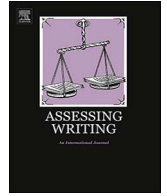


Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Assessing Writing

journal homepage: www.elsevier.com/locate/asw

Full length article

Automated formative writing assessment using a levels of language framework

Joshua Wilson^{a,*}, Rod Roscoe^b, Yusra Ahmed^c^a School of Education, University of Delaware, 213E Willard Hall Education Building, Newark, DE 19716, United States^b Human Systems Engineering, Arizona State University-Polytechnic, United States^c Department of Psychology and Texas Institute for Measurement, Evaluation, and Statistics (TIMES), University of Houston, United States

ARTICLE INFO

Keywords:

Automated writing evaluation (AWE)
Coh-Metrix
Formative assessment
Natural language processing
Writing assessment

ABSTRACT

This study investigates a novel approach to conducting formative writing assessment that involves evaluating students' writing skills across three levels of language (word, sentence, and discourse) using automated measures of word choice, syntax, and cohesion. Writing from students in Grades 6 and 8 ($n = 240$ each) was analyzed using Coh-Metrix. Multigroup confirmatory factor analysis evaluated a hypothesized three factor levels of language model, and multigroup structural equation modeling determined if these factors predicted performance on a state writing achievement test comprised of a Direct Assessment of Writing (DAW) and an Editing and Revising test (ER). Results indicated that a subset of 9 Coh-Metrix measures successfully modeled three latent levels of language factors at each grade level. Results also indicated that the DAW test was predicted by the latent Discourse factor and the ER test was predicted by the latent Discourse and Sentence factors. Findings provide a proof of concept for automated formative assessment using a levels of language framework. Furthermore, although not the primary goal of the study, results may lay the groundwork for new levels of language detection algorithms that could be incorporated within automated writing evaluation software programs to expand automated + teacher assessment and feedback approaches.

1. Introduction

Given the broad importance of writing in the 21st Century (Bellanca & Brandt, 2010; Pellegrino & Hilton, 2013), a daunting and disheartening finding is that many primary and secondary grade students in the United States are struggling or failing to achieve grade-level writing proficiency. Recent data from the U.S. National Assessment of Educational Progress (NAEP) shows that more than two thirds of students in grades four, eight, and twelve fail to achieve grade-level proficiency in writing (National Center for Education Statistics, 2012; Persky, Daane, & Jin, 2002). Such challenges have inspired a substantial effort to identify and research instructional and assessment practices that can help students achieve higher levels of writing proficiency (e.g., Graham, McKeown, Kiuahara, & Harris, 2012; Graham, MacArthur, & Fitzgerald, 2013; Graham & Perin, 2007). Research by Graham and colleagues, for instance, has repeatedly shown that students benefit from explicit strategy instruction that emphasizes core writing knowledge, goals, and processes (i.e., Self-Regulated Strategy Development; Harris, Graham, Mason, & Friedlander, 2008). We also know that writing development requires meaningful opportunities for deliberate practice over time (Kellogg, 2008; Kellogg & Raulerson, 2007; Kellogg & Whiteford, 2012).

Formative feedback and assessment represent a third set of key principles of writing instruction. Formative feedback refers to

* Corresponding author.

E-mail address: joshwils@udel.edu (J. Wilson).

messages that reveal underlying knowledge, gaps between current and expected performance, and actionable methods and strategies for improvement and success (Shute, 2008). In contrast to summative feedback that indicates students' success or failure in writing, formative feedback emphasizes how to grow as a writer (McGarrell & Verbeem, 2007). In order to provide such feedback, however, instructors must conduct formative assessments. That is, educators must evaluate students' work in ways that reveal key needs and areas for improvement that can be addressed via instruction (Black & William, 2009; Popham, 2008). In two recent meta-analyses, Graham and colleagues reviewed research on the effectiveness of formative writing assessment for students in Grades 1 through 8. The authors concluded that formative writing assessment indeed supports the teaching and learning of writing (Graham, Harris, & Hebert, 2011a; Graham, Hebert, & Harris, 2015), with positive effect sizes (see Cohen, 1992) as large as 0.87 for feedback given by adults.

Although formative assessment can be beneficial to writing, these benefits are not guaranteed (see Biber, Nekrasova, & Horn, 2011; Kluger & DeNisi, 1996) and appear to be dependent on the assessment tool that educators use (Graham et al., 2011a). For instance, holistic scoring is efficient, but insufficiently informative for guiding nuanced feedback. Multi-trait (Hamp-Lyons, 1986; Hamp-Lyons & Henning, 1991) or analytic scoring (Diederich, 1974) requires raters to assign separate scores for different aspects of text quality using multiple rubrics. However, this method is time consuming and evidence suggests that analytic trait ratings do not measure distinct dimensions of writing ability, limiting their utility for formative feedback (Carr, 2000; Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Schoonen, 2005).

Furthermore, formative assessments may be insufficiently reliable to inform valid, effective instructional feedback. Formative assessments that rely on human scoring are susceptible to issues related to poor inter-rater reliability, poor intra-rater reliability (i.e., rater drift), differences in rater severity or leniency, the tendency for one feature of a text to unduly influence raters' judgment (i.e., halo effect), and the tendency that raters overestimate the writing skills of students whose texts are written neatly (i.e., presentation effects) (Graham et al., 2011a). In other words, one challenge for formative assessment is that *humans* are typically the sources of evaluation. While training is essential for aligning raters' concepts of writing traits or rubric definitions, it may not eliminate all biases or challenges such as fatigue (Jonsson & Svingby, 2007; Panadero & Jonsson, 2013). Humans are not machines and assessing writing is not easy.

Therefore, in this paper, we consider how automated tools might be used to facilitate, and perhaps even augment, formative writing assessment for middle-grade students. Although writing may be a fundamental skill that both predates and participates in 21st-century success (see Dede, 2010), modern automated computational and linguistic tools may afford a 21st-century approach to writing assessment, feedback, and instruction. Specifically, we seek to provide a proof of concept for assessing writing using automated measures within a levels of language framework.

1.1. The levels of language framework

The *levels of language* framework is based on cognitive theories of writing that portray writing ability as hierarchically structured: component writing skills and cognitive processes are associated with the expression of writing ability at the word, sentence, and discourse levels (Abbott, Berninger, & Fayol, 2010; Berninger & Garvey, 1982; Berninger, Mizokawa, Bragg, Cartwright, & Yates, 1994; Hayes & Berninger, 2014; Whitaker, Berninger, Johnston, & Swanson, 1994). These levels allow for assessing and characterizing students' writing proficiency in ways that may effectively guide feedback and instruction.

Word-level writing skills include spelling and the use of word choice to convey precise meaning and desired stylistic effects (Olinghouse, & Wilson, 2013; National Assessment Governing Board, 2010). These skills are supported by the development of motor skills, orthographic and phonological skills, and linguistic skills such as vocabulary knowledge and morphological awareness (Abbott & Berninger, 1993; Berninger et al., 1992; Berninger & Rutberg, 1992; Dockrell, Lindsay, & Connelly, 2009; McCutchen, 1984). In terms of formative assessment and feedback, students' word-level challenges might be addressed through interventions that target nuanced word knowledge, such as morphological instruction (e.g., Bowers, Kirby, & Deacon, 2010) and explicit vocabulary instruction (McKeown & Beck, 2004; Loftus & Coyne, 2013).

Sentence-level writing skills refer to (a) constructing sentences that are syntactically, grammatically, and mechanically correct; and (b) varying sentence structure and length for stylistic effect (NAGB, 2010). These skills are supported by fluency in word-level skills, by linguistic skills related to syntax and semantics (McCutchen, 1988; Scott, 2009), and by phonological and orthographic skills (Berninger & Swanson, 1994). When formative assessments reveal difficulties at this level, students may benefit from feedback or instruction on strategies such as sentence combining (e.g., Datchuk & Kubina, 2012; Limpo & Alves, 2013).

Finally, discourse-level writing skills refer to producing texts that are well organized, cohesive, have adequate idea development and elaboration, and achieve communicative goals. Discourse-level writing skills are supported by fluency in lower level writing skills (i.e., both word- and sentence-level skills); by strategically enacting the cognitive processes of planning and proposing, translating, transcribing, and reviewing (Flower & Hayes, 1980; Graham, Harris, & Mason, 2005; Hayes, 2012); and by activating relevant topic and discourse knowledge stored in long-term memory (Bereiter & Scardamalia, 1987; Hayes, 2012). Formative assessments that detect discourse-level problems may address these issues via targeted feedback, such as focusing students' attention on content and audience awareness (Midgette, Haria, & MacArthur, 2008), better advanced planning (Limpo & Alves, 2013), and increasing students' knowledge of genre-specific rhetorical structures (Philippakos & MacArthur, 2016).

The levels of language framework has been used productively to guide instructional diagnosis in samples of struggling writers and students with disabilities (Berninger, Mizokawa, & Bragg, 1991; Berninger, Stage, Smith, & Hildebrand, 2001; Nelson and Van Meter, 2007; Scott, 2002, 2009). For example, in two seminal studies, Berninger and colleagues documented evidence of intraindividual differences in writing ability across levels of language for middle schoolers. Specifically, the authors reported small, and non-

Download English Version:

<https://daneshyari.com/en/article/4935751>

Download Persian Version:

<https://daneshyari.com/article/4935751>

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