

Evaluation of an Automated Procedure for Scoring Patient Notes as Part of a Clinical Skills Examination

Kimberly Swygert

Melissa Margolis, Tim Siftar, Stephen Clyman, Brian Clauser
National Board of Medical Examiners
3750 Market Street, Philadelphia, PA 19104
kswygert@nbme.org

Keywords: clinical skills assessment, automated essay scoring, latent semantic analysis, performance assessment, predicting expert judgements

Abstract

In the movement towards the use of standardized-patient-based (SP-based) tests in large-scale high-stakes medical assessment, the issues of accuracy, reliability, cost efficiency, and scoring efficiency are of great concern. This is especially the case with the scoring of patient encounter notes that are completed after an examinee has interacted with an SP. Considerable efficiency and cost reduction could be achieved if automated procedures, which have been developed over the past few decades, could be used for scoring these notes.

The purpose of the present study was to examine the extent to which an automated scoring procedure could predict expert human ratings of patient notes that are part of the clinical skills examination (CSE) currently planned for implementation in addition to Step 2 of the United States Medical Licensure Examination (USMLE). An automated scoring procedure that utilizes Latent Semantic Analysis (LSA) was applied to the patient notes generated by several hundred examinees across four CSE cases. Results show that the automated scores were able to predict a substantial portion of the variance in expert judgments, but results were inconsistent across case. For three of the four cases, the predicted scores generated from automated component scores predicted more of the variability in the mean human holistic ratings than did a single expert human rater. Implications for further research are discussed.