Lexical Sophistication, Learning, and Engagement in Math Problems

Abstract

This study uses correlation mining to investigate relationships between the linguistic properties of math problems and student outcomes. We find that linguistic properties associated with boredom were negatively associated with engaged concentration, an emotion which is boredom's inverse in terms of activation (intensity of emotion) and valence (positivity of emotion). However, few of the linguistic features associated with gaming the system correlated with poorer moment-by-moment-learning. These findings have potential implications for mathematics problem design.

Lexical Sophistication, Learning, and Engagement in Math Problems

Introduction:

Research shows the relationships between the linguistic properties of texts and student measures of text comprehension (McNamara et al., 2014) and affective states (Bixler & D'Mello, 2013) to be of critical importance, but these relations have not been fully explored in the domain of mathematics. This paper aims to address this gap by exploring how the lexical sophistication (Kyle & Crossley 2015) of mathematics problems written by teachers and curriculum developers relates to student learning and engagement. Specifically, we use data from an online tutor to examine how these linguistic measures correlate with boredom, confusion, engaged concentration, frustration, gaming the system, and moment-by-moment calculations of learning—all measures that have been previously used to make long-term predictions about student outcomes (e.g., Ocumpaugh et al., 2014). Findings from this study can provide useful feedback for teachers and curriculum developers in the area of problem design.

Methods:

Data. Mathematics word problems, which were written by teachers and curriculum developers, (n=114,168) were collected from ASSISTments, an online learning system for middle school mathematics (Heffernan & Heffernan, 2014).

Learning/Engagement (Outcome) Measures. Learning and engagement measures were collected from 22,225 students who used ASSISTments during the 2012-13 school year. These outcomes are assessed retroactively using six previously-validated automated measures that are based on students' interactions with ASSISTments. Affective engagement was assessed using Ocumpaugh et al.'s (2014) models of *boredom, confusion, engaged concentration*, and *frustration*. Learning was assessed for each mathematics problem using Baker et al.'s (2011) *Moment-By-Moment Learning* (MBML) model. Disengaged behavior, namely *gaming the system* by trying to complete problems by systematic guessing or rapidly asking for hints, was assessed using Pardos et al.'s (2013) model. These models were originally validated using *in situ* classroom observations.

Lexical Sophistication Measures. Lexical characteristics of the word problems were assessed using the Tool for Automated Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015). We calculated 137 of 485 TAALES indices for each word problem in our corpus.

Analyses: The six learning and engagement measures listed above were applied to each student/problem combination. Then, values for each measure were aggregated across every student who saw each problem, generating problem-level averages. Next, each average for a given problem was correlated to the TAALES measures for the same problem. Due to non-normality, Spearman's $\boldsymbol{\rho}$ was used, and due to the large number of correlations generated, Benjamini & Hochberg's correction was used to prevent spurious results, which still resulted in 718 of 822 significant correlations.

Results:

All Outcome Measures. Table 1 presents the ten TAALES indices that show the strongest correlations across all outcomes, calculated by averaging the absolute value of Spearman **ρ** across all learning/engagement measures.

Table 1. Largest Correlations by Avg. |ρ|.

TAALES Feature	Avg. |ρ|	Confusion	Frustration	Boredom	Engaged Concentration	MBML	Gaming
Trigram Type Count	0.117	0.083	0.104	0.089	-0.169	0.193	0.061
Bigram Type Count	0.117	0.082	0.103	0.090	-0.170	0.193	0.062
Word Count	0.116	0.084	0.104	0.089	-0.168	0.193	0.060
Word Count for N-grams	0.116	0.084	0.104	0.089	-0.168	0.192	0.061
BNC Written Bigram Freq Log	0.103	0.080	0.093	0.072	-0.149	0.178	0.047
BNC Spoken Bigram Freq Log	0.103	0.077	0.091	0.072	-0.147	0.180	0.048
BNC Spoken Trigram Freq Log	0.087	0.069	0.093	0.052	-0.102	0.167	0.036
BNC Written Trigram Freq Log	0.083	0.066	0.088	0.050	-0.098	0.163	0.034
BNC Spoken Trigram Normed (tri) Freq	0.069	-0.070	-0.075	-0.051	0.097	-0.089	-0.029
BNC Written Trigram Freq Normed (tri) Log	0.065	0.047	0.063	0.048	-0.098	0.115	0.022

Similarities and Differences across outcome Measures.

The results of our analyses revealed that the TAALES features that correlated positively with MBML were negatively correlated with *engaged concentration*, suggesting that students may not experiencing this affective state at key learning moments. This finding is surprising, given that better learners tend to experience more positive emotions during learning overall (Pardos et al., 2013).

Further, although *gaming the system* is often negatively correlated with long-term learning outcomes (e.g. Pardos et al., 2013), we did not find that the lexical features associated with gaming the system were associated with poorer immediate learning. In fact, common written trigrams were negatively correlated with both gaming and MBML. MBML was instead correlated with features using N-gram calculations, whereas gaming was correlated with features related to function words.

In comparisons across affective states, we find that *confusion* and *frustration* overlap considerably in terms of strongly correlated TAALES features, especially those associated with concreteness. Interestingly, Liu et al., (2013) shows that moderate levels of both affective states are associated with improved learning outcomes (Liu et al., 2013), while McNeil et al., (2009) finds that maintaining moderate levels of concreteness improves learning. As such, it follows that varying the level of concreteness could be used to help maintain optimal levels of confusion and frustration.

Likewise, *boredom* and *engaged concentration*, which are the inverse of each other both in activation (intensity of emotion) and valence (positiveness of emotion), show expected parallels. Features with the strongest negative correlations to boredom were positively related to concentration (and *vice versa*). Of particular interest, common content words were associated with lower boredom and more concentration, suggesting that perhaps problems should be designed to have more common content words when possible.

Conclusion and Implications:

Word problems are written by many mathematics teachers and curriculum developers, with different ideas and expectations for design. This study examines how the lexical sophistication of such math problems correlates with student learning and engagement. Our findings suggest that simple adjustments to design may be able to improve engagement and learning. We intend to follow up this research with a series of experiments to modify the text of mathematics problems in ASSISTments in line with our findings, to see if engagement and learning outcomes are indeed enhanced.

References

Baker, R.S., Goldstein, A.B., Heffernan, N.T. (2011). Detecting learning moment-by-moment. *International J. Artificial Intelligence in Ed*, *21*(1-2), 5-25.

Bixler R., D'Mello. S. (2013). Towards Automated Detection and Regulation of Affective States During Academic Writing. *Artificial Intelligence in Education*, H. Chad Lane, K. Yacef, J. Mostow, P. Pavlik. Springer Berlin Heidelberg, 904-7.

Heffernan, N., Heffernan, C. (2014). The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International J. Artificial Intelligence in Ed*, *24*(4), 470-497.

Kyle, K., Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*(4), 757-786.

Ocumpaugh, J., Baker, R.S., Gowda, S., Heffernan, N., Heffernan, C. (2014). Population Validity for Educational Data Mining models: A case study in affect detection. *British J. Ed. Tech.*, *45*(3), 487-501.

Pardos, Z.A., Baker, R.S., San Pedro, M.O., Gowda, S.M., Gowda, S.M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proc.3rd International Conf. Learning Analytics & Knowledge,* 117-124.

Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R. (2013). Sequences of frustration and confusion, and learning. *Proc. 6th International Conf. EDM*, 114-120.

McNamara, D., Graesser, A., McCarthy, P., Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge: Cambridge U.P.

McNeil, N., Uttal, D., Jarvin, L., Sternberg, R. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning & Instruction*, *19*(2), 171-184.