Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality

Joshua Wilson[*], Amanda Czik

University of Delaware, United States

ARTICLE INFO

ABSTRACT

Automated Essay Evaluation (AEE) systems are being increasingly adopted in the United States to support writing instruction. AEE systems are expected to assist teachers in providing increased higher-level feedback and expediting the feedback process, while supporting gains in students' writing motivation and writing quality. The current study explored these claims using a quasi-experimental study. Four eighth-grade English Language Arts (ELA) classes were assigned to a combined feedback condition in which they received feedback on their writing from their teacher and from an automated essay evaluation (AEE) system called PEG Writing®. Four other eighth-grade ELA classes were assigned to a teacher feedback-only condition, in which they received feedback from their teacher via GoogleDocs. Results indicated that teachers gave the same median amount feedback to students in both condition, but gave proportionately more feedback on higher-level writing skills to students in the combined PEG + Teacher Feedback condition. Teachers also agreed that PEG assisted them in saving one-third to half the time it took to provide feedback when they were the sole source of feedback (i.e., the GoogleDocs condition). At the conclusion of the study, students in the combined feedback condition demonstrated increases in writing persistence, though there were no differences between groups with regard to final-draft writing quality.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

A commonly used method for teaching writing is to provide instructional feedback (Biber, Nekrasova, & Horn, 2011; Black & William, 2009; Hattie & Timperley, 2007; Kellogg & Whiteford, 2009). Instructional feedback is information provided by an agent—such as a teacher, peer, or computer—that indicates both correctness/incorrectness and ways to improve performance or understanding (Hattie & Timperley, 2007; Parr & Timperley, 2010). Struggling writers, in particular, need targeted instructional feedback because they tend to produce shorter, less-developed, and more error-filled and problem-laden texts than their peers (Troia, 2006).

However, the role of instructional feedback in the teaching of writing is not without controversy. Proponents advocate its role in supporting motivation and writing quality by (a) indicating to the author his/her position relative to a desired level of

---

* Corresponding author. University of Delaware, School of Education, 213E Willard Hall Education Building, Newark, DE 19716, United States.
E-mail address: joshwils@udel.edu (J. Wilson).

quality; (b) identifying areas in need of improvement related to low-level writing skills (spelling, word choice, mechanics, grammar) or high-level skills (idea development and elaboration, organization, rhetoric); and (c) prompting additional practice attempts in which the author incorporates and eventually internalizes the feedback (Ferster, Hammond, Alexander, & Lyman, 2012; Kellogg, Whiteford, & Quinlan, 2010; Parr & Timperley, 2010). In contrast, others argue that providing instructional feedback is (a) too time consuming and leads to teacher burnout (Anson, 2000; Baker, 2014; Lee, 2014); (b) too difficult for teachers to provide given the complexity of writing ability (Marshall & Drummond, 2006; Parr & Timperley, 2010); and (c) ineffective or incapable of achieving substantial, generalizable gains in students' writing performance (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Biber et al., 2011; Kluger & DeNisi, 1998). Nevertheless, instructional feedback continues to be recommended as a method for teaching writing (APA, 2015; Graham, Harris, & Santangelo, 2015; Graham, Hebert, & Harris, 2015; Graham, McKeown, Kiuhara, & Harris, 2012; Graham & Perin, 2007).

In the U.S., an increasingly common form of instructional feedback for writing is feedback provided by automated essay evaluation systems (AEE) (Warschauer & Grimes, 2008). AEE are web-based formative writing assessment software programs which provide students with immediate automated feedback in the form of essay ratings and individualized suggestions for improvement when revising (Shermis & Burstein, 2013). Some of the principal benefits of AEE are efficiency and flexibility. While there is no consensus regarding the optimal timing of feedback (see Shute (2008) for review), immediate feedback is often preferred (Chan, Konrad, Gonzalez, Peters, & Ressa, 2014; Ferster et al., 2012) especially in classroom settings (Hattie & Timperley, 2007; Shute, 2008). In addition, unlike teacher or peer feedback, automated feedback allows students to control feedback timing. Students receive feedback when they request it, either in the middle of, or after having completed, an essay draft. This enables feedback to be immediately actionable, accelerating the practice-feedback loop (Foltz, Streeter, Lochbaum, & Landauer, 2013; Kellogg et al., 2010).

While automated feedback addresses some of the barriers faced by teachers when providing instructional feedback, the intended use of AEE systems is to complement and not replace teacher feedback (Foltz, 2014; Foltz et al., 2013; Kellogg et al., 2010). Indeed, AEE is thought to free up instructional time and allow teachers to be more selective in the type of feedback they provide, thereby improving students' writing motivation and writing performance (Grimes & Warschauer, 2010). For instance, after implementing AEE in her high school, a school administrator reported that, "[AEE] has helped motivate our students to write while making it easier for educators to provide the feedback needed to ensure growth in writing" (Schmelzer, 2004, p.34).

Yet, the growing adoption of AEE in the U.S. has been accompanied by a number of concerns and fears. For instance, despite its intended role as a complement to teacher feedback, some fear that AEE will come to replace the teacher as primary feedback agent (Ericcson & Haswell, 2006; Herrington & Moran, 2001), and thereby negate the social communicative function of writing (National Council of Teachers of English [NCTE], 2013). Others are concerned that AEE can be easily fooled to assign high scores to essays which are long, syntactically complex, and replete with complex vocabulary (Bejar, Flor, Futagi, & Ramineni, 2014; Higgins & Heilman, 2014). Concerns such as these have led some groups to summarily reject the use of AEE (Conference on College Composition and Communication, 2014; NCTE, 2013).

This debate over AEE's virtues and ills is compounded by two related issues. First, there is a dearth of research on AEE used for the purpose of formative assessment—i.e., assessment for, rather than of learning (Black & William, 2009). By far, the majority of research has focused on the psychometric properties of the automated scoring engine, rather than documenting evidence that automated feedback is associated with desired changes in teacher feedback practices or students' writing motivation or writing quality (Stevenson & Phakiti, 2014). Indeed, a recent chapter on the formative use of AEE in the Handbook of Writing Research still primarily discusses the features of the AEE scoring systems and the reliability and agreement of those systems with human essay ratings. The chapter authors acknowledge that research "still needs to be conducted to gain a more comprehensive understanding of the impact of [automated] feedback than can guide best-use practice" (Shermis, Burstein, Elliot, Miel, & Foltz, 2015, p.406). Second, previous research has most often examined the effects of automated feedback in isolation of teacher feedback (Stevenson & Phakiti, 2014). Such designs lack ecological validity and may inadvertently bolster fears that adoption of AEE will replace teachers as feedback agents.

Given the controversies surrounding the use of instructional feedback and AEE, as well as the dearth of prior research focusing on the intended usage of AEE, the current study was designed to explore the implications for instruction and student performance when teacher feedback on writing was combined with automated feedback. Specific outcomes of interest included the amount, type, and level of teacher feedback; students' writing motivation; and final-draft writing quality. To further provide context for this study, three areas of prior research will be discussed: (1) categorization of teacher feedback on writing, (2) effects of teacher and automated feedback on writing motivation, and (3) effects of teacher and automated feedback on writing quality.

2. Background

2.1. Categorizing teacher feedback on writing

Teacher feedback on writing is commonly categorized as having at least two components: feedback type and feedback level. *Feedback type* relates to the manner in which feedback is presented to the student. A common distinction is between direct and indirect feedback (Biber et al., 2011; Black & William, 1998; Cho, Schunn, & Charney, 2006; DeGroff, 1992; Shute, 2008). Direct feedback (i.e., directives) involves teachers making a correction or directly telling students what needs to be