

Grant R305F100007
Year of Study: 2012

Title: Computer-based assessment of student constructed responses

Authors: Magliano, J. P., & Graesser, A. C.

Citation: Magliano, J. P., & Graesser, A. C. (2012). Computer-based assessment of student constructed responses. *Behavioral Research Methods*, 44, 608-621.

Strand of Work: Tools

Abstract

Purpose and Questions Investigated

It makes sense that tasks that allow students to express complex understanding of texts, such as writing an essay or to thinking aloud about a text, provide opportunities for students to acquire skills necessary to succeed in college and beyond. However educators are faced with the task of grading these exercises, which can be both time consuming and cognitively challenging. Consequently, they may opt for tests that require simple responses, such as true/false or multiple-choice tests, because these are relatively easy to grade. Fortunately, researchers have been developing systems to automatically score student essays, making them a more viable choice for educators. The goal of this paper was to take stock in the variety of approaches that have been developed to evaluate student constructed responses and to make recommendations about the optimal approaches that are available.

Research Context or Methodology

This journal article is a literature review of the various approaches that have been proposed in the literature for scoring student constructed responses. The research reviewed in this article reflects a variety of applications in the computer-based coding of language: (1) scoring protocols in the context of computer-based tutoring systems, (2) developing stand alone assessments of student writing or reading comprehension proficiency, and (3) the automated coding of authored texts (e.g., textbooks). This literature review focuses on two dimensions that need to be considered when developing new systems. The first is type of response provided by the student—namely, meaning-making responses (e.g., think-aloud protocols, tutorial dialogue) and products of comprehension (e.g., essays, open-ended questions). The second corresponds to considerations of the type of natural language processing systems used and how they are applied to analyze the student responses.

General statement of findings

We argue that the appropriateness of computer-based assessment protocols is, in part, constrained by the type of response. For example if one is developing a system for evaluating responses that reflect an understanding of specific aspects of a text, then comparing student response to specific content in that text (e.g., specific sentences in the text) is appropriate. However, if one is developing a system to code essays that are not based on any source, then automated coding must focus on the features of the essay (e.g., how coherently related the sentences in the essay are, frequency of the words used, complexity of the sentences) or comparisons to exemplar essays that reflect different

levels of student proficiency. However, with respect to the issue of optimal natural language processing, the conclusions drawn from the literature review were more definitive. Researchers should use hybrid systems that rely on multiple, convergent natural language algorithms.

Implications

This literature review provides an important step toward consolidating our understanding of the best approaches for analyzing student written responses. It provides researchers with system for determining the nature of the responses that they need to score, the challenges that one faces when scoring these responses, and the best approaches for doing so.

Acknowledgements

The research was supported by the National Science Foundation (ALT-0834847, DRK12-0918409), the Institute of Education Sciences (R305B070349, R305F100007), the Center for the Interdisciplinary Studies of Language and Literacy, and the Institute for Intelligent Systems. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these funding sources.