations can sometimes be included, even if less linked to clinical care.

Some procedural professions and specialties yield clear products (dentistry, e.g., a crown; plastic surgery, e.g., a facial reconstruction; orthopedic surgery, e.g., fractured bone reconstruction, as evidenced by imaging). While these may still reflect team efforts, the product is visible. In non-procedural specialties, such as in psychiatry, the suicidal status must be documented, and written patient handovers, coherence of patient records, and therapeutical adherence can be measured.

While surgical complication rates and patient readmissions are other outcome measures, they may be even less easily directly linked to individual effort, while some examples, for instance in obstetric care, do exist.⁵¹ Some scholars argue that in health care the quality or competence of teams is more important than those of individuals,⁵² but teams do not graduate from programs and are not licensed to practice. To link individual competence to the products in health care, it has been suggested that contribution analysis may be used to determine each individual's contribution to the team outcome.^{53,54} Resident-sensitive or trainee-attributable quality measures are being explored, developed, and researched for WBA linked to EPAs^{27,55}; more research and development needs to be done before these measures can be used for product evaluation.

Summative entrustment decision-making in clinical competency committees can include, besides observations and discussions, an assessment of products of care, but that means these should be made visible to the assessors. As with other elements of WBA, this requires a two-stage procedure: formative evaluation with feedback to trainees and documentation in a portfolio, and weighing of these for summative decisions by the decision-making team.

Documentation

WBAs conducted in daily clinical practice must be documented to enable data aggregation, and support validity and transparency in summative entrustment decision-making. The ambition of programmatic assessment, combined with the constraints of a clinical environment, is high-quality, efficient documentation. Electronic portfolios and mobile apps⁵⁶ now facilitate realization of these ambitions through dashboards that systematically synthesize multiple sources of information. Technology should cause minimal disruption to clinical workflow, not only to minimize additional effort but also to maintain the authenticity of the workplace context. Oral feedback can be recorded, transcribed, sent to portfolios,⁵⁷ and aggregated with scores to produce insightful visual displays for program directors and competency committees. Rating forms must include carefully designed questions and items,⁸ to optimize the response process by observers without requiring instruction that is too detailed. Other chapters discuss the use of rating scales (Chapter 19) and technology support (Chapter 20) for WBA more deeply. Narrative information is becoming more prominent in WBA, both for feedback conversations and for documentation to support summative decisions by the educational team.⁵⁸ New advances in learning analytics and AI are likely to optimize conclusions and decisions about trainees and programs in the near future.^{53–55}

Feedback

Assessment, evaluation, feedback, and teaching in the workplace cannot be disentangled. Recommendations to separate coaching and mentoring from assessment to improve objectivity⁵⁹ are debatable, simply because a preceptor or tutor has more relevant knowledge about a trainee's readiness for entrustment than does an independent assessor, who may not know the trainee well enough. Role conflicts between assessing versus coaching are better resolved by: (a) regarding direct observation as low stakes and formative; (b) requiring feedback conversations to be more than providing comments and support improvement as an inherent component of observation; (c) leaving high-stakes decision-making regarding progress and summative entrustment to an educational team; (d) involving trainees in the assessment process as much as possible through feedback conversations; and (e) cocreation of developmental paths. Assessment in the clinical workplace is largely meant to support learning^{60–62} as a continuous endeavor, and only use distinct moments for making summative decisions.

Faculty development

Skills increase with training and deliberate practice.⁶³ That is true for clinical skills as well as skills for observation and assessment of trainees. Experts easily forget how much clinical practice experience is needed to gain expertise;^b this holds similarly to observation and assessment. Faculty development for observing and evaluating trainees must be supplemented with regular, deliberate assessment practice.

Assessment is not only done by clinical supervisors but also by peers (e.g., residents evaluating medical students) and others in the clinical environment, such as for multisource feedback. While the recommendation is to create procedures that are simple and self-explanatory, instruction and training are also needed for high-quality observation and feedback,¹⁹ including so-called 'frame-of-reference' training, where the educational team shares norms and criteria to build the quality of their judgments. Chapter 23 provides more information on faculty development.

Conclusion

WBA with the purpose to support summative entrustment decision-making needs to draw on various sources of information available in the clinical workplace. The collection of this information should happen as an integral component of daily clinical business and belongs to the culture of any clinical setting that is responsible for educating health professionals. Four sources of information (direct observation, longitudinal observation, conversations, and evaluation of care produced by trainees) all have their place in a program of assessment, one of the important components of competency-based education.⁶⁵

Figure justifications

Figure 17.1 is derived from Miller (1990).¹

^b As an example, machine learning studies reveal that up to 100,000 annotated specimens in nephropathology may be needed before an algorithm has learned to recognize a glomerulus;⁶⁴ senior pathologists may forget how many they saw to become an expert.

Competing interests

The authors declare that they have no competing interests.

References

1. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9 Suppl):S63-S67. DOI: <https://doi.org/10.1097/00001888-199009000-00045>
2. Norcini J, Burch V. Workplace-based assessment as an educational tool: AMEE guide no. 31. *Med Teach*. 2007;29(9):855–871. DOI: <https://doi.org/10.1080/01421590701775453>
3. Norcini JJ, Zaidi Z. Workplace-based assessment. In: Swanwick T, Forrest K, O'Brien B, eds. *Understanding Medical Education*. 3rd ed. Wiley; 2018:319–334.
4. Holmboe ES, Yamazaki K, Hamstra SJ. The evolution of assessment: thinking longitudinally and developmentally. *Acad Med*. 2020;95(11S):S7-S9. DOI: <https://doi.org/10.1097/ACM.0000000000003649>
5. Govaerts MJB, van der Vleuten CPM, Schuwirth LWT, Muijtens AMM. Broadening perspectives on clinical performance assessment: rethinking the nature of in-training assessment. *Adv Health Sci Educ Theory Pract*. 2007;12(2):239–260. DOI: <https://doi.org/10.1007/s10459-006-9043-1>
6. Ryan MS, Gielissen KA, Shin D, et al. How well do workplace-based assessments support summative entrustment decisions? A multi-institutional generalisability study. *Med Educ*. January 2, 2024. DOI: <https://doi.org/10.1111/medu.15291>
7. Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv Health Sci Educ Theory Pract*. 2013;18(5):1087–1102. DOI: <https://doi.org/10.1007/s10459-013-9450-z>
8. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ*. 2011;45(6):560–569. DOI: <https://doi.org/10.1111/j.1365-2923.2010.03913.x>
9. ten Cate O, Carraccio C, Damodaran A, et al. Entrustment decision-making: extending Miller's pyramid. *Acad Med*. 2021;96(2):199–204. DOI: <https://doi.org/10.1097/ACM.0000000000003800>
10. Schumacher DJ, Michelson C, Winn AS, Turner DA, Martini A, Kinnear B. A realist synthesis of prospective entrustment decision-making by entrustment or clinical competency commitments. *Med Educ*. December 13, 2023. DOI: <https://doi.org/10.1111/medu.15296>
11. van Manen M. *Pedagogical Tact - Knowing What to Do When You Don't Know What to Do*. Routledge; 2016. DOI: <https://doi.org/10.4324/9781315422855>
12. ten Cate O, Schwartz A, Chen HC. Assessing trainees and making entrustment decisions: on the nature and use of entrustment-supervision scales. *Acad Med*. 2020;95(11):1662–1669. DOI: <https://doi.org/10.1097/ACM.0000000000003427>
13. ten Cate O, Hart D, Ankel F, et al. Entrustment decision-making in clinical training. *Acad Med*. 2016;91(2):191–198. DOI: <https://doi.org/10.1097/ACM.0000000000001044>
14. Hauer KE, O'Sullivan PS, Fitzhenry K, Boscardin C. Translating theory into practice: implementing a program of assessment. *Acad Med*. 2018;93(3):444–450. DOI: <https://doi.org/10.1097/ACM.0000000000001995>
15. ten Cate O, Chen HC. The ingredients of a rich entrustment decision. *Med Teach*. 2020;42(12):1413–1420. DOI: <https://doi.org/10.1080/0142159X.2020.1817348>
16. ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. Curriculum development for the workplace using entrustable professional activities (EPAs): AMEE guide no. 99. *Med Teach*. 2015;37(11):983–1002. DOI: <https://doi.org/10.3109/0142159X.2015.1060308>

17. McBride ME, Adler MD, McGaghie WC. Workplace-based assessment. In: Yudkowsky R, Park YS, Downing SM, eds. *Assessment in Health Professions Education*. 2nd ed. Routledge; 2020:160–172.
18. Kogan JR, Holmboe ES, Hauer KE. Tools for direct observation and assessment of clinical skills of medical trainees - a systematic review. *JAMA*. 2009;302(12):1316–1326.
19. Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: the do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ*. 2017;6(5):286–305. DOI: <https://doi.org/10.1007/s40037-017-0376-7>
20. Holmboe ES, Durning SJ, eds. *Assessment of Clinical Competence*. 3rd ed. Elsevier; 2025.
21. ten Cate O, Hoff RG. From case-based to entrustment-based discussions. *Clin Teach*. 2017;14(6):385–389. DOI: <https://doi.org/10.1111/tct.12710>
22. Irby DM, Wilkerson L. Teaching when time is limited. *BMJ*. 2008;336(7640):384–387. DOI: <https://doi.org/10.1136/bmj.39456.727199.AD>
23. Lockyer J. Multisource feedback in the assessment of physician competencies. *J Contin Educ Health Prof*. 2003;23(1):4–12. DOI: <https://doi.org/10.1002/chp.1340230103>
24. Alofs L, Huiskes J, Heineman MJ, et al. User reception of a simple online multisource feedback tool for residents. *Perspect Med Educ*. 2015;4(2):57–65. DOI: <https://doi.org/10.1007/s40037-015-0173-0>
25. Buis CAM, Eckenhausen MAW, ten Cate O. Processing multisource feedback during residency under the guidance of a non-medical coach. *Int J Med Educ*. 2018;9:48–54. DOI: <https://doi.org/10.5116/ijme.5a7f.169d>
26. Fava GA, Tomba E, Sonino N. Clinimetrics: the science of clinical measurements. *Int J Clin Pract*. 2012;66(1):11–15. DOI: <https://doi.org/10.1111/j.1742-1241.2011.02825.x>
27. Schumacher DJ, Holmboe ES, van der Vleuten C, Busari JO, Carraccio C. Developing resident-sensitive quality measures: a model from pediatric emergency medicine. *Acad Med*. 2018;93(7):1071–1078. DOI: <https://doi.org/10.1097/ACM.0000000000002093>
28. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 2012;87(10):1401–1407. DOI: <https://doi.org/10.1097/ACM.0b013e3182677805>
29. Zwischenberger JB, Hatchett SP, Rahman NA, Chadha P, Sebok-Syer SS, George BC. Implementing workplace-based assessments at scale: The SIMPL approach. *Ann Surg Open*. 2023;4(4):e353. DOI: <https://doi.org/10.1097/AS9.0000000000000353>
30. Holmboe ES, Osman NY, Murphy CM, Kogan JR. The urgency of now: rethinking and improving assessment practices in medical education programs. *Acad Med*. 2023;98(8S):S37–s49. DOI: <https://doi.org/10.1097/ACM.00000000000005251>
31. Burch V. The changing landscape of workplace-based assessment. *Journal of Applied Testing Technology*. 2019;20(S2):37–59.
32. Park YS, Hicks PJ, Carraccio C, Margolis M, Schwartz A, PMAC Module 2 Study Group. Does incorporating a measure of clinical workload improve workplace-based assessment scores? Insights for measurement precision and longitudinal score growth from ten pediatrics residency programs. *Acad Med*. 2018;93(11S):S21–S29. DOI: <https://doi.org/10.1097/ACM.00000000000002381>
33. ten Cate O. The false dichotomy of quality and quantity in the discourse around assessment in competency-based education. *Adv Health Sci Educ Theory Pract*. 2015;20(3):835–838. DOI: <https://doi.org/10.1007/s10459-014-9527-3>
34. Harvey A, Paget M, McLaughlin K, et al. How much is enough? Proposing achievement thresholds for core EPAs of graduating medical students in Canada. *Med Teach*. 2023;45(9):1054–1060. DOI: <https://doi.org/10.1080/0142159X.2023.2215910>
35. Holleran C, Konrad J, Norton B, Burlis T, Ambler S. Use of learner-driven, formative, ad-hoc, prospective assessment of competence in physical therapist clinical education in the United States: a prospective cohort study. *J Educ Eval Health Prof*. 2023;20:36. DOI: <https://doi.org/10.3352/jeehp.2023.20.36>

36. Albanese MA. Challenges in using rater judgements in medical education. *J Eval Clin Pract.* 2000;6(3):305–319. DOI: <https://doi.org/10.1046/j.1365-2753.2000.00253.x>
37. Ras T, Stander Jenkins L, Lazarus C, et al. ‘We just don’t have the resources’: supervisor perspectives on introducing workplace-based assessments into medical specialist training in South Africa. *BMC Med Educ.* 2023;23(1):832. DOI: <https://doi.org/10.1186/s12909-023-04840-x>
38. Cheung K, Rogoza C, Chung AD, Kwan BYM. Analyzing the administrative burden of competency based medical education. *Can Assoc Radiol J.* 2022;73(2):299–304. DOI: <https://doi.org/10.1177/08465371211038963>
39. Ott MC, Pack R, Cristancho S, Chin M, Van Koughnett JA, Ott M. ‘The most crushing thing’: understanding resident assessment burden in a competency-based curriculum. *J Grad Med Educ.* 2022;14(5):583–592. DOI: <https://doi.org/10.4300/JGME-D-22-00050.1>
40. Ahn E, LaDonna KA, Landreville JM, Mcheimech R, Cheung WJ. Only as strong as the weakest link: resident perspectives on entrustable professional activities and their impact on learning. *J Grad Med Educ.* 2023;15(6):676–684. DOI: <https://doi.org/10.4300/JGME-D-23-00204.1>
41. Schut S, Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten C, Driessen E. Where the rubber meets the road - an integrative review of programmatic assessment in health care professions education. *Perspect Med Educ.* 2021;10(1):6–13. DOI: <https://doi.org/10.1007/s40037-020-00625-w>
42. Rencic J, Schuwirth LWT, Gruppen LD, Durning SJ. A situated cognition model for clinical reasoning performance assessment: a narrative review. *Diagnosis (Berl).* 2020;7(3):227–240. DOI: <https://doi.org/10.1515/dx-2019-0106>
43. Pascoe JM, Nixon J, Lang VJ. Maximizing teaching on the wards: review and application of the One-Minute Preceptor and SNAPPS models. *J Hosp Med.* 2015;10(2):125–130. DOI: <https://doi.org/10.1002/jhm.2302>
44. Boursicot K, Roberts TE, Burdick WP. Structured assessment of clinical skills. In: Swanwick T, Forrest K, O’Brien BC, eds. *Understanding Medical Education*. 3rd ed. John Wiley & Sons Inc; 2019:335–346.
45. LaDonna KA, Hatala R, Lingard L, Voyer S, Watling C. Staging a performance: learners’ perceptions about direct observation during residency. *Med Educ.* 2017;51(5):498–510. DOI: <https://doi.org/10.1111/medu.13232>
46. Lockyer J. Multisource feedback. In: Holmboe ES, Durning SJ, Hawkins RE, eds. *Evaluation of Clinical Competence*. 2nd ed. Elsevier; 2018:204–214.
47. ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ.* 2005;39(12):1176–1177. DOI: <https://doi.org/10.1111/j.1365-2929.2005.02341.x>
48. Taylor DR, Park YS, Egan R, et al. EQual, a novel rubric to evaluate entrustable professional activities for quality and structure. *Acad Med.* 2017;92(11S Association of American Medical Colleges Learn Serve Lead: Proceedings of the 56th Annual Research in Medical Education Sessions):S110–S117. DOI: <https://doi.org/10.1097/ACM.0000000000001908>
49. Kirkpatrick DL, Kirkpatrick JD. *Evaluating Training Programs: The Four Levels*. 3rd ed. Berrett-Koehler Publishers, Inc; 2006.
50. Yardley S, Dornan T. Kirkpatrick’s levels and education ‘evidence.’ *Med Educ.* 2012;46(1):97–106. DOI: <https://doi.org/10.1111/j.1365-2923.2011.04076.x>
51. Epstein AJ, Nicholson S, Asch DA. The production of and market for new physicians’ skill. *American Journal of Health Economics.* 2016;2(1):41–65. DOI: https://doi.org/10.1162/AJHE_a_00033
52. Lingard LA. Rethinking competence in the context of teamwork. In: Hodges BD, Lingard L, eds. *The Question of Competence: Reconsidering Medical Education in the Twenty-First Century*. Cornell University Press; 2012.

53. ten Cate O, Carrie Chen H. The parts, the sum and the whole - evaluating students in teams. *Med Teach*. 2016;38(7):639–641. DOI: <https://doi.org/10.3109/0142159X.2016.1170794>
54. Schumacher DJ, Dornoff E, Carraccio C, et al. The power of contribution and attribution in assessing educational outcomes for individuals, teams, and programs. *Acad Med*. 2020;95(7):1014–1019. DOI: <https://doi.org/10.1097/ACM.0000000000003121>
55. Schumacher DJ, Kinnear B, Burk-Rafel J, Santen SA, Bullock JL. The next era of assessment: can ensuring high-quality, equitable patient care be the defining characteristic? *Acad Med*. December 18, 2023. DOI: <https://doi.org/10.1097/ACM.0000000000005603>
56. Marty AP, Linsenmeyer M, George B, Young JQ, Breckwoldt J, ten Cate O. Mobile technologies to support workplace-based assessment for entrustment decisions: guidelines for programs and educators: AMEE guide no. 154. *Med Teach*. 2023;45(11):1203–1213. DOI: <https://doi.org/10.1080/0142159X.2023.2168527>
57. Young JQ, McClure M. Fast, easy, and good: assessing entrustable professional activities in psychiatry residents with a mobile app. *Acad Med*. 2020;95(10):1546–1549. DOI: <https://doi.org/10.1097/ACM.0000000000003390>
58. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: validity evidence for qualitative educational assessments. *Acad Med*. 2016;91(10):1359–1369. DOI: <https://doi.org/10.1097/ACM.0000000000001175>
59. Cavalcanti RB, Detsky AS. The education and training of future physicians - why coaches can't be judges. *JAMA*. 2011;306(9):993–994.
60. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach*. 2011;33(6):478–485. DOI: <https://doi.org/10.3109/0142159X.2011.565828>
61. Wiliam D. What is assessment for learning? *Studies in Educational Evaluation*. 2011;37(1): 3–14. DOI: <https://doi.org/10.1016/j.stueduc.2011.03.001>
62. Castanelli D. Sociocultural learning theory and assessment for learning. *Med Educ*. 2023;57(5):382–384. DOI: <https://doi.org/10.1111/medu.15028>
63. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev*. 1993;100(3):363–406. DOI: <https://doi.org/10.1037/0033-295X.100.3.363>
64. Fogo AB, Kronbichler A, Bajema IM. AI's threat to the medical profession. *JAMA*. 2024;331(6):471–472. DOI: <https://doi.org/10.1001/jama.2024.0018>
65. Van Melle E, Frank JR, Holmboe ES, et al. A core components framework for evaluating implementation of competency-based medical education programs. *Acad Med*. 2019;94(7):1002–1009. DOI: <https://doi.org/10.1097/ACM.0000000000002743>