



Full length article

Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation

Rod D. Roscoe^{a,*}, Joshua Wilson^b, Adam C. Johnson^a, Christopher R. Mayra^a^a 7271 E. Sonoran Arroyo Mall, Santa Catalina Hall, Suite 150, Arizona State University-Polytechnic, Mesa, AZ 85212, United States^b 213E Willard Hall, School of Education, University of Delaware, Newark, DE 19716, United States

ARTICLE INFO

Article history:

Received 1 June 2016

Received in revised form

12 December 2016

Accepted 28 December 2016

Available online 2 January 2017

Keywords:

Automated writing evaluation

Formative feedback

Technology adoption

User experience

Human-computer interaction

Writing

ABSTRACT

Automated writing evaluation (AWE) is a popular form of educational technology designed to supplement writing instruction and feedback, yet research on the effectiveness of AWE has observed mixed findings. The current study considered how students' perceptions of automated essay scoring and feedback influenced their writing performance, revising behaviors, and future intentions toward the technology. The manner in which the software was presented—claims about the accuracy and quality of the automated scoring and feedback—were modestly related to students' expectations and perceptions. However, students' direct experiences with the software were most strongly associated with their perceptions. Importantly, students' perceptions seemed to have minimal impact on their "in the moment" use of the software to write and revise successfully. Students revised and improved their essays regardless of their positive or negative views of the system. However, positive and negative perceptions significantly predicted future intentions to use the software again or to recommend the software to a friend. Implications for AWE design, implementation, and evaluation are discussed.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Automated writing evaluation (AWE) is a popular educational technology that saves teachers time in assessing writing, enables more writing practice, and supplements writing instruction. Commercially available systems have been deployed across thousands of classrooms, such as Educational Testing Service's *Criterion* (Burstein, Tetreault, & Madnani, 2013), Pearson's *WriteToLearn* (Foltz, Streeter, Lochbaum, & Landauer, 2013), and Measurement Incorporated's *Project Essay Grade* (PEG, Wilson, Olinghouse, & Andrada, 2014). Other systems serve as test beds for research on writing and AWE design, such as *Writing Pal* (Allen, Crossley, Snow, & McNamara, 2014; Roscoe & McNamara, 2013).

Each technology employs different algorithms but the underlying premises are similar (Dikli, 2006; Shermis & Burstein, 2013). Natural language processing (NLP) tools extract linguistic, structural, semantic, and rhetorical text features related to text quality, and these relationships can be statistically modeled to assign holistic writing scores and assess writing traits. Many systems exhibit

high scoring accuracy (Bridgeman, Trapani, & Attali, 2012; Shermis, 2014) and increasingly offer constructive, formative feedback on qualities such as usage, mechanics, organization, and development (e.g., Burstein et al., 2013). These scoring and feedback functions, along with the ability to process thousands of essays in seconds, can enable teachers to offer more writing assignments without a concomitant increase in workload.

Research on the effectiveness of AWE—the extent to which students improve in writing proficiency after using the software—has produced mixed findings (Stevenson & Phakiti, 2013). In one experimental evaluation of *Criterion*, Kellogg, Whiteford, and Quinlan (2010) asked freshman composition students to write and revise three essays with varying levels of feedback (i.e., no feedback versus feedback on one or more essay). Although students made fewer mechanical errors, overall writing quality was not affected by the amount of feedback. Other evaluations have examined patterns of revising and proficiency in larger datasets. Wilson et al. (2014) examined the performance of over 4000 students who used PEG to write and revise essays with feedback. Less than a quarter of students ($n = 955$) submitted more than one revision of their first drafts. Among those who did revise, students achieved small score increases with each draft, but the rate of growth decreased over time and reached a plateau around the 11th

* Corresponding author.

E-mail addresses: rod.roscoe@asu.edu (R.D. Roscoe), joshwils@udel.edu (J. Wilson), acjohn17@asu.edu (A.C. Johnson), cmayra@asu.edu (C.R. Mayra).

or 12th revision. In their review, [Stevenson and Phakiti \(2013\)](#) concluded that “there is only modest evidence that AWE feedback has a positive effect on the quality of texts that students produce using AWE, and that as yet there is little clarity about whether AWE is associated with more general improvements in writing proficiency” (p. 62).

One cause of these mixed efficacy findings may be similarly mixed beliefs about the appropriateness of automated scoring and feedback. Concerns about validity have been at the heart of long-standing debates about AWE ([Anson et al., 2013](#); [Condon, 2013](#); [Deane, 2013](#); [Hearst, 2000](#)), and one overarching critique is that automated approaches do not capture the complete writing construct. Computers can only respond to features of writing that can be automatically detected, which might exclude nuanced and subjective dimensions that even humans find difficult to assess ([Deane, 2013](#)). For instance, the National Council of Teachers of English released a position statement in 2013 ([Anson et al., 2013](#)) arguing that “computers are unable to recognize or judge those elements that we most associate with good writing (logic, clarity, accuracy, ideas relevant to a specific topic, innovative style, effective appeals to audience, different forms of organization, types of persuasion, quality of evidence, humor or irony, and effective uses of repetition, to name just a few).”

Negative perceptions of AWE may have consequences with respect to adoption, use, and effectiveness. Users make decisions about technology adoption based on ease of use and utility, and software that is difficult to access, incomprehensible, or seems to provide few benefits may be rejected ([Vinkatesh & Davis, 2000](#)). Ertmer and colleagues ([Ertmer, 1999](#); [Ertmer, Ottenbreit-Leftwich, Sadik, Sendurur, & Sendurur, 2012](#)) have specifically studied barriers to teachers' use of educational technologies. Logistical barriers, such as a poor student-to-computer ratio or a lack of reliable Internet access, may be substantial when hundreds of students are simultaneously writing and submitting essays to web-based AWE services. Other barriers are grounded in teachers' beliefs, such as the belief that certain domains (e.g., writing) cannot be taught using automated approaches. Instructors who possess necessary resources and expertise ([Koehler & Mishra, 2009](#); [Voogt, Fisser, Pareja Roblin, Tondeur, & van Braak, 2013](#)) may nonetheless reject AWE due to skepticism ([Curran, Draus, Maruschock, & Maier, 2009](#)).

Students also express mixed views about AWE ([Grimes & Warschauer, 2010](#); [Warschauer & Grimes, 2008](#)). In one study ([Grimes & Warschauer, 2010](#)), students rated *My Access* favorably in terms of ease, enjoyment, usefulness, and fairness, and reported that they revised more and increased their confidence after using the system. However, students also focused their attention on low-level writing feedback and were sometimes overwhelmed by the amount of feedback. McNamara and colleagues ([Roscoe & McNamara, 2013](#); [Roscoe, Allen, Weston, Crossley, & McNamara, 2014](#)) conducted similar feasibility assessments of early versions of *Writing Pal* (W-Pal), an intelligent tutoring system for writing instruction. Their study was conducted in several high school English classrooms over one school year. After students had interacted with W-Pal for several months, the researchers probed their perceptions of the feedback system. Most students (about 80%) rated the writing tools as easy to use, but some students critiqued the system with regards to quantity of feedback (i.e., either too much or not enough). About 60% of students found the feedback to be easy to understand, and about 40% reported that the feedback was “often” or “always” useful. In open-ended responses, students noted that a lack of feedback specificity and personalization were key concerns.

Overall, students and teachers seem to cautiously embrace the potential of AWE for providing summative and formative feedback

on writing, but also express doubts regarding scoring accuracy, specificity and personalization, clarity, and quantity of the feedback. When students' preferences for human versus automated feedback have been probed directly, students tended to prefer comments from teachers or peers rather than computers ([Curran, Draus, Maruschock, & Maier, 2013](#); [Lai, 2010](#); [Lipnevich & Smith, 2009](#)).

The current study explores college students' perceptions of automated essay scoring and feedback, and examines the effects of perceptions on writing performance, writing process (i.e., revising), and future intentions toward the technology. Specifically, we assess students' initial expectations about AWE scoring and feedback, immediate reactions to received scores and feedback, and final impressions. Importantly, we also manipulate how system capabilities are presented to student users. As noted above, teachers possess conflicting views about the validity of AWE and may communicate these views to their students ([Li, Link, & Hegelheimer, 2015](#)). With this manipulation, we can consider how messaging from authority figures—in this case, developers and researchers associated with the AWE itself—might influence user perceptions and outcomes. However, the actual functioning or quality of the system is not manipulated; all students interact with the same scoring and feedback tools.

1.1. Research questions

Research Question 1 (RQ1). When AWE scoring and/or feedback capabilities are presented as either “well established” versus a “work in progress,” how does this presentation influence students' expectations about software performance, immediate perceptions of feedback received on their own work, and final impressions of the system?

Research Question 2 (RQ2). Do differences in presentation, expectations, and experience contribute to positive or negative shifts in final perceptions of the system? One possibility is that initial expectations strongly anchor subsequent interpretations of the system. Alternatively, direct experiences and interactions might override original expectations.

Research Question 3 (RQ3). How do positive and negative perceptions of AWE influence writing behaviors and future intentions regarding the system? In terms of writing behaviors, we examine how perceptions of the software and feedback quality relate to revising. Students who believe that automated feedback is more accurate, relevant, or useful may be more inclined to use that feedback and revise extensively. With future intentions, we test whether perceptions influence willingness to use the system again or recommend it to a friend.

2. Method

2.1. Participants

We recruited 110 undergraduate students enrolled in Introduction to Psychology courses at a large university in the southwestern United States. Students received course credit for their participation. Demographically, 35.5% of students self-identified as female with an average age of 22 ($M = 21.8$, $SD = 5.7$). Most students self-identified as Caucasian (40.0%) or Hispanic (20.9%) although other races and ethnicities were represented (see [Table 1](#)). Most students spoke primarily English (69.1%) or were fluent in English and another language (28.2%). A small number of students reported another language as their primary language (2.7%) but possessed sufficient English proficiency to participate. Academically, students reported a relatively high average GPA of about 3.5

Download English Version:

<https://daneshyari.com/en/article/4937292>

Download Persian Version:

<https://daneshyari.com/article/4937292>

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