Exploring Validity Evidence for the use of Immersive Virtual Environments for Formative and Summative Assessment Purposes

Advances in research, technology, and faculty expertise have contributed to impressive educational gains using immersive virtual environments (IVEs)–such as virtual, augmented or mixed reality–but the validity evidence for their use as an assessment modality is less certain.

Threats to validity and formative value of IVE for assessment may emerge in two important ways. First, the assumption that the observed trainee performance will be shaped by underlying attributes and not something else. Second, that observers (i.e., raters, faculty) can generate meaningful assessment data informed by trainee behaviors, and again, not by something else.

In both cases, the “something else” is referred to as construct irrelevant variance (CIV), reflecting any factor other than the construct that has an impact on the indicator. For example, relative to physical simulations, or actual clinical environments, do IVEs prompt trainees to shift the way they make decisions, communicate, or conduct themselves? Do observers shift their observations and interpretations when judging IVE performance, relative to physical performances in the simulation- or workplace-based settings?

Hence, the interest in adopting modern computing for assessment of clinical competence may outpace the evidence supporting its use. Efforts to date have been on designs without fully appreciating validity implications. To address such risks and gaps, we will explore the validity argument and formative value of IVE while also demonstrating a process others might use in their own contexts.