Letting the Genie out of the Lamp: Using Natural Language Processing Tools to Predict Math Performance

Scott Crossley[1], Victor Kostyuk[2]

[1] Georgia State University, Applied Linguistics & ESL, Atlanta, GA USA
scrossley@gsu.edu
[2] Reasoning Mind, Houston, TX USA
vak@reasoningmind.org

Abstract. This study examines links between natural language processing and its application in math education. Specifically, the study examines language production and math success in an on-line, blended learning math program. Unlike previous studies that have relied on correlational analyses between linguistic knowledge tests and standardized math tests or compared math success between proficient and non-proficient speakers of English, this study examines the linguistic features of students' language production while e-mailing a virtual pedagogical agent. In addition, the study examines a number of non-linguistic features such as grade and objective met within the program. The findings indicate that linguistic features related to the use of standardized language use explain around 8% of math success. These linguistic features outperform non-linguistic features.

Keywords: Natural Language Processing, Online Tutoring Systems, Math Education, Text Analytics

1 Introduction

A number of cognitive skills are necessary for young students to be successful in the math classroom. Primarily, research has focused on skills that strongly overlap with math knowledge including spatial attention and quantitative ability [1]. Less attention has been paid to supporting cognitive skills such as language ability. However, some researchers argue that language skills are needed to transfer cognitive operations between math and language domains.

In support of this notion, researchers have begun to examine links between language skills and math success with the understanding that student with greater language abilities are likely able to better engage with math concepts and problems. More specifically, this research is premised on the notion that success in the math classroom is at least partially explained through language development that allows students to constructively participate in math discussions, understand and solve word problems, as well as quantitatively engage with math problems that arise outside of the classroom [2, 3]. In a similar fashion, it is argued that math literacy is not just having knowledge of numbers and symbols, but also having the language skills to understand the discourse of math (i.e., the words surrounding the numbers and symbols) [4]. While other cognitive skills are also critical to math success

Until recently, studies linking language skills to math success in the classroom generally relied on correlational analyses among standardized tests of math and linguistic knowledge. For instance, several studies have examined links between tests of language proficiency (e.g., syntax, knowledge, verbal ability, and phonological skills) and success on tests of math knowledge (e.g. algebraic notation, procedural arithmetic, and arithmetic word problems [2, 5]. Other studies have compared success on standardized math tests between first language (L1) speakers of English and second language speakers of English, who have lower linguistic ability [6, 7, 8]. However, the majority of studies have not examined the actual language produced by students and the relationship between the complexity of this language and success on math assessments (see [9] for an exception).

This study builds on the work of Crossley et al. [9] by examining links between the affect and complexity of language produced by students in an e-mail system used within a math intelligent tutoring system and students' math success within the system. To do so, we examine students' emails within the systems for a number of linguistic features related to text cohesion, lexical sophistication, and affect derived from natural language processing (NLP) tools. The goal of this study is to examine the extent to which the language features produced by students are predictive of their math success within the tutoring system. In addition to the linguistic features, we also examine a number of non-linguistic factors that are potentially predictive of math success including grade, number of messages sent and received by the students, hours spent on-line, and number of objective met within the platform.

1.1 Linking Language and Math

Previous studies have examined links between language proficiency and math skills in native speakers (NS) of English. These studies generally demonstrate strong links between math ability and language ability. As an example, Macgregor and Price [5] analyzed relations between three cognitive indicators of language proficiency (syntax, metalinguistic awareness of symbols, and language ambiguity) and understanding of algebraic notation. Those students who scored high on the algebra test also scored well on language tests. A follow-up study using more difficult algebra found a stronger relationship between algebraic notation and language ability. The authors concluded that low metalinguistic awareness was negatively related to algebra learning. Similarly, Vukovic and Lesaux [2] examined links among arithmetic knowledge (arithmetic word problems and procedural arithmetic), symbolic number skills, and linguistic skills (i.e., phonological skills and general verbal ability) in NS students. They also included control variables consisting of visual–spatial ability and working memory. The participants comprised 287 third graders using the same math curriculum from five different schools. Their results showed links between linguistic and math skills, but that the linguistic skills differed in their degree of relation with arithmetic knowledge. For example, general verbal ability was indirectly related through symbolic number skills while phonological skills were found to be directly related to arithmetic knowledge. Vukovic and Lesaux argued that general verbal ability was related to how children reason numerically but that phonological skills were related to executing arithmetic problems.

Other research has investigated indirect links between math and language skills. Hernandez [10] examined relationships between reading ability and math achievement levels under the presumption that there was a positive correlation between math scores and reading skills. In his study, he analyzed 652 ninth-grade students' scores from the reading and math sections of the *Texas Assessment of Knowledge and Skills.* Correlations between the math scores and the reading scores were calculated for tests taken in sixth, seventh, and eighth grades. The results demonstrated significant positive correlations between reading ability and math achievement. These findings led Hernandez to recommend that students' reading skills should factored to provide more effective math instruction, especially for poor readers.

Not all studies have found significant links between math knowledge and language skills. For instance, LeFevre et al. [1] conducted a longitudinal study of 182 NS children ages 4 to 8 (37 in preschool and 145 in kindergarten) that followed the children's math progress. Data collection from a year including non-linguistic skills such as spatial attention, early numeracy skills (nonlinguistic arithmetic and number naming) quantitative knowledge and linguistics skills (phonological awareness and receptive vocabulary). Dependent variables included research-based and standardized tests of math ability. The results indicated that linguistic skills were significantly related to number naming, that quantitative abilities were related to processing numerical magnitudes, and that spatial attention was related to a variety of numerical and math tests. However, the quantitative abilities and spatial attention results reported that non-linguistic features were stronger predictors of math ability.

In terms of language production, only one study to our knowledge has examined links between the language produced by NS students and their success in the math classroom. Crossley et al. [9] examined the linguistic and affective features of student discourse while students were engaged in collaborative problem solving within an on-line math tutoring system. Student speech was transcribed and natural language processing tools were used to extract linguistic information related to text cohesion, lexical sophistication, and affect. They examined links between the linguistic features and pretest and posttest math performance scores as well as links with a number of non-linguistic factors including gender, age, grade, school, and content focus (procedural versus conceptual). The results indicated that non-linguistic factors are not predictive of math scores but that linguistic features related to cohesion, affect, and lexical proficiency explained around 30% of the variance in the math scores such that higher scoring students produced more cohesive texts that were more linguistically sophisticated.

Beyond studies examining NS math students, a rich source of evidence for connecting language and math abilities are non-native speakers (NNS) of English who are learning math skills in an English classroom. NNS, unlike NS, generally have lower language skills in English and it is argued that these lower language skills will result in lower math skills. The general notion behind this theory is that most NNS have not reached a threshold of language proficiency that allows them the resources to perform on par with NS [11]. This notion is supported by the US Department of Education [7], which reports that over a five-year period (from 1st to 5th grade), NS report higher math scores than proficient NNS who report higher math scores lower proficiency NNS.

A number of studies have supported the findings reported by the US Department of Education report. For instance, Alt et al. [6] investigated relations between language achievement and math among school-age children (ages 7-10) who were grouped into NS, NNS who spoke Spanish as a first language (L1), and students with specific language impairment (SLI). Data included two standardized math tests (one in Spanish and the other in English) and three experimental tasks (quantity comparison, number comparison, and concept mapping games). The tests and tasks were categorized in terms of language, symbol, and visual working memory as either heavy processing (the English math test for the NNS) or light processing (the visual working memory for the NNS). The results indicated that SLI students performed significantly worse than NS in all tests and tasks and that NS significantly outperformed the NNS only in language-heavy tests and games. From these results, Alt et al. concluded that language proficiency is a key component of math success for NNS.

Martinello [8] analyzed item difficulty differences across math tests between NS and NNS by examining pictures and schemas for different levels of linguistic complexity (grammatical and lexical complexity) and contextual support. The items examined comprised 39 questions that assessed knowledge of patterns and relations, algebra, measurement, geometry, probabilities, and number sense and operations. The results indicated that the non-linguistic representations and linguistic complexity of the items accounted for around 66% of the variation in scores between native and non-native students. The findings demonstrated that items with complex grammatical structures and low frequency non-math words were more difficult for NNS. Non-linguistic representations (especially schemas) were found to decrease the difficulty of more linguistically complex items for NNS. Similar findings have been reported in a number of studies [12, 13, 14], all of which indicate that NNS are at a disadvantage in math performance when compared to NS, providing support for the threshold hypothesis [11] that proficiency in English is necessary for achievement in other academic disciplines such as math. These results may hold across NNS of different proficiency levels as well [15].

1.2 Current Study

A number of studies have demonstrated strong links between linguistic knowledge and success in math. Studies examining these links in L1 speakers have traditionally relied on correlational analyses between linguistic knowledge tests and standardized math tests [1, 4, 5]. For L2 speakers, the majority of studies have compared math success between proficient and non-proficient speakers of English [6, 7, 12, 13, 14]. In this study, we take a novel approach and examine the language features of students' language production while e-mailing a virtual pedagogical agent in an on-line math intelligent tutoring system (Reasoning Mind). To derive our language features of interest, we analyzed the language produced by the students using a number of natural language processing tools to extract linguistic information related to text cohesion, lexical sophistication, and sentiment. Thus, in contrast to most previous studies (see [9] for an exception), our interest is not on language performance as measured by standardized tests, but on language performance as a function of language production in student e-mails.

Our criterion variables are students' accuracy on beginning level math problems within the Reasoning Mind system. In addition to examining relations between linguistic features of student language production and math scores, we also control for a number of non-linguistic factors, including grade level, number of messages sent by the student to the avatar, number of messages sent to the teacher and number received by the teacher, hours spent online in the Reasoning Mind platform, and number of objectives met within the Reasoning Mind platform. Thus, in this study, we address two research questions:

1. Are non-linguistic factors significant predictors of math performance in the Reasoning Mind on-line tutoring environment?
2. Are linguistic factors related to lexical sophistication, cohesion, and affect significant predictors of math performance in the Reasoning Mind on-line tutoring environment?

2 Method

2.1 Reasoning Mind

Data was collected from Reasoning Mind *Foundations*, which is a blended learning math program used primarily in grades 2-5. In Foundations classrooms, teachers facilitate the class, while students study on computers, allowing the teacher to conduct both one-on-one and small-group interventions. *Foundations* includes a sequenced main curriculum divided into objectives, each of which introduces a new topic (e.g., the distributive property) using interactive explanations, presents problems of increasing difficulty on the topic, and reviews previously studied topics. (The algorithms and pedagogical logic underlying *Foundations* are described in detail by Khachatryan et al. [16]) All students complete first difficulty level ("Level A") problems – these problems address the basic knowledge and skills in the objective. Students who do well progress to problems of greater difficulty. Other modes in *Foundations* allow students to play math games against classmates, tackle challenging problems and puzzles, and use points earned by solving math problems to buy virtual prizes.

Foundations uses many animated characters to provide a backstory to the mathematics being learned and to deliver emotional support. The main character is the Genie, who helps students when they struggle with independent problem solving by guiding them through the solution and provides praise when they get many problems right in a row. The Genie, who appears in many of the animated stories in the

Foundations system, has a virtual house where students can play games, and students have the option to send email messages to the Genie. These messages are answered in character by Reasoning Mind employees (Student Mentors) who project an empathetic persona (i.e., consistent, warm, and encouraging), model a positive attitude toward learning, and emphasize the importance of practice and hard work for success. Messages reflect an extensive Genie biography, which includes political beliefs (fractions over factions), educational interests (mathematics, physics, disappearing & reappearing, literature, flying, and musicology), and other facts about the Genie. The Genie email system is popular. For instance, the Genie received 157,346 messages from about 40,000 different students in 2014-15.

2.2 Participants and Corpus

The students sampled in this study consists of all *Foundations* students in two academic years, who had written at least 50 words to the Genie through the email system from August 1, 2013 to July 1, 2015. We used the messages sent from the students to the Genie as our language sample for this analysis. Because many of the samples contained few words, we aggregated all e-mails sent by each student to create a representation of individual student's linguistic knowledge. The 50-word threshold provides a sample with enough linguistic coverage for normal distribution of most linguistic features reported by the natural language processing tools used in this study. All samples were cleaned of non-ASCII characters. Misspellings were kept in the data. There were a total of 13,983 such students, in grades 1 through 6 with the majority of students in grades 2 through 5. The students were from 546 different schools located in 107 different districts. Most districts were located in Texas.

2.3 Natural Language Processing Tools

Each transcript was run through a number of natural language processing tools including the Tool for the Automatic Analysis of Lexical Sophistication (TAALES) [17], the Tool for the Automatic Analysis of Cohesion (TAACO) [18] and the SEntiment ANalysis and Cognition Engine (SEANCE) [19]. The selected tools reported on language features related to lexical sophistication, text cohesion, and sentiment analysis respectively. The tools are discussed in greater detail below.

TAALES. TAALES [17] is a computational tool that is freely available and easy to use, works on most operating systems (Windows, Mac, Linux), allows for batch processing of text files, and incorporates over 150 classic and recently developed indices of lexical sophistication. These indices measure word frequency, lexical range, n-gram frequency and proportion, academic words and phrases, word information, lexical and phrasal sophistication, and age of exposure. In terms of word frequency, TAALES reports frequency counts retrieved from Thondike-Lorge [20], Kucera-Francis [21], Brown [22], and SUBTLexus databases [23]. In addition, TAALES derives frequency counts from the British National Corpus (BNC) [24] and the Corpus of Contemporary American English (COCA) [25]. TAALES calculates scores for all words (AW), content words (CW), and function words (FW). In addition to frequency information, TAALES includes a number of range indices which calculate how many texts within a corpus a word appears (i.e., specificity). Range indices are calculated for the spoken (574 texts) and written (3,083 texts) subsets of the BNC, SUBTLEXus (8,388 texts) and Kucera-Francis (500 texts).

TAALES also calculates a number of phrasal indices. These include bigram and trigram frequencies and proportion scores (i.e., the proportion of n-grams in a text that are common in a reference corpus) from BNC and COCA. TAALES also computes strength of association indices between words to measure the conditional probability that words will occur together. These include mutual information scores, *T* values, Delta P, and approximate Colexeme Strength.

Lastly, TAALES reports on a number word information and psycholinguistic scores. The word information scores are derived from the MRC Psycholinguistic Database [26], Kuperman norms [27], Brysbaert norms [28], WordNet [29], the Edinburgh Associative Thesaurus (EAT) [30], the University of South Florida (USF) norms [31], and the English Lexicon Project (ELP) [32]. Word information scores are calculated for word familiarity, concreteness, imageability, meaningfulness, age of acquisition, word association norms, polysemy, hypernymy, orthographic, phonographic, and phonologic neighborhoods.

TAACO. TAACO [18] incorporates over 150 classic and recently developed indices related to text cohesion. For a number of indices, the tool incorporates a part of speech (POS) tagger from the Natural Language Tool Kit [33] and synonym sets from the WordNet lexical database [29]. The POS tagger affords the opportunity to look at content words (i.e., nouns, verbs, adjectives, adverbs) as well as function words (i.e., determiners, propositions). TAACO provides linguistic counts for both sentence and paragraph markers of cohesion and incorporates WordNet synonym sets. Specifically, TAACO calculates type token ratio (TTR) indices (for all words, content words, function words, and n-grams), sentence overlap indices that assess local cohesion for all words, content words, function words, POS tags, and synonyms, paragraph overlap indices that assess global cohesion for all words, content words, function words, POS tags, and synonyms, and a variety of connective indices such as logical connectives (e.g., *moreover, nevertheless*), causal connectives (*because, consequently, only if*), sentence linking connectives (e.g., *nonetheless, therefore, however*), and order connectives (e.g., *first, before, after*).

SEANCE. SEANCE [19] is a sentiment analysis tools that relies on a number of pre-existing sentiment, social positioning, and cognition dictionaries. SEANCE contains a number of pre-developed word vectors to measure sentiment, cognition, and social order. These vectors are taken from freely available source databases. For many of these vectors, SEANCE also provides a negation feature (i.e., a contextual valence shifter) that ignores positive terms that are negated (e.g., not happy). SEANCE also includes a part of speech (POS) tagger.

2.4 Statistical Analysis

We calculated linear models to determine if linguistic features in the students' language output along with other fixed effects (grade, number of messages sent and received, hours online, and objectives met) could be used to predict the students' math scores. In this study, we use accuracy on first difficulty level problems as a proxy to student mastery of the taught curriculum. Prior to LME analysis, we first checked that the linguistic variables were normally distributed. We also controlled for multicollinearity between all the linguistic variables ($r \geq .900$). We used R [34] for our statistical analysis and the package *relaimpo* [35] to report the importance of the individual features in the linear models. Final model selection and interpretation was based on t and p values for fixed effects and visual inspection of residuals distribution. To obtain a measure of effect sizes, we computed correlations between fitted and observed values, resulting in an overall R^2 value for the fixed factors. We first developed a baseline linear model that included non-linguistic fixed effects (e.g., grade, number of messages, objectives met). We next developed a second model that included only linguistic factors (e.g., frequency, word neighborhood effects, number of determiners). We then created a final model that include both linguistic and non-linguistic effects. We compared the strength of each model using Analyses of Variance (ANOVAs) to examine which models were most predictive.

3 Results

3.1 Non-linguistic Linear Model

A linear model considering all non-linguistic fixed effects revealed significant effects for grade level, number of messages sent to Genie, number of messages received from Genie, number of messages from the teacher, and the number of objectives met on math scores. Table 1 displays the coefficients, standard error, t values, p values, and relative importance for each of the non-linguistic fixed effects. The overall model was significant, $F(4, 13597) = 220.80$, $p < .001$, $R^2 = .061$. Inspection of residuals suggested the model was not influenced by homoscedasticity. The non-linguistic variables explained around 6% of the variance of the math scores and indicated that students in lower grades who sent more messages to the Genie and received more messages from their teachers had low math proficiency. In addition, students who met a greater number of objectives, had higher math proficiency.

Table 1

Non-linguistic model for predicting math scores

Fixed Effect	**Coefficient**	**Std. Error**	***t***	***r***
(Intercept)	77.796	0.340	228.87**	
Grade Level	-0.423	0.086	-4.942**	0.001

Number of messages to Genie	-0.082	0.01	-9.964**	0.004
Number of messages from teacher	0.217	0.030	7.184**	0.006
Objectives met	0.145	0.005	27.607**	0.051

**p < .001

3.2 Linguistic Linear Model

A linear model including linguistic fixed effects revealed significant effects for a number of features related to proportion of n-grams used, phonographic and orthographic neighborhoods, number of unique function words, word frequency, incidence of determiners, strength of word associations, and term certainty. Table 2 displays the coefficients, standard error, *t* values, *p* values, and relative importance for each of the linguistic fixed effects. The overall model was significant, $F(11, 13590) = 112.70$, $p < .001$, $r = .290$, $R^2 = .084$. Inspection of residuals suggested the model was not influenced by homoscedasticity. The linguistic variables explained around 8% of the variance of the math scores and indicated that students that used more common n-grams with stronger associations and used words with fewer orthographic and phonological neighbors that were more frequent had higher math proficiency. In addition, students that used more unique function words, more determiners, and were more certain also had higher math proficiency scores. An ANOVA comparison between the non-linguistic model and the linguistic found a significant difference between the models, ($F = 47.867$, $p < .001$), indicating that linguistic features contributed to a better model fit than non-linguistic features.

Table 2

Linguistic model for predicting math scores

Fixed Effect	**Coefficient**	**Std. Error**	***t***	***r***
(Intercept)	88.656	3.710	23.894**	
Bigram proportion (COCA news)	3.447	1.261	2.734*	0.01
Phonological Neighbors	-0.244	0.039	-6.255**	0.014
Average Levenshtein Distance of closest orthographic neighbors	13.437	1.221	11.005**	0.013
Trigram proportion (BNC spoken)	5.533	1.913	2.892*	0.006
Unique function words	0.07	0.012	6.066**	0.007
Content word frequency (BNC written)	2.062	0.409	5.044**	0.006
Average frequency of closest orthographic neighbors	-3.549	0.428	-8.29**	0.009

Incidence of determiners	14.674	3.75	3.913**	0.004
Trigram association strength (COCA spoken MI)	0.381	0.072	5.293**	0.005
Certainty words	4.312	1.431	3.013*	0.003
Bigram association strength (COCA spoken T)	0.022	0.004	5.803**	0.005

* p < .010, **p < .001

3.3 Full Linear Model

A linear model considering non-linguistic and linguistic fixed effects revealed significant effects for all the non-linguistic features in the first model and all the linguistic features in the second model. Table 3 displays the coefficients, standard error, *t* values, *p* values, and relative importance for each of the fixed effects. The overall model was significant, $F(15, 13586) = 154.60$, $p < .001$, $r = .381$, $R^2 = .145$. Inspection of residuals suggested the model was not influenced by homoscedasticity. The non-linguistic and linguistic variables explained around 15% of the variance of the math scores and followed the same trends as reported in the first two models. An ANOVA comparison between the full model and the linguistic model found a significant difference between the models, ($F = 122.65$, $p < .001$), indicating that a combination of non-linguistic and linguistic features contributed to a better model fit than linguistic features alone.

Table 3

Full model for predicting math scores

Fixed Effect	**Coefficient**	**Std. Error**	***t***	***r***
(Intercept)	88.436	3.704	23.877**	
Bigram proportion (COCA news)	2.387	1.261	1.894	0.010
Phonological Neighbors	-0.237	0.038	-6.293**	0.014
Average Levenshtein Distance of closest orthographic neighbors	13.680	1.185	11.543**	0.013
Trigram proportion (BNC spoken)	4.539	1.851	2.453*	0.006
Unique function words	0.126	0.013	9.394**	0.008
Content word frequency (BNC written)	2.357	0.396	5.950**	0.006
Average frequency of closest orthographic neighbors	-3.401	0.414	-8.207**	0.009
Incidence of determiners	13.300	3.633	3.661**	0.004

Trigram association strength (COCA spoken MI)	0.393	0.070	5.647**	0.005
Certainty words	5.317	1.397	3.805**	0.004
Bigram association strength (COCA spoken T)	0.021	0.004	5.811**	0.005
Grade Level	-1.277	0.086	-14.823**	0.007
Number of messages from Genie	-0.103	0.009	-10.909**	0.005
Number of messages from teachers	0.182	0.029	6.316**	0.004
Objectives met	0.133	0.005	26.470**	0.046

* $p < .010$, **$p < .001$

4 Discussion

Cognitive skills are important indicators of math success. Most previous studies have examined indicators related to spatial attention and quantitative ability. In this study, we take an innovative approach and examine links between math success and language production. This approach is unlike many previous studies examining math success and language skills in that we did not examine standardized assessment of language ability or differences in math success between native and non-native speakers of English. In addition to language production, we also co-varied a number of non-linguistic factors to better assess the relationships between language production and math success. We found that a linguistic features model was a stronger predictor of math success than a model based on non-linguistic features. A blended model containing both linguistic and non-linguistic features outperformed both models.

The linguistic linear model showed that about 8% of the variation in math scores was accounted for by linguistic features. The results indicated that students who used more common n-grams (words and phrases), phrased in which the words were more strongly associated, more difficult words (e.g., words with fewer neighbors), a greater number of unique function words, more determiners, and more certainty words received higher math scores. These findings indicate that students with more conventional language production (i.e., students that follow standardized language patterns) scored higher in math assessments. In addition, students that expressed more certainty scored higher. These findings likely indicate that students that have acquired greater standardized patterns of language (i.e., academic language) perform better in math because those types of language skills allow for greater transfer of cognitive operations between language and academic domains such as math.

These findings differ from those reported by [9] in that greater cohesion and linguistic complexity were strong predictors of math success their study. In addition, the previous analysis indicated that linguistic features predicted about 30% of the variance in math scores. These differences likely stem from context in that student language production in [9] was recorded in the context of mathematics problem solving, while in the current study the context of language production was personal communication with an emphatic avatar. A qualitative analysis of the Genie data indicated that much of this communication was non-mathematical in nature and informal. As verbalized reasoning in the process of problem solving is a closer reflection of the mental processes of mathematical problem solving than informal communication, it's not surprising that the linguistic features of that communication were more strongly associated with language complexity and explained a greater percentage of mathematics performance.

The covariates used in the non-linguistic linear model, with the exception of Objectives Met, accounted for very little of the variation in math scores, despite the statistical significance of their coefficients. However, they did indicate that students in higher grades scored better and that students who wrote fewer messages to the Genie and received more messages from their teacher scored better. The negative co-efficient reported between math success and messages to Genie may relate to the number of off topic messages sent to the Genie and the informal nature of those messages. In terms of objectives met, the strength of the relation to math scores is not surprising: students who have higher accuracy in answering problems are generally able to complete objectives faster, even though, due to the adaptive logic of *Foundations*, they are presented with harder problems [16].

5 Conclusions and Future Work

In conclusion, we present additional evidence that linguistic features in language production can predict math success such that students with more standardized language perform better on math assessments. Future iterations of this work will use a wider range of outcomes, both more proximal to language production and semantic content -- like student self-efficacy, topic knowledge interest in mathematics, and mastery orientation – and more distal – like scores on summative assessment of mathematics, scores on standardized state tests, and academic outcomes (graduation, placement in advanced courses, etc.). In addition, future studies will use additional covariates related to language exposure in the models, like geographic location, ELL status, ethnicity, socio-economic status, to investigate how those factors interact with language production features to account for variation in attitudinal and achievement outcomes. Such studies will strengthen the analyses presented here and extend our knowledge of how language skills can transfer cognitive operations between math and language domains.

Acknowledgements

This research was supported in part by NSF 1623730. Opinions, conclusions, or recommendations do not necessarily reflect the views of the NSF.

References

[1] LeFevre JA, Fast L, Skwarchuk SL, Smith-Chant BL, Bisanz J, Kamawar D, Penner-Wilger M (2010) Pathways to math: Longitudinal predictors of performance. Child Development 81(6): 1753–1767. doi:10.1111/j.1467-8624.2010.01508.x

[2] Vukovic, RK, Lesaux NK (2013) The relationship between linguistic skills and arithmetic knowledge. Learning and Individual Differences *23*: 87-91. Doi: 10.1016/j.lindif.2012.10.007

[3] Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, *78*(2), 333-368.

[4] Adams TL (2003) Reading math: More than words can say. The Reading Teacher 56(8): 786–795.

[5] MacGregor M, Price E (1999) An exploration of aspects of language proficiency and algebra learning. Journal for Research in Math Education 449–467. doi: 10.2307/749709

[6] Alt M, Arizmendi GD, Beal CR (2014) The relationship between math and language: Academic implications for children with specific language impairment and English language learners. Language, Speech, and Hearing Services in Schools 45(3): 220–233. doi:10.1044/2014_LSHSS-13-0003

[7] Hampden-Thompson G, Mulligan G, Kinukawa A, Halle T (2008) Math Achievement of Language-Minority Students During the Elementary Years. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

[8] Martiniello M (2009) Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. Educational Assessment 14(3–4): 160–179. doi:10.1080/10627190903422906

[9] Crossley, SA, Liu R, McNamara D (2017) Predicting Math Performance Using Natural Language Processing Tools. Proceedings of the 7th International Learning Analytics and Knowledge (LAK) Conference. New York, NY: ACM .

[10] Hernandez F (2013) The Relationship Between Reading and Math Achievement of Middle School Students as Measured by the Texas Assessment of Knowledge and Skills (Doctoral dissertation).

[11] Cummins, J (1979) Linguistic interdependence and the educational development of bilingual children. Review of Educational Research 49(2): 222-251.

[12] Ardasheva Y, Tretter TR, Kinny M (2012) English language learners and academic achievement: Revisiting the threshold hypothesis. Language Learning 62(3): 769–812. 10.1111/j.1467-9922.2011.00652.x

[13] Mosqueda E, Maldonado SI (2013) The effects of English language proficiency and curricular pathways: Latina/os' math achievement in secondary schools. Equity & Excellence in Education 46(2): 202–219. 10.1080/10665684.2013.780647

[14] Wang J, Goldschmidt P (1999) Opportunity to learn, language proficiency, and immigrant status effects on math achievement. The Journal of Educational Research 93(2): 101–111.doi: 10.1080/00220679909597634

[15] Chen, F., & Chalhoub-Deville, M. (2015). Differential and long-term language impact on math. *Language Testing*, *33*(4) 577–605.

[16] Khachatryan, GA, Romashov, AV, Khachatryan, AR, Gaudino, SJ, Khachatryan, JM, Guarian, KR, & Yufa, NV (2014) Reasoning Mind Genie 2: An intelligent tutoring system as a vehicle for international transfer of instructional methods in

mathematics. International Journal of Artificial Intelligence in Education, 24(3): 333-382. doi: 10.1007/s40593-014-0019-7
[17] Kyle K, Crossley SA (2015) Automatically assessing lexical sophistication: Indices, tools, findings, and application. TESOL Quarterly 49(4): 757–786. doi:10.1002/tesq.194
[18] Crossley SA , Kyle K , McNamara DS (in press). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. Behavior Research Methods.
[19] Crossley SA, Kyle K , McNamara DS (in press). Sentiment Analysis and Social Cognition Engine (SEANCE): An Automatic Tool for Sentiment, Social Cognition, and Social Order Analysis. Behavior Research Methods.
[20] Thorndike EL, Lorge I (1944) The teacher's wordbook of 30,000 words. New York: Columbia University, Teachers College: Bureau of Publications.
[21] Kučera H , Francis N (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.
[22] Brown GD (1984) A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. Behavior Research Methods, Instruments, & Computers 16(6): 502-532. doi:10.3758/BF03200836
[23] Brysbaert M, New B (2009) Subtlexus: American word frequencies. *http:/subtlexus. lexique. org*.
[24] The British National Corpus, version 3 (BNC XML Edition) 2007 Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: http://www.natcorp.ox.ac.uk/
[25] Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*, 159-190. doi:10.1075/ijcl.14.2.02dav
[26] Coltheart M (1981) The MRC psycholinguistic database. The Quarterly Journal of Experimental Psychology 33(4): 497–505.
[27] Kuperman V, Stadthagen-Gonzalez H, Brysbaert M (2012). Age-of-acquisition ratings for 30,000 English words. Behavior Research Methods 44(4): 978–990. doi: 10.3758/s13428-012-0210-4
[28] Brysbaert M , Warriner AB , Kuperman V (2014) Concreteness ratings for 40 thousand generally known English word lemmas. Behavior Research Methods 46(3): 904–911. doi:10.3758/s13428-013-0403-5
[29] Miller GA (1995). WordNet: A lexical database for English. Communications of the ACM 38(11): 39–41. doi 10.1145/219717.219748
[30] Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. *The computer and literary studies*, 153–165.
[31] Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Available online at: http://www.usf.edu/FreeAssociation/.
[32] Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.
[33] Bird S, Klein E, Loper E (2009) Natural language processing with Python. O'Reilly Media, Inc.
[34] R Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013: ISBN 3-900051-07-0.
[35] Grömping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software,* 1 (1).