



# Comparing the accuracy of different scoring methods for identifying sixth graders at risk of failing a state writing assessment<sup>☆</sup>



Joshua Wilson<sup>a,\*</sup>, Natalie G. Olinghouse<sup>b</sup>, D. Betsy McCoach<sup>b</sup>,  
Tanya Santangelo<sup>c</sup>, Gilbert N. Andrada<sup>d</sup>

<sup>a</sup> School of Education, University of Delaware, Newark, DE, United States

<sup>b</sup> Department of Educational Psychology, University of Connecticut, Mansfield, CT, United States

<sup>c</sup> Department of Education, Arcadia University, Glenside, PA, United States

<sup>d</sup> Psychometrics and Applied Research, Bureau of Student Assessment, Connecticut State Department of Education, Hartford, CT, United States

## ARTICLE INFO

### Article history:

Received 20 March 2015

Received in revised form 15 June 2015

Accepted 27 June 2015

Available online 11 August 2015

### Keywords:

Classification accuracy

At-risk students

Automated essay scoring

Scoring methods

Project Essay Grade

Benchmark tests

ROC curve analysis

## ABSTRACT

Students who fail state writing tests may be subject to a number of negative consequences. Identifying students who are at risk of failure affords educators time to intervene and prevent such outcomes. Yet, little research has examined the classification accuracy of predictors used to identify at-risk students in the upper-elementary and middle-school grades. Hence, the current study compared multiple scoring methods with regards to their accuracy for identifying students at risk of failing a state writing test. In the fall of 2012, students composed a persuasive prompt in response to a computer-based benchmark writing test, and in the spring of 2013 they participated in the state writing assessment. Predictor measures included prior writing achievement, human holistic scoring, automated essay scoring via Project Essay Grade (PEG), total words written, compositional spelling, and sentence accuracy. Classification accuracy was measured using the area under the ROC curve. Results indicated that prior writing achievement and PEG Overall Score had the highest classification accuracy. A multivariate model combining these two measures resulted in only slight improvements over univariate prediction models. Study findings indicated that choice of scoring method affects classification accuracy, and automated essay scoring can be used to accurately identify at-risk students.

© 2015 Elsevier Inc. All rights reserved.

## 1. Introduction

In light of evidence that the majority of U.S. 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), a growing body of research has focused on methods of identifying struggling writers in need of intervention in the early grades (K–2), before writing difficulties

<sup>☆</sup> The authors would like to thank Trish Martin, Fran Brown, Andrew O'Neill, and Tiwana Bazemore of Measurement Incorporated for their assistance in preparing the datasets used for analysis. Thanks also to Andrew Petsche and Samantha Evans for assistance scoring.

\* Corresponding author. Tel.: +1 3028312955.  
E-mail address: [joshwils@udel.edu](mailto:joshwils@udel.edu) (J. Wilson).

become severe and intractable (Coker & Ritchey, 2014; Fewster & McMillan, 2002; Gansle et al., 2004; McMaster, Parker, & Jung, 2012; Parker, Tindal, & Hasbrouck, 1991; Ritchey & Coker, 2014). However, the need to develop predictive models that identify struggling writers does not disappear as students enter upper-elementary and middle-school grades.

As students progress through these grades (G4–8), they transition from developing lower-level writing skills – handwriting, spelling, sentence construction, grammar, and punctuation – to developing higher-level writing skills, such as utilizing genre-specific methods of idea-development, organization, and word choice (Berninger, Abbott, Whitaker, Sylvester, & Nolen, 1995). Coordinating these low and high-level skills strains working memory and may impact writing performance (Flower & Hayes, 1980; Kellogg & Whiteford, 2009; McCutchen, 1996, 2011). Consequently, students in upper elementary and middle school may be at risk of developing, or worsening, writing difficulties.

In addition, students in these grades must participate in state and national accountability assessments which are used to determine whether students have attained grade-level standards (Hamilton et al., 2007). Performance on these assessments has a number of consequences for students, such as: (a) being assigned to particular schools, programs or classes (i.e., academic tracking) (Decker & Bolt, 2008; Goertz & Duffy, 2003); (b) being referred for additional instructional support (Graham, Hebert, & Harris, 2011b; Jones et al., 1999); and (c) being retained or promoted to the next grade (Darling-Hammond, 2004; Hamilton et al., 2007). Indirect consequences associated with repeated failure of accountability assessments include increased risk for school dropout (Heubert & Hauser, 1999) or referral to special education (Fiiglio & Getzler, 2002; Haney, 2000).

Thus, given the developmental challenges faced by upper-elementary and middle-grade students with regards to writing proficiency, and the direct and indirect consequences associated with poor performance on accountability assessments, it is important to develop predictive models that accurately identify students at risk for writing failure. Once identified, at-risk students may receive intervention to prevent and remediate their writing difficulties. While an emerging body of research has focused on identifying at-risk writers in the early grades (K–2), there is little research to guide educators in developing accurate prediction models for students in upper elementary or middle school. Hence, the present study compared measures of sixth-graders' writing ability with regard to classification accuracy, i.e., the accuracy of predicting which students passed or failed a state writing test.

### 1.1. *Prior research examining predictors of performance on state writing tests*

The majority of prior research examining predictors of performance on state writing tests has focused on assessment procedures and scoring measures associated with curriculum-based measurement for writing (W-CBM). For example, a study of fourth-grade students examined correlations between writing quality measures derived from two three-minute CBM writing probes and subtest scores of the Louisiana Educational Assessment Program writing test (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). Writing probes were scored for 12 W-CBM measures, two computer-scored measures of text readability – the Flesch Reading Ease score and Flesch–Kincaid grade-level – and computer-scored measures of sentence and vocabulary complexity. Only number of verbs, words spelled correctly, and correct word sequences (CWS) demonstrated statistically significant correlations with scores of the state test:  $r = .33$ ,  $.29$ , and  $.41$ , respectively.

McMaster and Campbell (2008) sampled fifth-grade students who completed two passage copying tasks, two picture prompts, two narrative prompts, and two expository prompts. Each of the writing tasks were scored for total words written (TWW), words spelled correctly (WSC), correct word sequences (CWS), and correct minus incorrect word sequences (C-IWS). Evidence of criterion validity with the Minnesota state writing test differed by scoring metric: TWW and WSC were not statistically significantly correlated with state test performance for any of the writing tasks; CWS was moderately correlated (range  $r = .54$ – $.56$ ) for the three-minute and five-minute narrative writing tasks and the five-minute expository writing task, but for no other writing task; and C-IWS was moderately correlated (range  $r = .54$ – $.68$ ) for the three-minute and five-minute narrative and expository tasks. Similar results were reported in a study of eighth-grade students (Espin et al., 2000): moderate correlations were found between TWW, WSC, CWS, IWS, and C-IWS scored from two story writing and two descriptive writing samples and a district writing test.

Finally, Lopez and Thompson (2011) sampled students in grades 6–8 who responded to a story starter scored for CWS and who participated in the Arizona state writing assessment. CWS was not a statistically significant predictor for grade six, but demonstrated moderate correlations with the criterion measure for grades seven and eight. The authors also examined how accurately a CWS cutscore of one standard deviation below the mean identified students who scored at the “Does not meet expectations” level on the Arizona state writing test. They reported classification accuracy of 75% for grade six, 87% of for grade seven, and 96% for grade eight. Classification accuracy is the percent of students correctly classified as true positives or true negatives, i.e., as truly at risk or truly not at risk. However, classification accuracy is a misleading measure of diagnostic accuracy when the base rate (i.e., prevalence) of a condition is low (Meehl & Rosen, 1955; Wilson & Reichmuth, 1985). When the base rate is low, it is possible to achieve high classification accuracy by simply diagnosing all students as not at risk. In the Lopez and Thompson study, the base rates of students failing the state writing test were 17%, 22%, and 4%, respectively, for the sixth, seventh, and eighth-grade samples. Thus, classification accuracy rates of 83%, 78%, and 96% would have been achieved by simply assuming that no students were at-risk. To warrant utility for making selection/screening decisions, a measure should yield classification accuracy values significantly better than those obtained by identifying no at-risk students (Johnson, Jenkins, Petscher, & Catts, 2009), which was not the case in this study.

Download English Version:

<https://daneshyari.com/en/article/344199>

Download Persian Version:

<https://daneshyari.com/article/344199>

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