

CHAPTER 19

Scales, numbers, and words to support entrustment decisions about trainees

Vanessa C. Burch, Jacqueline de Graaf, Jennie B. Jarrett,
Severin Pinilla, Alan Schwartz, Olle ten Cate

Abstract

This chapter focuses on reporting trainee performance in workplace-based assessment programs. Tools that translate observations into assessment documentation typically include several components of observed activities, or competencies, and rating scales with multiple gradations of proficiency. Over the past two decades, introduction of the concept of entrustment and entrustable professional activities has led educators to create scales that focus on the amount of supervision, support, or help trainees need to complete a workplace-based activity. More recently, entrustment–supervision (ES) scales have shifted the focus to reporting readiness for future activities, using specified levels of supervision to be recommended for trainees. This chapter describes the use of these scales for ad hoc and summative entrustment decision-making, including adaptations for profession-specific or context-specific circumstances and variable levels of trainee proficiency.

Scales for entrustment and supervision are more holistic than those for reporting observed behavior, and include both specific and general trainee features. Expert supervisor judgment includes an inherent subjective element, because experts show legitimate differences, but unwanted bias must be excluded. Narrative feedback comments can supplement and explain scale scores providing a holistic picture, guided by tools such as the 'A RICH' framework. Entrustment decisions are crucial in an EPA-based curriculum, and ES

How to cite this book chapter:

Burch VC, de Graaf J, Jarrett JB, Pinilla S, Schwartz A, ten Cate O. Scales, numbers, and words to support entrustment decisions about trainees. 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 19, pp. 225–235. [2024] London: Ubiquity Press. DOI: <https://doi.org/10.5334/bdc.s>

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>

scale values must be embedded in a program of assessment in which feedback conversations with trainees remain essential.

Authors

- Vanessa C. Burch, MD, PhD. Colleges of Medicine of South Africa, Cape Town, South Africa. Correspondence: vanessa.burch@cmsa.co.za
- Jacqueline de Graaf, MD, PhD. Radboud University Medical Center, Nijmegen, the Netherlands.
- Jennie B. Jarrett, PharmD, MMedEd, PhD. University of Illinois Chicago College of Pharmacy, Chicago, Illinois, USA.
- Severin Pinilla, MD, MEd, PhD. University Hospital of Old Age Psychiatry and Psychotherapy, Bern, Switzerland.
- Alan Schwartz, PhD, JD. University of Illinois Chicago, Chicago, Illinois, USA.
- Olle ten Cate, PhD. University Medical Center Utrecht, Utrecht, the Netherlands and University of California, San Francisco, USA.

Introduction

Reporting on the performance of trainees within a program of workplace-based assessment requires tools that translate observations into assessment documentation. This chapter explores the evolving conversation about such documentation, specifically the concept of entrustment with a focus on determining future supervision likely to be required by a trainee or licensed professional, rather than just reporting on currently observed competencies. The chapter also addresses the quantity of workplace-based observations needed, the role of narrative comments in holistic assessment decision-making, and the need to accommodate inherent expert observer subjectivity while remaining cognizant of unwanted bias. Finally, the chapter reflects on feedback that advances trainee autonomy in the workplace.

Traditional workplace-based assessment scales

Many paper forms and, more recently, digital forms have been proposed to document the performance of trainees in the workplace. These include, among many, the mini-CEX, direct observation of procedural skills (DOPS), and case-based discussion.¹ Common scales to assess performance often include three or four general values (below expectations–meets expectations–above expectations; or poor–marginal–good–excellent). Other examples use nine or 10 scale values with a cut-off between five and six for fail and pass. Scales can focus on a single dimension (how did the trainee do in general?) or pertain to several dimensions (history, physical examination, knowledge and reasoning, communication, professional conduct, or other) or combine a series of skills with an overall judgment. Rating forms often include ancillary information (names of ratee and rater, specialty, program year, setting, case complexity) and space for narrative feedback. The core purpose is to record the observed performance and report on trainee performance using the scale provided. As a record of observed performance, this is a retrospective report.

A common concern with these traditional performance assessment scales is eloquently illustrated by Crossley when he says:

*[M]y judgement about the performance of my trainee, based on my interpretation of his performance, with a particular patient or client, with a particular problem, in a particular context today, is always highly meaningful, [but does] this judgement have anything to say about my judgement with the same trainee in a completely different context, or anyone's judgement about the same trainee in any context?*²

As Crossley contends, maybe we are asking the wrong questions to expect a reliable answer³ and we should instead ask ‘is this trainee ready for more autonomy?’⁴

Entrustment–supervision scales

Over the past two decades, introduction of the concept of entrustment and EPAs has led educators to create scales focusing on the amount of supervision, support, or help needed to complete an activity in the workplace. Entrustment–supervision (ES) scales differ from traditional scales in their focus on (a) entrustment with tasks and (b) level of supervision or support required.^{5,6} ES scales reflect a shift in focus from the pursuit of ‘objective proficiency ratings,’ using predominantly numerical scales with brief anchoring statements, to decision-making about the level of supervision trainees require to safely complete workplace-based activities. From the perspective of clinical training, autonomy can be described in terms of a required level of supervision. The

most generic entrustment–supervision scale includes five levels: (1) observe only, (2) act under direct supervision, (3) act under indirect supervision, (4) act unsupervised, and (5) act as a supervisor. The literature provides many variations of ES scales.⁵

ES scales can be retrospective or prospective. While supervision is factored into both scales, the first regards the supervision or help provided during an activity, and is not particularly focused on entrustment regarding a future activity. In contrast, prospective ES scales focus on readiness to trust future performance at a specified level of supervision.⁵ They ask supervisors to estimate readiness using a scale based on autonomy. These scales, and variations thereof, frame the judgment to include an estimation of readiness and risks associated with entrustment. Paired with narrative feedback, ES scales can highlight specific actions or gaps that are relevant when considering entrustment. Prospective ES scales are a good preparation for summative entrustment decisions, which should be based on multiple evaluations of observations and case-based discussions. Table 19.1 shows published examples of retrospective and prospective ES scales.

Prospective ES scales, which focus on *decisions* of entrustment, also differ from proficiency scales by their ordinal nature. Entrustment decisions reflect discrete steps toward granted autonomy and are not a continuous scale of ability. Note that, when entrustment decisions about increased trainee autonomy do not lead to the actual granting of increased autonomy (‘To what extent *would* I trust the trainee with Task X, in theory only’),¹⁰ they are better called ‘entrustment determinations.’¹¹ Entrustment determinations are problematic in that they run the risk of reducing ordinal ES scales to continuous proficiency scales of trustworthiness, and just add more scales to the existing ones. True entrustment scales with concrete consequences cannot be continuous, because the decision to entrust a trainee with a task is a discrete act.

Working with prospective ES scales is not easy, either for clinicians¹² or for members of a clinical competency committee (CCC)¹³ because it requires a deeper understanding of trainees than is needed when just documenting an observation.¹⁴ However, entrustment decisions should not be avoided just because they are difficult to make. To ask a supervisor about the readiness of a trainee for less supervision involves weighing benefits and risks,¹⁵ which requires thoughtful consideration. Preparing trainees for a qualification or license to practice (for an EPA or a full profession respectively) is a key role of educators and programs, and it is exactly what such entrustment decisions entail. It may be easy to give a score for observed proficiency but hard yet important to estimate the consequences of entrustment for patient care.

Critical in ES scale use is the distinction between ad hoc and summative decisions of entrustment. Ad hoc judgments and decisions, made and evaluated by individual supervisors for

Table 19.1: Examples of retrospective and prospective entrustment–supervision scales.

	Explanation	Examples used with direct observation
Retrospective entrustment–supervision scale (examples are the O-SCORE ⁷ or OCAT ⁸ scale)	This scale uses supervision levels to indicate how much support was provided in an observed performance	<ol style="list-style-type: none"> 1. I had to do it 2. I had to talk them through 3. I had to prompt them from time to time 4. I needed to be there just in case 5. I did not need to be there
Prospective entrustment–supervision scale	This scale uses supervision levels to indicate how much supervision the trainee should receive in future performances of a similar activity ⁹	<p>Based on this observation, my judgment is that this trainee, for this activity, is:</p> <ol style="list-style-type: none"> 1. not yet ready for direct supervision 2. ready for direct, proactive supervision 3. ready for indirect, reactive supervision 4. ready to perform without supervision 5. ready to supervise junior trainees

specific patient care encounters in the workplace, lead to reports and scale use, as the examples in Table 19.1 show. Summative decisions are made by a team—often a CCC in postgraduate medical education or entrustment committee in undergraduate education—building on a variety of information sources, and are true decisions for qualification and privileging.⁹ Following principles of programmatic assessment, summative entrustment decisions must be based on multiple data points. Permission to practice health care and execute specified EPAs under indirect supervision is an example that could apply to senior medical students, aligning well with the autonomy of a medical license.

Context-specificity of entrustment–supervision scales

Various ES scales and, more generally, levels of supervision, can deviate from the common, original framework of five levels (Table 19.2, column 1). Depending on the context, specifications or adaptations of this generic framework are useful. A modified, expanded scale created by Chen et al.¹⁶ for undergraduate medical education is frequently used (Table 19.2, column 2). This expanded scale is useful for reporting on trainees' progress in the early stages of training. Similarly, the scale created by Jarrett et al. (Table 19.2, column 3) utilizes the expanded model further modified for pharmacy trainees, by noting the physical location of the supervisor to quickly orient them to the use of the scale in relation to the performance of the trainee.¹⁷ Expansion allows faculty to better

Table 19.2: Generic and expanded entrustment–supervision scales.

Generic ES scale	Expanded scales	
	Chen-modified ES scale ¹⁶	Jarrett-modified ES scale ¹⁷
1. Not allowed to practice the EPA	1a. Not allowed to observe 1b. Allowed to observe	1a. Would not trust, not allowed to observe 1b. Trust to thoughtfully observe, has foundational knowledge
2. Allowed to practice under direct, proactive supervision (supervisor in the room)	2a. As co-activity with supervisor 2b. With supervisor in room ready to step in as needed	2a. Trust to perform task with the supervisor, requiring direction, guidance and help 2b. Trust to perform task with the supervisor present and ready to step in and is new in performing the task alone
3. Allowed to practice under indirect, reactive supervision (supervisor not in the room)	3a. With supervisor immediately available; all findings/decisions double-checked 3b. With supervisor immediately available; key findings/decisions double-checked 3c. With supervisor distantly available; findings/decisions reviewed on request	3a. Trust to perform task with supervisor nearby and all findings are immediately checked 3b. Trust to perform task with supervisor nearby and key findings are immediately checked 3c. Trust to perform task with supervisor remote, but on demand and work is audited soon after completion
4. Allowed to practice unsupervised	4a. With supervisor not around but available by phone 4b. With supervisor not available by phone	4. Trust to perform task independently and unsupervised
5. Allowed to act as supervisor for more junior trainees		5. Trust to perform task independently as well as supervise and teach others

represent their opinion about supervision needs in the future and provides ample opportunity to document progress of junior trainees. Since graduates of several health professions become licensed to enter full unsupervised clinical practice upon graduation, with no further postgraduate training requirements, this expanded scale example accurately describes assessment of the work completed.

Many discipline-specific retrospective ES scales are in use internationally. Examples include the ‘Zwisch scale,’ used with the mobile app SIMPL in surgical education—the scale has four values: ‘show and tell,’ ‘active help,’ ‘passive help,’ and ‘supervision only.’¹⁸ It is somewhat similar to the O-SCORE scale, developed for postgraduate surgery training in Canada,⁷ which uses the values one to five as depicted in the upper right cell of Table 19.1. An ES scale used in one anesthesia program frames supervision as the duration a supervisor can leave the operating theater (for coffee, for lunch, etc.).¹⁹ A triple-tool scale in pathology using procedures, situations, or reporting²⁰ and a retrospective ES scale under consideration in internal medicine uses the supervisor’s expectation as a benchmark (‘to ensure safe care I (1) had to step in much more than expected, (2) stepped in a little more than expected, (3) provided usual supervision, (4) could step back a little more than usual, (5) stepped back much more than usual’).^a Stepping in (‘hands-on’) and out (‘hands-off’) also depends on the context, the trainee, and the tasks expected to be performed in a given role.²¹

In other words, clinical specialties need to adapt ES scales depending on the nature of the work and expectations of the workplace. For example, level 2 entrustment (supervisor present during the activity) is infrequently reported for junior internal medicine residents, while this level predominates in surgical specialties, where even senior trainees spend many hours under direct supervision in the operating room,²² and may be characterized by different steps, such as permission to do part of an operation (‘Open entry to the abdomen’ and ‘Fascial and abdominal closure’) as early EPAs, nested later within full surgery EPAs, or to act with a supervisor present but unscrubbed, which could be characterized as a new level 2c (supervisor present to instruct, but not able to step in immediately). It is recommended that, for each specialty and profession, a logical scale of supervision and entrustment be established to reflect a stepwise, deliberate increase in autonomy based on relevant, documented observations. This requires a shared mental model of ES scales by both supervisors and trainees.

Making trustworthy entrustment decisions

Since the 20th century, education in the health professions has pursued objectivity over subjectivity in assessment.²³ Assessment outside the workplace, including standardized testing of canonical (i.e., non-disputed) knowledge and skills, should meet traditional reliability criteria. Workplace-based assessment, however, which focuses on the application of knowledge and skills in patient- and context-specific practice, often cannot meet the criteria established for ‘high-stakes’ standardized tests, in particular reliability criteria proposed in classical test theory. As a result, no single instrument will meet the goals of workplace-based assessment—to accurately identify areas for trainee development, work collaboratively with trainees through self-reflection, and make decisions about future practice. Instead, robust workplace-based assessment requires the integration of different types of assessments involving multiple observations by multiple observers in varying contexts. Because learning is not linear, these multiple assessments may cause confusion when a trainee with generally positive assessments in one context performs poorly in another context. CCCs must determine how to make summative decisions with variation in workplace-based

^a Personal communication Stephen Gauthier and David Taylor.

assessment, considering the upcoming tasks the trainee will face, their agility to work in various contexts of practice, with associated risks, and the support available for development (for more details see Chapter 21).

Although programmatic assessment emphasizes collecting many observations, the quality of assessment is not necessarily driven by the quantity of assessments. One observation by an experienced clinician who is trained in assessment and knows the trainee well can be more trustworthy than several observations by junior faculty who are not acquainted with the trainee and have little time or motivation for trainee assessment.²⁴ In competency-based education, success is measured by demonstrating performance and ability, not just the number of performances nor amount of time in training. Programs should evaluate what types, frequency and modes of assessment will best support decision-making in summative evaluations for their profession, practice, and context, and find a balance between quantity and quality.

Subjectivity and bias in observational judgment

Workplace-based assessment focuses on the contextual competence of trainees. Unlike much of the assessment of canonical competence, determining the contextual competence of trainees requires expert judgment—the opinion of experienced professionals who can appreciate the importance of their judgments, evaluate performance in context, and facilitate reflection and discussion with the trainee for growth (see Chapter 2 for a discussion of this multilayered competence model). One expert's judgment, based on their personal clinical and supervisory experiences, will show similarities but also differences with other experts' judgments. In other words, experts develop opinions that are both meaningfully 'subjective' and highly relevant.²⁵ This subjectivity is characterized by differences with other experts, which in the past has been considered 'error variance' or 'bias.' Valuing subjectivity also means acknowledging that some judgments cannot be fully expressed in numerical scales.^{26,27} CCCs should, therefore, take into account both rating of readiness for autonomy and supportive narrative information, and at the same time acknowledge that contextual competence implies the ability to work in particular contexts, which may require different knowledge and skills that cannot always be characterized as generally 'right' or 'wrong.'

On the other hand, unwanted bias, or subjectivity that stems from prejudice that is unrelated to the proficiency of the trainee or readiness for a decrease in supervision, must be avoided. Psychometrically this is regarded as 'construct-irrelevant variance.'²⁸ The remaining legitimate subjective judgment must be retained. A recent AI analysis of narrative feedback highlighted the presence of unintended bias,²⁹ underscoring the need for observer vigilance and the importance of specifically addressing unconscious bias when training observers to give feedback. Although individual supervisors might find it hard to recognize their own bias,³⁰ the exchange of subjective judgments in a team, such as a CCC, may help to keep unwanted bias off the table.

Holistic evaluation of trainees

Most ES rating forms contain both numbers and words because numbers alone lack contextual detail. Narrative information can be distinguished in brief comments generated by direct observation, and more elaborate narratives, such as those generated by multisource feedback procedures (see Chapter 17). Together, integrated and synthesized with ratings, they provide the necessary story-type information needed to inform decisions made by CCCs. These committees need to take various professional attributes into account, besides clinical and technical proficiency, to allow for

entrustment decisions. The A RICH framework provides an overview of these attributes in five categories: agency, reliability, integrity, capability, and humility.³¹ The recommendation is not to translate the A RICH framework into a rating scale but to consider these attributes when writing narrative comments.

Words captured on entrustment rating forms serve two distinct purposes: (a) summatively, they contribute toward promotion decisions for trainees, and/or (b) formatively, they provide developmental feedback to trainees.³² Word choice in narrative comments is important and should be aligned with the primary purpose of the assessment event, which should be clarified with the trainee beforehand.

A challenge of combining entrustment ratings and narrative text within a program of assessment is the accumulation of an overwhelming volume of data. This is where technology, with e-portfolios and mobile tools, may be of great help by assisting with data collection and aggregation to provide accessible overviews of trainee development. For more information see Chapter 20. And not everything needs to be documented. If supervisors concisely document narrative information, they also have time for undocumented, psychologically safe, oral communication with trainees.

Feedback

High-quality, meaningful feedback may be enhanced by a focus on EPAs and ES scales. However, trainee and supervisor perspectives may differ. The benefit of ES scales from a supervisor's perspective stems from the connection of direct observation in the workplace with a decision about readiness for autonomy. The trainee perspective may be complicated by blurring the value of feedback for further growth and development with judgment decisions about competence, the more traditional view of assessment.³³

It is important to agree on the purpose of an observation followed by feedback. Since the latter should help a trainee improve, the search and focus on inadequacies, which does not directly feed intrinsic motivation³⁴ should be avoided. This can turn workplace-based assessment events into stressful and burdensome moments. Since giving and receiving feedback is a complex interpersonal process, supervisors and trainees need training (see Chapter 23).

A useful starting point is a definition of helpful feedback, which frames it as a 'supportive conversation that clarifies the trainee's awareness of their developing competencies, enhances their self-efficacy for making progress, challenges them to set objectives for improvement, and facilitates their development of strategies to enable that improvement to occur.'³⁵ Considering that entrustment decisions may also be informed by interprofessional feedback, guidelines for such feedback may be taken into account (Table 19.3).

Table 19.3: Guidelines for interprofessional feedback using the Westerfeld framework.³⁶

Open, respectful	Participants are open to each other's input and communicate respectfully
Relevant	Agreed-upon goals for observed performance (e.g., EPAs)
Timely	Feedback is given in a phase of learning (hours or days, not months)
Dialogical	The conversation reflects a two-way communicative exchange
Responsive	The feedback is adapted to specific context at stake
Sense making	Feedback is explored and elaborated as needed to help trainees make sense of it
Actionable	Feedback contributes to its usability and leads to concrete action plans

Other useful feedback models are ask-and-tell approaches, self-assessment with encouragement and direction, or coaching (R2C2: relationship–reaction–content–coaching) frameworks.³⁷ While frameworks are useful to structure feedback conversations, there is no quick fix to generate helpful feedback that fulfills the promise of assessment for learning. A persistent complaint heard in clinical education is a lack of true constructive feedback, a complaint that should force supervisors to remain attentive toward a genuine interest in trainee development and progression toward entrustment. Supervisors need to consider the nature of effective communication with varying personalities and develop an understanding of how feedback is given and received.

Conclusion

Entrustment–supervision scales are useful for several reasons. They force observers to think about trainee readiness for patient care activities, support CCCs in making decisions about trainees' progress and summative entrustment, and serve as a focus for feedback to trainees. They also inform trainees about what is expected of them to achieve more autonomy and less supervision. Trainees have the responsibility to work on their progress in autonomy and reflect on it. Self-reflection by trainees helps supervisors understand what major gaps or problems are limiting trainees from achieving greater autonomy. ES scales are needed for making decisions about trainees' progress but scales alone are not enough. Narrative information gives much-needed depth and rich information to support such decisions. Finally, in the everyday workplace environment it is also critical to cherish the regular occurrence of undocumented, psychologically safe, oral communication in daily activities between supervisors and trainees.

Competing interests

The authors declare that they have no competing interests.

References

1. 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>
2. Crossley J. Validity and truth in assessment. *Med Educ*. 2013;47(12):1152–1154. DOI: <https://doi.org/10.1111/medu.12317>
3. 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>
4. Weller JM, Misur M, Nicolson S, et al. Can I leave the theatre? A key to more reliable workplace-based assessment. *Br J Anaesth*. 2014;112(6):1083–1091. DOI: <https://doi.org/10.1093/bja/aeu052>
5. 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>
6. Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med*. 2016;91(2):186–190. DOI: <https://doi.org/10.1097/ACM.0000000000001045>
7. 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>

8. Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A new instrument for assessing resident competence in surgical clinic: the Ottawa Clinic Assessment Tool. *J Surg Educ.* 2016;73(4):575–582. DOI: <https://doi.org/10.1016/j.jsurg.2016.02.003>
9. 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>
10. ten Cate O, Schumacher DJ. Entrustable professional activities versus competencies and skills: exploring why different concepts are often conflated. *Adv Health Sci Educ Theory Pract.* 2022;27(2):491–499. DOI: <https://doi.org/10.1007/s10459-022-10098-7>
11. ten Cate O, Jarrett JB. *Would I trust or will I trust?* The gap between entrustment determinations and entrustment decisions for trainees in pharmacy and other health professions. *Pharmacy (Basel).* 2023;11(3). DOI: <https://doi.org/10.3390/pharmacy11030107>
12. Postmes L, Tammer F, Posthumus I, Wijnen-Meijer M, van der Schaaf M, ten Cate O. EPA-based assessment: clinical teachers' challenges when transitioning to a prospective entrustment–supervision scale. *Med Teach.* 2021;43(4):404–410. DOI: <https://doi.org/10.1080/0142159X.2020.1853688>
13. 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 committees. *Med Educ.* December 13, 2023. DOI: <https://doi.org/10.1111/medu.15296>
14. Jonker G, Ochtman A, Marty AP, Kalkman CJ, ten Cate O, Hoff RG. Would you trust your loved ones to this trainee? Certification decisions in postgraduate anaesthesia training. *Br J Anaesth.* 2020;125(5):e408–e410. DOI: <https://doi.org/10.1016/j.bja.2020.07.009>
15. ten Cate O. Managing risks and benefits: key issues in entrustment decisions. *Med Educ.* 2017;51(9):879–881. DOI: <https://doi.org/10.1111/medu.13362>
16. Chen HC, van den Broek WES, ten Cate O. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med.* 2015;90(4):431–436. DOI: <https://doi.org/10.1097/ACM.0000000000000586>
17. Jarrett JB, Goliak KL, Haines ST, Trolli E, Schwartz A. Development of an entrustment-supervision assessment tool for pharmacy experiential education using stakeholder focus groups. *Am J Pharm Educ.* 2022;86(1):8523. DOI: <https://doi.org/10.5688/ajpe8523>
18. 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>
19. Weller JM, Castanelli DJ, Chen Y, Jolly B. Making robust assessments of specialist trainees' workplace performance. *Br J Anaesth.* 2017;118(2):207–214. DOI: <https://doi.org/10.1093/bja/aew412>
20. Bryant BH, Anderson SR, Brissette M, et al. Leveraging faculty development to support validation of entrustable professional activities assessment tools in anatomic and clinical pathology training. *Acad Path.* 2024;11(2):100111. DOI: <https://doi.org/10.1016/j.acpath.2024.100111>
21. Hatala R, Ginsburg S, Gauthier S, Melvin L, Taylor D, Gingerich A. Supervising the senior medical resident: entrusting the role, supporting the tasks. *Med Educ.* 2022;56(12):1194–1202. DOI: <https://doi.org/10.1111/medu.14883>
22. Sandhu G, Thompson-Burdine J, Nikolian VC, et al. Association of faculty entrustment with resident autonomy in the operating room. *JAMA Surg.* 2018;153(6):518–524. DOI: <https://doi.org/10.1001/jamasurg.2017.6117>
23. van der Vleuten CPM, Norman GR, De Graaff E. Pitfalls in the pursuit of objectivity: issues of reliability. *Med Educ.* 1991;25:110–118.
24. Kulasegaram K, Paradis E. Better data >> Bigger data. *Med Teach.* 2014;36(11):1008–1009. DOI: <https://doi.org/10.3109/0142159X.2014.917761>

25. ten Cate O, Regehr G. The power of subjectivity in the assessment of medical trainees. *Acad Med*. 2019;94(3):333–337. DOI: <https://doi.org/10.1097/ACM.0000000000002495>
26. van Enk A, ten Cate O. ‘Languaging’ tacit judgment in formal postgraduate assessment: the documentation of ad hoc and summative entrustment decisions. *Perspect Med Educ*. 2020;9(6):373–378. DOI: <https://doi.org/10.1007/s40037-020-00616-x>
27. 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>
28. Thomas MH, Steven MD. Construct-irrelevant variance in high-stakes testing. *Educ Meas Issues Pract*. 2004:17–27.
29. Gin BC, ten Cate O, O’Sullivan PS, Boscardin C. Assessing supervisor versus trainee viewpoints of entrustment through cognitive and affective lenses: an artificial intelligence investigation of bias in feedback. *Adv Health Sci Educ Theory Pract*. February 23, 2024. DOI: <https://doi.org/10.1007/s10459-024-10311-9>
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.0000000000005251>
31. 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>
32. Ginsburg S, Watling CJ, Schumacher DJ, Gingerich A, Hatala R. Numbers encapsulate, words elaborate: toward the best use of comments for assessment and feedback on entrustment ratings. *Acad Med*. 2021;96(7S):S81–S86. DOI: <https://doi.org/10.1097/ACM.0000000000004089>
33. 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>
34. ten Cate OTJ. Why receiving feedback collides with self determination. *Adv Health Sci Educ Theory Pract*. 2013;18(4):845–849. DOI: <https://doi.org/10.1007/s10459-012-9401-0>
35. Lefroy J, Watling C, Teunissen PW, Brand P. Guidelines: the do’s, don’ts and don’t knows of feedback for clinical education. *Perspect Med Educ*. 2015;4(6):284–299. DOI: <https://doi.org/10.1007/s40037-015-0231-7>
36. Tielemans C, de Kleijn R, van der Schaaf M, van den Broek S, Westerveld T. The Westerveld framework for interprofessional feedback dialogues in health professions education. *Assess & Eval in Higher Educ*. 2023;48(2):241–257. DOI: <https://doi.org/10.1080/02602938.2021.1967285>
37. Liakos W, Keel T, Ellen Pearlman R, Fornari A. Frameworks for effective feedback in health professions education. *Acad Med*. 2023;98(5):648. DOI: <https://doi.org/10.1097/ACM.0000000000004884>