

## 2022 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 2022 projects.

### Development and Validation of a Machine Learning Model for Automated Workplace-Based Assessment of Resident Clinical Reasoning Documentation



**Verity Schaye, MD, MPHE**

*Associate Professor,  
Department of Medicine,  
Assistant Dean for Education in  
the Clinical Sciences, Assistant  
Director, Curricular Innovation  
in Medical Innovation; Director,  
Integrated Skills at the NYU  
Grossman School of Medicine*



**Sally Santen, MD, PhD**

*Professor, Evaluation,  
Assessment, Scholarship  
Senior Associate Dean at  
the Virginia Commonwealth  
University*



**Jesse Burk-Rafel, MD, MRes**

*Assistant Professor, Internal  
Medicine at the NYU  
Grossman School of Medicine*

#### Project Summary

This project aims to develop and validate a machine learning model for automated workplace-based assessment of clinical reasoning documentation.

#### Abstract

Diagnostic errors annually contribute to 17% of adverse hospital events and 10% of patient deaths. Premature closure and failure to establish a sufficient differential diagnosis are major sources of diagnostic error. It was anticipated that the advent of the electronic health record (EHR) would help

decrease diagnostic errors by fostering thoughtful assessments in a centralized place. However, since EHR implementation, there has been a decline in documentation quality including documentation of clinical reasoning (CR). A contributing factor to the decline is trainees receive infrequent feedback on notes, partly due to faculty time limitations. In this project, we aim to develop and validate a machine learning (ML) model for automated workplace-based assessment (WBA) of CR documentation. Previously we developed an innovative WBA using ML and natural language processing for formative feedback on medicine residents' CR documentation. We are seeking funding to support two subsequent aims: (1) improving the ML model to provide more specific feedback; and (2) generating additional validity evidence to support use of the WBA in summative assessment. To demonstrate generalizability, we will conduct these aims in parallel at two institutions. Our current ML model has high performance but provides only dichotomous output, rating notes as low- or high-quality CR documentation. To generate more specific feedback, we will label a dataset of notes using a human assessment rubric we developed and validated in prior phases of this work, and train new ML models to provide feedback. We will gather validity evidence for the new ML model using Messick's framework. After sufficient validity evidence is collected, this novel WBA can be integrated into CR competency-based assessment programs to help facilitate achievement of the sub-competency "appropriate utilization and completion of health records." Our multi-site collaboration will also generate a process for implementation to facilitate dissemination of this tool across EHRs.

## Predictive Validity and Workplace-Based Assessments: A Rigorous Framework for Evaluating Operative Performance



**Andrew Krumm, PhD**

*Assistant Professor of  
Learning Health Services  
at the University of  
Michigan Medical School*

### Project Summary

This project will analyze operative performance evaluations from a popular workplace-based assessment, SIMPL OR, to help surgical training programs identify trainees who are on- and off-track in their development.

### Abstract

There is a critical need for developing systematic scoring approaches for assessing the operative skill of surgical trainees over time. This project will analyze operative performance evaluations from a popular workplace-based assessment, SIMPL OR, to help surgical training programs identify trainees who are on- or offtrack in their development. It will (1) examine the predictive validity of three distinct models applied to SIMPL OR evaluations and (2) examine the reliability and utility of scores derived from predictive models applied to SIMPL OR evaluations.