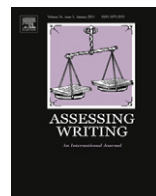




ELSEVIER

Contents lists available at SciVerse ScienceDirect

Assessing Writing



English language learners and automated scoring of essays: Critical considerations

Sara Cushing Weigle*

Georgia State University, P.O. Box 4099, Atlanta, GA 30302-4099, United States

ARTICLE INFO

Article history:

Available online 6 December 2012

Keywords:

Automated essay scoring
Second language writing
English language learners
Writing assessment

ABSTRACT

This article presents considerations for using automated scoring systems to evaluate second language writing. A distinction is made between English language learners in English-medium educational systems and those studying English in their own countries for a variety of purposes, and between learning-to-write and writing-to-learn in a second language (Manchón, 2011a), extending Manchón's framework from instruction to assessment and drawing implications for construct definition. Next, an approach to validity based on articulating an interpretive argument is presented and discussed with reference to a recent study of the use of e-rater on the TOEFL. Challenges and opportunities for the use of automated scoring system are presented.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

In the US and internationally, automated scoring is increasingly being used or considered for use in writing assessment. Automated scoring is, of course, faster and potentially less expensive than human scoring, and vendors of automated scoring systems frequently point out that their product is at least as reliable as human scorers (e.g., Attali & Burstein, 2006). The use of automated scoring has generated controversy, particularly among scholars and practitioners in university-level composition programs (e.g., Cheville, 2004; Condon, 2013; Herrington & Moran, 2001), including some whose work focuses on second language writers (e.g., Crusan, 2010), but less so among second language assessment specialists. As an illustration of these contrasting views, a position statement adopted by the Conference on College

* Tel.: +1 404 413 5192.

E-mail address: sweigle@gsu.edu

Composition and Communication (CCCC, 2004), states unequivocally: “We oppose the use of machine-scored writing in the assessment of writing” (“A Current Challenge” section, para. 2). In contrast, the editorial for a recent special issue of *Language Testing* on automated scoring and feedback systems (Xi, 2010) takes a more moderate view, concluding that “Computer capabilities, if used appropriately and responsibly, can expand the resources and improve the efficiency of language learning and assessment. However, the current limitations of [Natural Language Processing] and speech technologies also call for responsible and cautious use of them and call into question the appropriateness of using them alone in scoring assessments for high-stakes decisions” (pp. 297–298).

The contrast between the perspectives on automated scoring between the disciplines of composition (including second language composition) and language assessment (including the assessment of second language writing) stem in large measure from their focus on different learner populations, leading to different conceptualizations of the construct of writing and different goals for writing instruction and assessment. In this article I lay out some of these differences and discuss their implications for automated scoring of writing for second language learners of English.

The article is organized as follows. First, I make a broad distinction between English language learners in English-medium educational systems such as the US, and those studying English in their own countries for a variety of purposes, and how this distinction plays out in formulating a construct of second language writing for instruction and assessment. Following Manchón (2011a), I draw a distinction between learning-to-write and writing-to-learn in a second language, extending Manchón’s framework from instruction to assessment and drawing implications for construct definition. Next, recent scholarship in language testing has moved from the validity framework of Messick (1998) and others toward an approach to validity based on articulating an interpretive argument (Bachman, 2005; Chapelle, Enright, & Jamieson, 2008; Kane, 1992, 2002). Using this approach, I discuss the most important inferences that need to be made in moving from a written performance to a score and the use of that score and the evidence that is needed to support these inferences, illustrating this approach with examples from my own research on automated scoring. Finally, I suggest some ways in which automated scoring systems could be improved to better serve these learners and those who teach them.

2. English language learners and writing assessment

English language learners (ELLs) make up an increasing number of students in the U.S. educational system, both in K-12 and higher education. ELLs comprised approximately 10.7% of public school students in 2007–8 (Migration Policy Institute, 2010); in higher education, there were close to 725,000 international students enrolled in colleges and universities (IIE, 2011), not including US residents who speak a language other than English at home. These students are often included in large-scale writing assessments, where they are evaluated alongside their native-speaking peers, using the same criteria. As such assessments become more prevalent even as education budgets shrink, automated scoring of writing becomes more attractive as a way to contain costs and reduce the time and resources needed to evaluate student writing.

Both the National Council of Teachers of English (NCTE) and the Conference on College Composition and Communication (CCCC) have issued position statements urging recognition and accommodation of the diverse cultural, linguistic, and socioeconomic backgrounds and learning needs of ELLs in English/Language Arts/Composition courses. Any discussion of automated scoring of writing in the US context, whether at the K-12 or postsecondary level, must therefore take into account the characteristics of L2 writers so that assessments are fair and unbiased.

While most of the literature on automated scoring has focused on automated scoring in English-based educational systems, it is important to recognize that the largest potential market for automated scoring of English writing may not be for students in the US or English-speaking countries at all, but may be for non-native English speakers (NNES) learning English in their own countries. Currently the major use of automated scoring for NNES is on the Test of English as a Foreign Language® (TOEFL®) iBT™ (internet-based test), required by many US universities as evidence of English language proficiency for international students. The TOEFL iBT uses the e-rater scoring engine, developed by Educational Testing Service, along with human raters to score two different writing tasks.

Download English Version:

<https://daneshyari.com/en/article/344280>

Download Persian Version:

<https://daneshyari.com/article/344280>

[Daneshyari.com](https://daneshyari.com)