# Assessment in Simulation versus Clinical Context

## A Different Lens for Different Moments

**Iva Bursac, M.D.[1,2] and Briseida Mema, M.D., M.H.P.E.[3,4]**

[1]Division of Pediatric Critical Care, Jim Pattison Children's Hospital, Saskatoon, Saskatchewan, Canada; [2]Department of Pediatrics, College of Medicine, University of Saskatchewan, Saskatoon, Saskatchewan, Canada; [3]Department of Critical Care Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; and [4]Department of Pediatrics, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

ORCID IDs: 0000-0002-3273-3124 (I.B.); 0000-0001-9622-0304 (B.M.)

I (B.M.) frequently take photographs to record important moments. In my spare time, I transform some of the more significant photographs into paintings, one of which was left on the floor of my office as a trainee entered hesitantly. She wanted to discuss an unfair assessment she had received in simulation, and as we started discussing, we were naturally drawn to look at my abandoned painting. What started as a distraction soon became a relevant topic of our conversation about the analogy of photography and painting to assessments. The photographer Dorothea Lange once said, "Photography takes an instant out of time, altering life by holding it still" (1). Taking a photograph might be akin to performing an assessment. We capture moments and freeze them so that we can revisit the past in the future. Sometimes a good photographer with the right camera can draw out certain aspects and nuances of a subject so they are evident to other viewers. A painting after a photograph is meant to reproduce or simulate the original picture. Its quality depends on many factors, including the *complexity* of what's being painted, the painter's intended *focus* of visual attention, and their degree of *skill*.

These reflections and analogies were the stimuli for this piece on the important topic of assessment in the era of competency-based medical education (CBME). Critical care training programs in North America have now largely transitioned to a CBME model. Thousands of publications have been dedicated to emerging challenges in trainee assessment, with learners and teachers discussing and

(Received in original form April 23, 2022; accepted in final form October 12, 2022)



**Author Contributions:** I.B. and B.M. contributed significantly to this work.

**Correspondence and requests for reprints should be addressed to** Iva Bursac, M.D., Division of Pediatric Critical Care, Jim Pattison Children's Hospital, 103 Hospital Drive, Saskatoon, SK, S7N 0W8 Canada. E-mail: iva.bursac@usask.ca.

ATS Scholar Vol 4, Iss 1, pp 12–19, 2023

DOI: 10.34197/ats-scholar.2022-0040CM

debating both the merits and shortcomings of CBME extensively (2, 3). CBME's implementation in the workplace has mandated multiple assessments of learners, and sometimes in acute care we have relied on simulation for clinical events that occur rarely or for convenience of deciding what and when to assess (4).

In a recent review examining the evidence behind many of the important assumptions that the medical community has formed about CBME (3), the authors found that the majority of studies (10 of 12) supported that competence is *not* adaptable to any context or task. This finding is important when we use competence in simulation as a surrogate for competence in real life, and vice versa, because these two contexts are different. Recent evidence supports the notion that validity in CBME assessments exists as threatened, uncertain, and assumed (5), and looking more specifically at simulation for assessment of learning, very few studies measure an important part of validity evidence: the correlation of performance in simulation with real life (6). Some groups offer frameworks to think about assessment in simulation that follows the development of the learner, but their validity has not yet been tested (7). In the current landscape, there is a huge need for assessment in simulation and sparse literature to support the validity of simulation assessments as a surrogate for real life. The community cannot stay paralyzed or abandon assessments in the simulation context; rather it should look carefully at its major benefits and potential drawbacks.

The medical community's focus on assessment is portrayed by the frequently updated frameworks about what constitutes a good assessment (8). Although the validity of assessment tools will always remain a key component, other important considerations are feasibility, acceptability, and educational and catalytic effect. Feasibility concerns the practicality of assessment; acceptability concerns its credibility; catalytic and educational effect concerns its ability to support further learning and improve education programs. In this piece, we examine how these elements that form a good assessment might be affected by two important issues when assessment in simulation is used as a surrogate for real-life assessments: the differences in learners' behaviors and performance in simulated environments and the validity of assessments that use simulation.

## DIFFERENCES IN LEARNERS' BEHAVIORS AND PERFORMANCE IN SIMULATION

The differences in learners' behaviors in simulation compared with real life may be affected by *1*) the realism or fidelity of the simulation; *2*) the fiction contract, which has replaced the suspension of disbelief whereby both the assessor and the learner enter a contract to make the simulation worthwhile; and *3*) the preplanning of simulation activities (9). Although these aspects of simulation are often discussed and researched in simulation-based learning, the implications when using simulation as an assessment tool are different and haven't received as much attention.

Realism or fidelity is described as multidimensional. Although many frameworks exist, they all include three dimensions of fidelity that are of particular importance: how simulations look (structural fidelity), what a simulator does (functional fidelity), and the social context in which simulation takes place. Classically, we have focused on structural fidelity, or the degree to which a simulator

or scenario superficially mimics a real patient or a real clinical setting as the be-all and end-all of fidelity. However, we have learned that although important, the relationship of structural fidelity and learning outcomes is not linear but more complex depending on other dimensions of fidelity (9, 10). Another perhaps more important dimension of fidelity is functional fidelity, which is what the simulator does in response to a learner's actions combined with what the simulation is supposed to do so that learning objectives for the scenario are satisfied, or the learner can exhibit the skills we want to assess (10). If the fidelity perceived by the learner differs from that by the assessor (the assessor has an idea of the design of the scenario and what's expected), it can be reflected negatively on the assessment despite the fact that the learner's performance might have been appropriate in another context. This needs to be taken into account when using simulation to perform an assessment, as the assessment needs to reflect the trainee's performance as it would in an actual clinical scenario, and the effects of structural fidelity should be minimized on the learner.

A further element that affects fidelity is the social context in which the learning or assessment takes place. Engagement is required from both the learner and the assessor to form a fictional contract, and the degree of engagement will affect the quality of the simulation and therefore of the assessment (9). To engage learners, there must be a balance between structural fidelity, functional fidelity, and how engaged learners are emotionally. Simulation, when used for assessment, is a social learning activity demanding both the learner and the assessor. Rather than a play in which only the actor is in action, the assessor must also participate in this complex scenario and share their representation of the simulation for the learner to understand the intentions of the simulation, as well as try to understand the intentions of the learner during the assessment. This bidirectional dynamic is described as intentionality and is required for the educator to be able to have a fair assessment of the trainee (10).

To continue with our analogy of painting and assessments, painting is a process that starts with a blank canvas and progresses to represent and interpret scenes, feelings, and emotions for the viewer. Developing simulation scenarios is also a painting process, in which the close representation of reality (structural fidelity) might be too complex for the painter (simulation developer and assessor) to reproduce, or the person looking at the painting to appreciate, because of the cognitive load (on the learner who is being assessed), or not even needed. Just as an impressionist painting that captures reality differently can elicit powerful feelings and emotions, the simulation developer and assessor should analyze what is needed in the simulation scenario to elicit the intended responses for the learner (functional fidelity). Last, for the viewer (learner) to appreciate the intended focus (goals of assessment) of the painter (assessor), sometimes a conversation is needed (intentionality).

Unlike in real life, part of the learner's performance will be affected by the fact that the scenario is preplanned, and rather than engaging fully, they will be attempting to figure out what the agenda for this scenario is. The following paragraph from Fenwick and Dahlgren highlights this thoughtfully: "Participation in all educative simulation is an imaginative act, regardless of the degree of fidelity. The experience is always and

already located in a pedagogical frame with particular objectives and performative expectations. The scenarios are often pre-planned narratives: the only permitted uncertainty may be the students' own performance: they are positioned as plugged into an existing situation rather than co-emergent with it. Students are observed, are often assessed and are most likely aware of the backstage orchestrations at work to maintain the illusion" (11).

Although all the above points are very important when simulation is used for learning and assessment, the body of literature is focused on how these factors affect learning and less on how they affect assessment. One can see how fidelity can affect feasibility when complex situations must be replicated in simulation. Preplanning of the scenarios and learners' performance can affect the educational and catalytic effects by unintendedly focusing the learning on figuring out the "backstage orchestration." It is up to the educator to keep an open mind, be aware of factors that influence performance, and be ready to discuss the performance and assessment with the learner.

## VALIDITY OF SIMULATION ASSESSMENT

Validity has been described as "the single most important attribute of all assessment data" (12). Indeed, without validity we cannot know whether the picture painted by our assessment is truthful to our subject. Multiple validity frameworks are described in the literature, one of the most known being Messick's five sources of evidence validity (13).

Despite its obvious importance in any type of assessment and especially in high-stakes assessments or in medical education models that rely on serial assessments such as CBME, there is a gap in valid assessment as it pertains to simulation. A systematic review of simulation-based assessments from 2014 revealed that at the time, only 3% of studies referenced Messick's validity framework, and nearly a quarter of studies did not reference any validity framework at all (14). That review is outdated, and the explosion of the use of simulation both for learning and assessment has been naturally followed by an influx of studies describing simulation-based assessment in conducting validity research. Although the attention and effort dedicated to research on validity are to be applauded, it is imperative for instructors using simulation-based assessment tools to be informed users. Often published assessment tools have incomplete evidence (14). Understanding validity frameworks will help educators understand this vast literature and select the best assessment tools for their goals. This is of critical importance, as without valid assessment tools we cannot with confidence trust the results. As Cook and Hatala explained, "Validation does not give a simple yes/no answer regarding trustworthiness (validity); rather, a judgment of trustworthiness or validity depends on the intended application and context and is typically a matter of degree" (15). When reviewing simulation-based assessment tools, the educator needs to consider the intention of the assessment in context. They can review if simulation is needed for this assessment or if this can be done in an authentic workplace setting. They also need to take into consideration who the learner is and whether the cognitive load of this assessment is appropriate to their level. When scenarios are standardized to increase reliability, they might reduce authenticity. Validity may suffer because the learner is trying to figure out the simulation agenda rather than fully engaging with the scenario. These are some issues

**Table 1.** Summary of Messick's validity framework and potential issues with simulation-based assessment

| What Aspects Does the Evidence Address? | Evidence Source | Description | How a Simulation-based Scenario May Affect the Evidence |
|---|---|---|---|
| Are we measuring what we believe we are measuring? | Content | The content of the simulation scenario reflects what assessment is intended to measure | Design of the scenario<br>Fidelity (structural and functional) adapted to what is intended to be assessed<br>Focus on structural fidelity may increase cognitive load for a novice learner and distract from the purpose of assessment |
| | Response process | Quality control: raters' actions | Intentionality: understanding intention of learner and assessor in simulation<br>Preplanning of scenarios |
| | Internal structural reliability | Reproducibility of scores | Strong evidence in standardized scenarios |
| | Relationship with other variables | How the performance on this assessment correlates with other assessments (e.g. workplace-based assessments) or expertise | Because context is different, there is paucity of data correlating performance in simulation and real life<br>Standardization, while increasing reliability, decreases authenticity as the scenario differs from workplace setting |
| Impact of assessment | Consequences | Benefits and harms of assessment | Someone might have a great performance in simulation and less so in real life, or someone might perform very well in real life and is not familiar with simulation |

among many that can affect validity. We describe the different elements of Messick's framework and how simulation-based scenarios may affect the evidence in Table 1.

Furthermore, another emerging topic in health professions education is using assessment beyond measuring performance or competence but rather as a process that affects learning itself, or assessment for learning (12). Although it is the assessor who decides what is being assessed and which tool has been validated for this use, the consequences will affect the learner, and thus the learner should be an active participant in judging the impacts of the assessment. This resonates even more so in the era of CBME, when one of the advantages often put forward is that trainees are given more control of their learning. Despite this, the assessment literature remains centered on validity data for assessments of learning.

To conclude, assessments consist of judgments and interpretations. Although simulation is an indispensable tool for training and assessing, we should nevertheless be informed users of this powerful tool. In this short commentary, we have addressed some of the issues educators should keep in mind when using simulation for assessment, and these are summarized in Table 2. From an educator's perspective, a single assessment is just one in the sea of hundreds received by a trainee throughout their program, but from a trainee's perspective, each assessment weighs heavily, especially if they do not believe it represents their performance. Going back to that conversation in my office with the learner, we discussed the points raised here using the analogy of painting and photos. The conversation was an extension of the debrief that happens after the simulation but this time focused on aspects of the assessment. The learner's concerns of having successfully managed this scenario in real life a couple of times, and the issues with fidelity and understanding the intention of assessment, were carefully explored together with the

**Table 2.** Summary of challenges of using simulation as assessment tool in competency-based medical education

| | |
|---|---|
| Variables that can affect learners' performance in simulation | • Fidelity or realism: reframing focus on functional fidelity rather than physical fidelity<br>• Social learning, shared representation, and intentionality: active participation by both trainee and assessor<br>• Fiction contract: degree of engagement by both assessor and learner will affect the quality of the simulation and of the assessment |
| Validity of assessments in simulation | • Paucity of validity data in assessment in simulation literature<br>• Familiarity of educators with validity frameworks and applying simulation-based assessment tools to their context<br>• Validity data based on expert opinion and for assessments of learning outweigh evidence on assessments for learning and as seen from the learner's perspective |

benefits of the feedback received from this experience in simulation.

As CBME with its outcome-based approach is the current reality, simulation-based assessment will most certainly augment workplace assessments. Understanding its benefit and drawbacks is key. The astronomer and educator Maria Mitchell was the first woman elected a fellow of the American Academy of Arts and Sciences, and as a scientist, she was quite comfortable with numbers and adhered to the most rigorous standards of scientific research. Yet she refused to grade her students (16). We know that it is inherently impossible to contain everything about a trainee in one assessment. Assessments are tools that we use to measure something as complex as competence. Bit by bit, each assessment reconfigures our understanding of a trainee's competence. This competence is revealed to us only in fragments with each "photograph" we take. The more "fragments" we capture and carefully organize, the more lifelike the mosaic of competence we make of them. But it is still a representation, imperfect and incomplete, subject to a continuous transformation.

Author disclosures are available with the text of this article at www.atsjournals.org.

## REFERENCES

1. Partridge E. Dorothea Lange: grab a hunk of lightning. San Francisco, CA: Chronicle Books; 2013.
2. Roberts LW. Emerging issues in assessment in medical education: a collection. *Acad Med* 2021;96: 159–160.
3. Brydges R, Boyd VA, Tavares W, Ginsburg S, Kuper A, Anderson M, *et al.* Assumptions about competency-based medical education and the state of the underlying evidence: a critical narrative review. *Acad Med* 2021;96:296–306.
4. Pandya A, Patocka C, Huffman J. Simulation for assessment of entrustable professional activities in an emergency medicine residency program. *Can J Emerg Med* 2022;24:84–87.
5. Tavares W, Rowland P, Dagnone D, McEwen LA, Billett S, Sibbald M. Translating outcome frameworks to assessment programmes: implications for validity. *Med Educ* 2020;54:932–942.
6. Scalese RJ, Obeso VT, Issenberg SB. Simulation technology for skills training and competency assessment in medical education. *J Gen Intern Med* 2008;23:46–49.
7. Roussin C, Sawyer T, Weinstock P. Assessing competency using simulation: the SimZones approach. *BMJ Simul Technol Enhanc Learn* 2020;6:262–267.
8. Norcini J, Anderson MB, Bollela V, Burch V, Costa MJ, Duvivier R, *et al.* 2018 consensus framework for good assessment. *Med Teach* 2018;40:1102–1109.
9. Rudolph JW, Simon R, Raemer DB. Which reality matters? Questions on the path to high engagement in healthcare simulation. *Simul Healthc* 2007;2:161–163.
10. Schoenherr JR, Hamstra SJ. Beyond fidelity: deconstructing the seductive simplicity of fidelity in simulator-based education in the health care professions. *Simul Healthc* 2017;12:117–123.
11. Fenwick T, Dahlgren MA. Towards socio-material approaches in simulation-based education: lessons from complexity theory. *Med Educ* 2015;49:359–367.
12. Yudkowsky R, Park YS, Downing SM. Assessment in health professions education, 2nd ed. New York: Routledge; 2020.

13. Messick S. Validity. In: Linn RL, editor. Educational measurement, 3rd ed. New York: American Council on Education and Macmillan; 1989. pp. 13–103.

14. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. *Adv Health Sci Educ* 2014;19:233–250.

15. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. *Adv Simul (Lond)* 2016;1:31.

16. Howe JW. Reminiscences 1818–1899. Boston: Houghton, Mifflin; 1900.