Exploring Writing Analytics and Postsecondary Success Indicators

Jill Burstein
Educational Testing Service
jburstein@ets.org

Daniel McCaffrey
Educational Testing Service
dmccaffrey@ets.org

Beata Beigman Klebanov
Educational Testing Service
bbeigmanklebanov@ets.org

Guangming Ling
Educational Testing Service
gling@ets.org

Steven Holtzman
Educational Testing Service
sholtzman@ets.org

ABSTRACT: Writing is a challenge and a potential obstacle for students in U.S. 4-year postsecondary institutions lacking prerequisite writing skills. Building on *Anonymous*, we collected authentic coursework writing from students enrolled at six 4-year colleges, extracted natural language processing (NLP) writing features (analytics), and examined relationships between analytics and college grade point average (GPA). Consistent with *Anonymous*, findings suggest that NLP writing analytics may contribute to college GPA prediction. Implications are that real-time NLP writing analytics from authentic coursework writing from students could be leveraged to efficiently track success and flag potential obstacles during students' college careers.

KEYWORDS: **natural language processing, writing analytics, higher education**

Research presented in this paper was supported by the Institute of Education Science, U.S. Department of Education, Award Number R305A160115.

Poster presentation published in *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge* (LAK19), 213-214, 2019.

Accessible online at: https://drive.google.com/file/d/1E4ry7M8C8Hoyb6M0KM8vybibeyGLiQfP/view

Exploring Writing Analytics and Postsecondary Success Indicators

Jill Burstein

Educational Testing Service

jburstein@ets.org

Daniel McCaffrey

Educational Testing Service

dmccaffrey@ets.org

Beata Beigman Klebanov

Educational Testing Service

bbeigmanklebanov@ets.org

Guangming Ling

Educational Testing Service

gling@ets.org

Steven Holtzman

Educational Testing Service

sholtzman@ets.org

ABSTRACT: Writing is a challenge and a potential obstacle for students in U.S. 4-year postsecondary institutions lacking prerequisite writing skills. Building on *Anonymous*, we collected authentic coursework writing from students enrolled at six 4-year colleges, extracted natural language processing (NLP) writing features (analytics), and examined relationships between analytics and college grade point average (GPA). Consistent with *Anonymous*, findings suggest that NLP writing analytics may contribute to college GPA prediction. Implications are that real-time NLP writing analytics from authentic coursework writing from students could be leveraged to efficiently track success and flag potential obstacles during students' college careers.

KEYWORDS: **natural language processing, writing analytics, higher education**

1 INTRODUCTION

Writing is a challenge and postsecondary students who lack prerequisite writing skills may not persist in U.S. 4-year postsecondary institutions (NCES, 2012). Previous work has found statistically-significant relationships between reading comprehension and writing features in postsecondary contexts (Allen et al, 2014). Studies related to reflective writing reveal relationships between reflective writing features, learning, and college success

outcomes (Gibson et al, 2017; Beigman Klebanov et al, 2017). Consistent with *Anonymous*, preliminary findings presented here further suggest that NLP writing analytics generated from authentic coursework writing assignments are predictors of college GPA. The broader implication is that analytics may be applied to authentic student writing in college and, in turn, may serve to efficiently track success and obstacles throughout college.

2 METHODS

Participants. Authentic coursework writing was collected from 693 students enrolled in first-year courses who participated across the 2017-18 academic year at 6 4-year postsecondary sites. Writing samples represented 7 academic disciplines across Social Sciences, Humanities and STEM.

Data. Nine-hundred and thirty-two assignments were collected. As this analysis represents a *slice* of a larger study, we examine writing submissions from a subset of students (N=369) who completed multiple required study tasks.

Table 1. College GPA Writing Analytics Predictors (N=369)

Variable	Standardized Coefficient	p-value	R^2	Inc. R^2
personal reflection	-0.17	0.00	0.27	0.02
vocabulary choice	0.20	0.00	0.28	0.03
vocabulary sophistication	0.18	0.00	0.28	0.03
discourse structure	0.17	0.00	0.26	0.01

Analysis. Thirty-six NLP features were automatically extracted from each writing assignment. Features represented writing construct features (e.g. *argumentation, coherence, discourse, grammar, and vocabulary*). Using the NLP feature values, we ran a separate hierarchical linear model analysis that contained: 1) one NLP analytics feature, plus 2) *length* (*square root of number of words in the text*), plus 3) *school site*. Each NLP feature plus *length* comprised the independent (or predictor) variables, and college GPA was the dependent variable. We control for length to ensure that features are not length proxies, and school to control for *site effects* in GPA.

Results and Discussion. Table 1 shows a subset NLP writing feature models as an illustration of NLP features (analytics) that were predictive of college GPA and where the p-value < 0.01. **The R^2 baseline (length+site-only model) for college GPA is equal to 0.25**. Table 1 illustrates features related to vocabulary (personal reflection, vocabulary choice and vocabulary sophistication), discourse structure, and mechanics errors were predictive of GPA. These analytics are aligned with writing domain knowledge that is essential to master for college writing (Anonymous). Implications of these findings suggest that real-time NLP writing analytics generated on authentic coursework writing from college students could be leveraged during students' college careers to track success and flag obstacles.

ACKNOWLEDGMENTS

Research presented in this paper was by the Institute of Education Science, U.S. Department of Education, Award Number [TBA]. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the IES. *Others TBA.*

REFERENCES

Anonymous.

Allen, L. K., Snow, E. L., Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014). Reading comprehension components and their relation to writing. *L'Année psychologique*, *114*(4), 663-691.

Beigman Klebanov, B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective Writing About the Utility Value of Science as a Tool for Increasing STEM Motivation and Retention–Can AI Help Scale Up?. *International Journal of Artificial Intelligence in Education*, *27*(4), 791-818.

Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). Reflective writing analytics for actionable feedback.

NCES. National Center for Education Statistics (2012). The nation's report card: Writing 2011 (NCES 2012–470). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf