NBME

STEMMLER GRANTS

2021 Stemmler Grants Projects

Established in 1995, the Stemmler Grant Program supports the research and development of innovative assessment approaches with the potential to advance assessment in medical education. Each year, as many as three awards of up to $150,000 each are given to research teams to support their efforts to drive innovations in the field.

Learn more about the 2021 projects.

Automated Assessment of Written Chart Notes: Generating Reliable, Timely, and Useful Feedback

William Bond, MD, MPH

Director of Research, Jump Simulation at the University of Illinois of Chicago School of Medicine

Suma Bhat, PhD

Assistant Professor at the University of Illinois at Urbana-Champaign

Project Summary

This project aims to develop and implement and evaluate new ways to extend the capabilities of ASAG systems using advances in natural language processing and machine learning to rapidly and accurately assess learners' responses in the patient note.

Abstract

Scoring patient notes (PNs) after standardized patient (SP) encounters is a time-consuming process, requiring significant faculty effort and affording only limited and delayed learner feedback. Improving faculty efficiency and creating a learner feedback mechanism presents the perfect opportunity to automate the scoring process using natural language processing (NLP). Our team with expertise in natural language processing (NLP), simulation, and assessment will develop, implement and evaluate new ways to extend the capabilities of ASAG systems using advances in NLP and machine learning (ML) to rapidly and accurately assess learners' responses in the PN. In addition to its assessment value, phrase-level ASAG represents a rich "deliberate practice with feedback" opportunity, which is

lost in the current grading method. The ASAG system will provide feedback to faculty and learners at the case section level (e.g., history score, differential diagnosis score) and case content domain area (e.g., cardiology, dermatology) for summative assessment of learning or phrase-level feedback at the item level (e.g., asked about chest pain quality, noted peritoneal signs) during assessment for learning. The system will use a human-in-the-loop (HITL) approach by seeking human input when the ASAG is below a threshold confidence level. This creates transparency of the system and uses human judgment to improve performance, allowing for faculty checks of the ASAG, thus leading to a learning ASAG system. By rapidly determining those who have clearly passed a case, the system will allow faculty to focus human oversight efforts on learners at the margin of passing.

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

Walter Tavares, PhD

Assessment Scientist and Assistant Professor at the University of Toronto

Fahad Alam, MD

Medical and Research Director at the Sunnybrook Canadian Simulation Centre

Project Summary

This project will explore the validity argument and formative value of immersive virtual environment while also demonstrating a process others might use in their own contexts.

Abstract

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.

Bias Reduction in Curricular Content: Using Machine Learning and Artificial Intelligence to Assess Bias in Medical Education

Project Summary

This project will explore the use of machine learning and artificial intelligence as ways to reduce bias in medical education.

Robert Montenegro, MD, PhD

Assistant Professor, Child and Adolescent Psychiatry; Associate Director, Bias Reduction in Curricular Content (BRICC) at Seattle Children's Hospital Foundation

Abstract

The field of medicine is marred by a long, painful, and deleterious history of overt and covert forms of social injustice, bias, and racism, as illustrated by the American Medical Association's recent pledge to take action to confront systemic racism. Identifying and reducing bias in medical curricula and assessment content is critical and fundamental to the education of future physicians. Studies continue to demonstrate that physicians possess implicit biases in a number of different areas such as race/ethnicity, gender, sex, age, weight, substance use and mental illness. The impact of the numerous and persistent biased faculty shortcomings will inevitably be reflected in the faculty's curricular and assessment content, and in turn, may affect the care that medical students ultimately provide for their future patients. A biased curriculum can also negatively impact the learning environment and well-being of medical students, especially students from underrepresented backgrounds. Despite numerous calls to action to deracialize and debias medical curricula and assessment content, most medical institutions continue to teach biased medicine in preclinical years. Many educators, for example, continue to inappropriately use race as a proxy for genetics or ancestry, or even as a "risk factor" for numerous health outcomes often erroneously associated with race (e.g. GFR race coefficient, Sickle Cell Disease, Salt Gene Hypothesis, HTN, or Schizophrenia) while ignoring social or structural determinants of health (SSDoH), such as systemic racism or income inequities. Many educators also continue to inappropriately use gender and sex terms and perpetuate the idea that sex and gender are binary and stagnant (vs fluid), which can potentially negatively impact gender nonconforming students and patients alike. Likewise, most medical educators are unaware of the numerous biases in the types of images they use in their lectures or assessment materials as well.