Entrustment Unpacked: Aligning Purposes, Stakes, and Processes to Enhance Learner Assessment

Benjamin Kinnear, MD, MEd, Eric J. Warm, MD, Holly Caretta-Weyer, MD, MHPE, Eric S. Holmboe, MD, David A. Turner, MD, Cees van der Vleuten, PhD, and Daniel J. Schumacher, MD, PhD, MEd

Abstract

Educators use entrustment, a common framework in competency-based medical education, in multiple ways, including frontline assessment instruments, learner feedback tools, and group decision making within promotions or competence committees. Within these multiple contexts, entrustment decisions can vary in purpose (i.e., intended use), stakes (i.e., perceived risk or consequences), and process (i.e., how entrustment is rendered). Each of these characteristics can be conceptualized as having 2 distinct poles: (1) purpose has formative and summative, (2) stakes has low and high, and (3) process has ad hoc and structured. For each characteristic, entrustment decisions often do not fall squarely at one pole or the other, but rather lie somewhere along a spectrum. While distinct, these continua can, and sometimes should, influence one another, and can be manipulated to optimally integrate entrustment within a program of assessment. In this article, the authors describe each of these continua and depict how key alignments between them can help optimize value when using entrustment in programmatic assessment within competency-based medical education. As they think through these continua, the authors will begin and end with a case study to demonstrate the practical application as it might occur in the clinical learning environment.

Maja is an intern on the final day of a rotation with her attending, Dr. Schurner, who will be giving feedback using the program’s entrustment-based assessment form. Maja is unsure what “entrustment” is or how these ratings will affect her progression through the residency program. The grade-centric environment of Maja’s medical school conditioned her to understand assessment as high stakes, focusing on the resulting consequences of a grade rather than the potential use for growth and development. She is also unaware of what skills have been assessed, who else has been involved in her assessment, and who is going to see her assessment data. Dr. Schurner, on the other hand, thinks that this conversation will be low stakes, with a primary goal of using the feedback to improve. Dr. Schurner believes he has put a significant amount of deliberate thought into his entrustment decisions, but Maja believes that attending assessments are “just their opinions,” and she is worried that one person’s assessment could have lasting effects on her residency experience and how she is viewed by others. Maja and Dr. Schurner have misaligned views on the purpose, stakes, and process of entrustment decisions in this moment. How could this problem be avoided?

Background

Medical education has shifted in recent decades to an intense focus on outcomes-based approaches instead of exclusively time-based, process-focused models. Competency-based medical education (CBME), the preeminent model for outcomes-based medical training, currently serves as the teleological foundation for multiple transformational national education projects. CBME strives to meet the needs of patients, learners, and other stakeholders by starting with the desired health and health care outcomes (e.g., the triple aim), determining the necessary competencies to prepare trainees to meet those health care outcomes, and designing curricula and assessment systems to meet these goals.

The centrality of valid measurement of learner outcomes makes assessment crucial to CBME implementation. van der Vleuten and others have advocated for programmatic approaches to assessment in which educators deliberately combine multiple instruments, assessors, and methods within a coordinated program. When measuring learner performance, no single data point or assessment method needs to be infallible, and the combination of data and methods within a program can be greater than the sum of the parts. Integrating varying assessment methods throughout a program can provide a more complete picture of performance and may also be better suited for assessing learners across differing contexts (e.g., clinical rotations, assessors, patient care tasks).

Programmatic assessment that employs differing assessment methods should have a unifying framework, and entrustment has gained significant traction as a framework in CBME. Entrustment emerged from the recognition that health care is rooted in public trust of physicians to deliver care that meets the high bar of the Institute of Medicine’s 6 domains of health care quality, and that this dynamic can be translated into a framework for assessment. The social contract (trust) undergirding entrustment aligns with the sociocultural public accountability movement that in part led to the rise of CBME. In that sense, entrustment intends to bring the quality and safety of patient care explicitly into assessment. Trust is an essential...
dynamic between supervisor and trainee in clinical environments, and use of entrustment in assessment may promote construct alignment between assessor and task.\textsuperscript{23,24} Entrustment can be used in multiple ways throughout a program of assessment, including frontline assessment decisions,\textsuperscript{25} back-end aggregation, or decisions at the level of a promotions or competence committee.\textsuperscript{26}

With such a diversity of potential applications, it is important that the concept of entrustment be used in a manner that fits the situation. Every entrustment decision can vary in its purpose (i.e., intended use), stakes (i.e., perceived risk or consequences), and process (i.e., how entrustment is rendered). Each of these characteristics can be conceptualized as having 2 distinct poles: (1) purpose has formative and summative, (2) stakes has low and high, and (3) process has ad hoc and structured. For each characteristic, entrustment decisions often do not fall squarely at one pole or the other, but rather lie somewhere along a spectrum. Thus, we use the model of a continuum to represent how entrustment decisions can vary in their proximity to each pole in different situations (Figures 1–3). The concepts of formative, low stakes, and ad hoc are often used together, as are summative, high stakes, and structured. However, formative assessment is not inherently low stakes or ad hoc, and summative is not inherently high stakes or structured. These 3 characteristics are separate, related continua that can each be manipulated to optimally integrate entrustment within a program of assessment. While distinct, these characteristics can, and sometimes should, influence one another. In this article, we define these 3 continua, describe how entrustment decisions can vary across them, and propose key alignments between them.

**Figure 1** This represents an ad hoc entrustment decision made by a supervisor and demonstrates a discrepancy between the perception of the purpose and stakes for the learner (1a) and supervisor (1b). In this scenario, the supervisor (1b) believes the entrustment decision is low stakes and serves a predominantly formative purpose, while the learner (1a) sees the decision as higher stakes and predominantly summative. This discrepancy can occur despite both understanding the entrustment decision was made using an ad hoc process.

**Purpose: From “Formative” to “Summative”**

While the literature covers the formative/summative continuum widely, the exact meanings of these terms are not always well defined, which may lead to confusion.\textsuperscript{27,28} The difference between formative and summative was first described in 1967 in the context of program evaluation\textsuperscript{29} and later for individual learners.\textsuperscript{30} The initial distinction focused on the process of assessment, with summative being a single process of making a judgment and formative involving the same summative process while adding a subsequent step of providing feedback to the learner.\textsuperscript{31} Later conceptualizations, which are widely held today, posit that the function, or intended use, of assessment distinguishes formative (to provide feedback) from summative (to render judgment).\textsuperscript{32} We agree that the distinction between formative and summative lies in the purpose or intended use of assessments as reflected in contemporary CBME vernacular, which has shifted away from “formative/summative” toward “assessment for learning” and “assessment of learning,” respectively.\textsuperscript{11,14,33–36} The oft-emphasized *<i>of</i>*/<i>for</i>* distinction draws a contrast in intent, and hence purpose.

CBME holds that programmatic assessment should have robust formative components rather than simply being summative in nature.\textsuperscript{4,10,14,26} Most assessments serve multiple purposes to varying degrees, falling along a continuum between formative and summative. Individual frontline assessments often have a predominantly formative purpose and are frequently aggregated over time for summative purposes. This dual purposing of assessment can make the intent unclear for learners\textsuperscript{37} and create a mismatch between educators’ intended assessment purposes and learners’ perceptions, with learners focusing on the summative aspects and potentially losing the formative benefits of the assessment.\textsuperscript{38–42} Setting clear expectations and providing strong learner support, such as with a longitudinal coach, is crucial to maximize the benefits of both purposes in CBME.\textsuperscript{43}
Like all assessments, entrustment can serve multiple purposes that fall somewhere along the summative/formative continuum. Entrustment decisions can have a predominantly summative purpose such as documenting and monitoring progress toward readiness for unsupervised practice. Some programs have developed entrustment-based assessment tools intended for more formative purposes to guide learner development. Using entrustment of specific skills and activities as a framework for feedback may facilitate formative conversations. For example, Warm et al use frontline entrustment decisions in a predominantly formative fashion, though they are aggregated to inform later summative decisions. The interdependent nature of performance within a health care team adds complexity to the use of entrustment for either formative or summative purposes at an individual level and is an area that deserves further investigation.

**Stakes: From “Low” to “High”**

In the context of CBME, stakes are the consequences that follow an assessment,
and these fall on a continuum from low to high.\textsuperscript{54,55} The concept of risk is inherent to the term \textit{stakes}, implying there is something to be lost or an opportunity to be missed.\textsuperscript{56} While risk is also inherent to the construct of trust,\textsuperscript{57\textendash}66 there is not a simple relationship between stakes and entrustment. There are multiple stakeholders in entrustment decision making, including patients, learners, supervising attendings, programs, and regulatory bodies. Therefore, one must specify: “stakes for whom?” Patients may have the highest stakes in entrustment decisions, as they arguably take the highest risk, being vulnerable to care from trainees and all health care professionals. Attendings make themselves vulnerable when trusting trainees to provide care under their supervision, and training programs face risk in being one of the primary gatekeepers to public safety when determining readiness for promotion or graduation. Regulatory bodies, whose missions focus on care safety and quality, must ultimately answer to the public through accreditation of training programs and certification of individual practitioners.

Stakes of entrustment decisions for learners are often conceptualized as being connected with the aforementioned continuum of purpose through phrases such as “low-stakes, formative” and “high-stakes, summative” as if stakes and purpose are necessarily linked.\textsuperscript{67} However, we suggest this approach is an oversimplification. If stakes are related to risk, then they may depend in part on the related consequences in a quantitative or qualitative manner. Quantitatively, stakes may relate to the proportion of contribution to a final grade or decision. This means that frontline entrustment-supervision ratings that have a larger influence on eventual summative decisions are inherently higher stakes. For example, if an entrustment rating contributes only 1\% to an eventual summative decision, then it is lower stakes than the one that contributes 10\%. In this sense, stakes are related to the structure of a program of assessment and can be lowered by increasing the overall number of assessments that contribute to entrustment decisions. Qualitatively, stakes may relate to the type of consequences, such as permission to progress through a training program, opportunities to participate in certain patient care experiences, or future training or job opportunities that depend on performance ratings. For example, decisions related to readiness for graduation may be inherently higher stakes than those related to supervision in a single clinical context, as the former has further-reaching consequences for the learner.

There is a growing body of literature that the stakes of assessment are not only inherent to subsequent consequences, but also the learner’s individual sense of control within a program of assessment.\textsuperscript{68} Schut et al found that stakes were perceived as lower when learners had opportunities to influence outcomes (e.g., opportunity to interact with the assessor), to collect and control data (e.g., initiate an assessment and control who sees it), and to show evidence of improvement over time.\textsuperscript{54} The teacher–learner relationship also plays a prominent role in learner perception of assessment stakes.\textsuperscript{64,62} If teachers emphasize their \textit{positional agency}, or authority, without also fostering \textit{high communion}, or personal relationships, then learners are more likely to see assessment as high stakes and hide their weaknesses to maintain a cloak of competence.\textsuperscript{62\textendash}64 These studies suggest that decisions viewed as lower stakes by a program may be viewed as higher stakes by learners if they do not feel in control of their learning experience. Thus, the continuum of stakes may be concordant or discordant between various stakeholders for which stakes can be considered.

Ratings scales used with entrustment decisions may also impact the stakes. Most entrustment-related scales use varying levels of supervision, which make the real-world consequences of ratings readily apparent to learners.\textsuperscript{65} This approach may inherently raise the stakes over other assessment frameworks that use narrative-based or simple numerical anchors that are not explicitly attached to consequences. The stakes may also vary by location of ratings along entrustment-supervision scales. These scales are ordinal in nature, and therefore do not constitute a true continuum with fixed intervals between levels.\textsuperscript{66} Aylward et al suggest that scores falling at the extremes such as Level 1 (unable to perform) or Level 4 (unsupervised practice) are inherently higher stakes than Levels 2 (direct supervision) or 3 (indirect supervision) as there is little difference “in practice” between the middle scores.\textsuperscript{67} Indeed, learners who receive Level 1 entrustment-supervision ratings may be more likely to receive formal remediation plans,\textsuperscript{68} which could be perceived by learners as higher stakes.

The very nature of entrustment may also impact the stakes. Multiple factors can affect entrustment,\textsuperscript{47,69\textendash}82 with a central component being learner “trustworthiness,” which is a multidimensional construct. Ability (e.g., knowledge, skill) is only one dimension of trustworthiness. Other dimensions include discernment, truthfulness, integrity, reliability, humility, conscientiousness, and benevolence.\textsuperscript{57\textendash}60\textsuperscript{,84} While an assessor might infer these attributes from behaviors, they are part of a learner’s internal professional identity\textsuperscript{65} that some argue are not observable.\textsuperscript{66} Perhaps this idea is why Gingerich wrote that trust, “can be assumed, inferred, felt, created, discovered, earned and lost, but not observed.”\textsuperscript{87} Pingree et al described behaviors that might actively demonstrate various domains of trustworthiness, such as “Recognize illness severity” or “Ask for help when unsure.”\textsuperscript{88} Whether these attributes of trustworthiness can truly be assessed by observing behavior is a question without definitive answer. Regardless, bringing such personal, questionably observable dimensions into competency-based assessment, even if done implicitly rather than explicitly, may affect the stakes for learners by attempting to assess aspects of their professional identities.

\textbf{Process: From “Ad Hoc” to “Structured”}

The third continuum focuses on process, where decisions range between \textit{ad hoc} and \textit{structured}.\textsuperscript{67} Ad hoc entrustment decisions are part of day-to-day clinical interactions between supervisors and learners. They are highly contextual and influenced by factors related to the supervisor, task, supervisor–learner relationship, clinical systems, environment, and patient.\textsuperscript{47,60\textendash}74 Ad hoc entrustment is often implicit, in the moment, and may or may not be captured by an assessment instrument. Structured decisions are made in a more systematic way and involve more reflection, deliberateness, and integration.

\textsuperscript{559}}
into a program of assessment. Supervisors make ad hoc entrustment decisions daily (e.g., implicitly determining the amount of supervision a learner needs to perform a procedure), but these decisions can become more structured through reflection on learner performance and deliberate expression of entrustment-supervision ratings on frontline assessment instruments. Entrustment can be made even more structured by informing judgments with robust, multisource assessment data and using group processes to arrive at decisions. An example of this would be a promotions or competence committee making entrustment decisions after review of longitudinal assessment data.

Fitting Entrustment to the Situation
When entrustment decisions are misaligned along these 3 continua or terms are conflated, pitfalls in CBME may emerge. The goal should be to find the best combination along these continua to produce the greatest good for both patients and learners over time. Figures 1-3 show sliders representing where an assessment may fall on each continuum in differing circumstances. Movement along continua is constant and expected in different situations. Three relationships between these continua deserve specific attention and include aligning: process with purpose, stakes with purpose, and process with stakes.

Aligning process with purpose
Programmatic assessment in CBME often uses the same data for formative and summative purposes, and the lack of clarity can be detrimental to assessment driving learning.43,48 The dominant purpose of any entrustment decision should be explicitly articulated to learners and assessors,52 with appropriate processes being used. It should not be assumed, as it often is, that all parties will perceive ad hoc entrustment to be formative and structured entrustment to be summative. Figure 1 shows a misalignment between the perspectives of a learner and assessor regarding the purpose of an ad hoc entrustment decision. From an education standpoint, ad hoc entrustment decisions should most often have a predominantly formative purpose, particularly given the relatively low reliability68 and the highly contextual nature of a single supervisor's entrustment rating 69,70,73,84. Linking ad hoc entrustment decisions with longitudinal coaching can reinforce the formative intent.11,17,43 Entrustment decisions that are more summative in purpose should be made in a systematic way using more structured processes. If frontline entrustment decisions are collated to inform larger summative decisions, supervisors should use a structured, deliberate process to report their assessments. Summative entrustment decisions that will affect future levels of supervision provided to a learner should be highly structured, informed by myriad data points from multiple assessors and contexts, and made using group processes.

Aligning stakes with purpose
The perception of high stakes to learners can inhibit the use of assessment for learning.48 Thus, every effort should be made to lower the perceived stakes of entrustment decisions that are designed to serve primarily formative purposes. We should not assume that learners' perception of stakes can be lowered only through programmatic structure (e.g., many data points) or formative feedback. Learner agency should be promoted through engagement in the assessment process, allowing for it to be a bidirectional conversation.21 Co-production is a concept that is gaining traction in health care and may serve as a useful model for this purpose.95–97 Holmboe defines co-production in assessment as “the interdependent work of learners, faculty, health professionals, and patients to design, create, develop, deliver, assess, and improve the relationships and activities that contribute to the effective assessment and professional development of learners.”98 Allowing for co-production at both the programmatic level (assessment systems) and the individual level (learner-initiated assessment) may help lower the perceived stakes for the learner and promote assessment for learning. Figure 2 shows how the stakes of an entrustment-based in-training evaluation report could be altered through learner engagement without changing the purpose.

Stakes for the learner can also be lowered by reducing the supervisor–trainee hierarchy and creating a supportive learning environment.44,99 Teachers may find it easier to flatten hierarchies and align their approaches with the low-stakes assessment principles of programmatic assessment when a third party, such as a promotion or competence committee, is given the responsibility of making higher-stakes entrustment decisions.100 Supportive learning environments can be promoted through longitudinal coaching relationships and by fostering community between educators and learners.

It may also be helpful to explicitly address the concept of entrustment with learners.93 Being transparent that entrustment is not intended to make a judgment on the learner's moral character may help reduce the stakes inherent to entrustment frameworks.87 To reinforce this perspective, terms such as “entrustment-supervision scales” or “readiness” for differing levels of supervision should be used rather than “trustworthy,” “entrusted,” or “trustworthy” when making summative decisions.66

Aligning process with stakes
Entrustment decisions with lower risk and lesser consequences can be made using ad hoc or structured processes. However, entrustment decisions with more significant consequences require more structured approaches. For example, supervisors in higher-risk (and hence higher-stakes) clinical contexts (e.g., supervising a complex surgical procedure) should use deliberate, explicit entrustment processes to determine appropriate levels of supervision and clearly communicate these decisions with the learner to ensure understanding and mitigate risk of patient harm. Decisions with higher educational stakes for the learner, such as entrustment that will affect promotion or graduation, should also be made using highly structured processes (e.g., competence committee) informed by myriad programmatic assessment data (Figure 3a).1420 Figure 3b shows a misalignment between stakes and structure, such as when competence committee decisions are made using unstructured group processes based on anecdotal data.

Case Study Revisited
Over the next year, Maja’s residency develops an integrated program of assessment using entrustment decisions for both formative and summative purposes. The dual purpose of end-of-rotation entrustment-supervision ratings is made clear to all residents, as well as the fact that all high-consequence decisions

S60
will occur using highly structured processes in the program’s competence committee. Maja is given agency to seek out multiple assessment opportunities and to show improvement over time with assessors. She is aware that the purpose of these entrustment-supervision ratings is mostly formative and perceives them as lower stakes both because of her agency in the process and the large number of assessment opportunities that she has over time. A longitudinal coach uses these assessments to help her set long-term improvement goals, while a competence committee representative provides her with transparent feedback on more summative entrustment decisions. In this context, Maja and her assessors approach their end-of-rotation feedback sessions with a shared understanding of the purpose, stakes, and process of entrustment decisions, allowing Maja to focus on her growth and improvement.

Conclusion

Entrustment decisions are complex and multifaceted. Some characteristics of entrustment decisions may be mismatched to their intended use, contributing to suboptimal assessment. Education leaders can mitigate these pitfalls by: (1) identifying the purposes, stakes, and processes involved in various types of entrustment decisions and (2) aligning them within a program in myriad ways that support assessment for learning.

Funding/Support: This article appears in a supplement that is funded by the American Board of Pediatrics Foundation.

Other disclosures: None reported.

Ethical approval: Reported as not applicable.

References

23 ten Cate O. Trust, competence, and the supervisor's role in postgraduate training. BMJ. 2006;333:748–751.
Focusing on the Learner in EPA-Based Assessment

83 ten Cate O. Entrustment as assessment: Recognizing the ability, the right, and the duty to act. J Grad Med Educ. 2016;8:261–262.
87 Gingerich A. What if the ‘trust’ in entrustable were a social judgement? Med Educ. 2015;49:750–752.
93 Batalden P. Getting more health from healthcare: Quality improvement must acknowledge patient coproduction—An essay by Paul Batalden. BMJ. 2018;362:k3617.