Yang, J. C. et al. (Eds.) (2018). Proceedings of the 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education

Identifying Changes in Math Identity Through Adaptive Learning Systems Use

Stefan SLATER[a*], Jaclyn OCUMPAUGH [a], Ryan BAKER [a], Junyi LI[b] & Matthew LABRUM[c]
[a]*Penn Center for Learning Analytics, University of Pennsylvania, United States*
[b]*Central China Normal University, China*
[c]*Imagine Learning, United States*
*slater.research@gmail.com

Abstract: Considerable evidence demonstrates that motivational constructs predict educational outcomes, but little research has examined how these constructs manifest within online learning systems. This study addresses this gap by surveying Math Identity measures (self-concept, value, and interest in mathematics) and correlating them to behavior and performance within Reasoning Mind's *Foundations* system for elementary mathematics.

Keywords: math identity, blended learning, self-concept, interest, value

1. Introduction

Studies of online learning environments primarily focus on short-term learning measures, performance on standardized tests, and other normative learning outcomes. Research shows that intelligent tutoring systems (ITS) produce learning improvements roughly equivalent to one-on-one tutoring, while functioning at a much larger scale (Kulik & Fletcher, 2016; Ma et al., 2014). Comparatively less research examines how student engagement predicts long-term development of interest (e.g. Ocumpaugh et al., 2016), value, or self-concept.

Early work suggests that positive self-concept is associated with goal setting (Bong & Skaalvik, 2003). Domain-specific self-perceptions, e.g., math self-concept, are known to predict achievement above and beyond measures of ability (Spinath et al., 2006). Self-efficacy is theorized to influence everything from self-regulatory behaviors in Expectancy-Value Theory (Eccles et al., 1983; Wigfield & Eccles, 2000), to interest in Social Cognitive Career Theory (Lent et al., 2002), the latter of which encompases both *intrinsic interest* (e.g., Gottfried, 1985) and *utility value* (e.g., Eccles et al., 1983).

Childhood development of math self-concept, interest, and value is not thoroughly understood. Self-concept measures are often only weakly correlated to performance (Gottfried, 1985; 1990, Steinmayr & Spinath, 2009). However, both self-concept and interest have reciprocal relationships with achievement (Guay et al., 2003; Marsh & Yeung, 1997), which suggests that even though they are separate constructs, they are each necessary components to understanding which students are most likely to show sustained engagement. The detailed documentation of ITS offer opportunities to explore how different experiences encourage students to value domain learning and to incorporate domain-specific success into their self-concept (e.g., math identity) in ways that lead to long-term interest. Yet developmental research shows that these scores are more often start high and decline rather than grow with experience (Frenzel et al., 2010).

This study examines the early development of math identity among elementary students working with Reasoning Mind, an ITS for mathematics. We operationalize this concept by adapting three survey scales (*self-concept*, *interest*, and *utility value*) used by Ryan & Ryan (2005) to explore how social stereotypes affect math idenity. We survey students at the start and end of the school year to examine how differences in these scales relate to student behaviors within the learning software.

2. Reasoning Mind

Reasoning Mind (RM) is an ITS for mathematics that is used by over 100,000 pre-K to 8^{th} grade students in the U.S. Prior work shows that it is associated with higher test scores (Boriack et al., 2015), interest (Waxman & Houston, 2012), and student engagement (Ocumpaugh, 2015). RM activities are organized within the context of *RM City*, a virtual town where students navigate from building to building using multiple modes: *Guided Study* (the main learning mode), *My Place* (where students decorate their virtual room) and *Game Room* (students participate in timed speed games, or solve math puzzles, like those found in the *Riddle Machine*). Content is classified by function and difficulty (Khachatryan et al., 2014). *Theory* problems teach mathematical concepts with animations and exercises. *Notes Test* problems require students to review Theory concepts while reinforcing good note-taking practices. *A-level* problems test basic material, while *B-level* problems may require multiple skills or steps. *C-level* problems are conceptually advanced, requiring higher-order thinking. While using RM students interact with virtual characters, including a pedagogical agent known as the Genie. Interactions are largely automated, but students can send email messages to the Genie (through RM software), which is answered in character by RM employees in a Wizard-of-Oz manner.

3. Methods

3.1 Students and Surveys

This study surveys 394 Texas students who used Reasoning Mind during their regular 2^{nd}-5^{th}-grade mathematics instruction in the 2016-2017 school year. Surveys questions were adapted from Ryan and Ryan's (2005) study on identity in mathematics, and they were administered using a 4-point Likert-style scale. Three different Math Identity scales were used, including *mathematics self-concept* (5 items capturing the degree to which students see themselves as a "math person," e.g., "I have always been good at math"), *interest in mathematics* (3 items capturing intrinsic curiosity or enjoyment of mathematics, e.g., "How much do you like math?"), and *value of mathematics* (5 items capturing the degree to which students find math useful, e.g., "How important is it to you to get good grades in math class?"). The Cronbach's α of each scale was 0.72, 0.69, and 0.72, respectively.

3.2 Feature Engineering

A total of 185 features were extracted from over 3.5 million interaction log events produced across the school year, and measures related to these actions were aggregated into monthly and yearly values. These include basic measures of performance, e.g. the percentage of correctly worked problems in *Guided Study*, the *Office*, and the *Notes Test*, across *A-level, B-level,* and *C-level* problems as well as activities in the *Game Room* and the *Riddle Machine*. We also assessed students' rates of contextual guess and slip (Baker et al., 2008). Other features included measures of how students spent the points that they earned in the system (for books, videos, and decorations) and features designed to capture goal-setting or challenge-seeking behaviors such as voluntarily working on difficult C-level problems.

Performance and temporal measures were used to contextualize some features, such as students' hint use (hints during poor vs. strong performance or followed by short or long pauses), the average time between an incorrect answer and a student's next response, and the proportion of incorrect answers in a moving six-second window. Some features were generated only at the yearly level , while other descriptive and summary features were generated at a monthly level and allowed us to examine yearly trends (slope, skewness, and kurtosis).

3.3 Analyses

Features were correlated against three identity scales: *value, self-concept,* and *interest* in mathematics. We examined both pre-year and post-year scores, as well as change from pre-year to post-year. The nonparametric Spearman Rho correlation coefficient was used, as normality assumptions were not universally met for our measures, and Benjamini and Hochberg's (1995) false discovery rate (FDR) control was applied due to the large number of tests (168 features x 9 outcomes). Only correlations that meet this adjusted significance criteria are discussed.

4. Results

4.1 Pre-Year and Post-Year Surveys

Paired-samples t-tests found no significant score differences between pre-year and post-year measures for *value*, t(393) = -0.284, p = 0.777, *self-concept*, t(393) = -0.620, p = 0.536, or *interest*, t(393) = 1.728, p = 0.085, though there was some evidence of a positive trend for interest. Ceiling effects may partially explain these findings: 81% of students had pre-year scores within a standard deviation of the maximum score for *value*, 40% for *self-concept*, and 58% for *interest*. As Table 1 shows, students above the median pre-year score ('high') for each subscale show slight decreases in their average post-year scores (-1.1 *self-concept*, -1.0 *value*, -1.2 *interest*), while those at are or below the median ('low') increase slightly (+1.0 *self-concept*, +1.1 *value*, and +0.6 *interest*).

Table 1

Averages for low and high-scoring students, as divided by median scores for each outcome measure.

	PRE-YEAR						POST-YEAR						CHANGE		
	Low		High		All		Low		High		All				
	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Low	High	All
S.C.	12.3	2.16	17.9	1.44	14.7	3.36	13.3	3.31	16.8	1.98	14.8	3.46	1	-1	0.1
Val.	17.7	2.33	20	0	18.2	2.22	17.8	3.16	19	2.69	18.2	2.19	1.1	-0.7	0
Int.	10.5	2.57	15.1	0.81	12.6	3.05	11	3.31	14	1.98	12.4	3.14	0.6	-0.9	-0.2

4.2 Correlations with Self-concept

Table 2 summarizes the features associated with pre-year and post-year self-concept scores. Of the 19 features that correlate to one or more of the outcome measures in this study, 18 correlated with the post-year scores of self-concept, 7 correlate to its pre-year scores, and none correlate to change.

Table 2

Correlations between student interaction features and self-concept; asterisks mark sig. p-values (after FDR corrections). Sig. negative correlations are in light gray; positive correlations are in dark gray.

-	Pre-year			Post-year			Change	
	rho	p		rho	p		rho	p
A-level Problems: %Correct (Avg)	**0.183**	**0**	*	**0.263**	**0.00**	*	0.105	0.04
A-level Problems: %Correct (SD)	-0.121	0.02		**-0.202**	**0.00**	*	-0.102	0.04
B-level Problems: %Correct (Avg)	**0.162**	**0.00**	*	**0.202**	**0.00**	*	0.03	0.56
B-level Problems: %Correct (SD)	**-0.2**	**0.00**	*	**-0.155**	**0.00**	*	0.055	0.29
C-level WoM: %Correct (Avg)	0.141	0.00		**0.196**	**0.00**	*	0.062	0.22
C-level WoM: Prop. Of Actions (Avg)	0.122	0.02		**0.168**	**0.00**	*	0.041	0.41
C-Level WoM: Prop. Of Actions (SD)	0.118	0.02		**0.16**	**0.00**	*	0.039	0.44
C-level WoM: Prop. Of Time (Avg)	0.124	0.01		**0.172**	**0.00**	*	0.061	0.22
C-Level WoM: Prop. Of Time (SD)	0.122	0.02		**0.168**	**0.00**	*	0.043	0.4
Known Skills: Cont. Guess (Avg)	**-0.169**	**0.00**	*	**-0.212**	**0.00**	*	-0.068	0.18

Known Skills: Cont. Slip (Avg)	**-0.168**	**0.00**	*	**-0.219**	**0.00**	*	0.074	0.14
Speed Game: %Correct (Avg)	**0.155**	**0.00**	*	**0.213**	**0.00**	*	-0.063	0.21
Speed Game: %Correct (SD)	-0.065	0.20		**-0.16**	**0.00**	*	-0.092	0.07
Speed Game: Completion Time (Avg)	-0.093	0.07		**-0.159**	**0.00**	*	-0.081	0.11
Theory Problems: %Correct (Avg)	0.142	0.00		**0.209**	**0.00**	*	0.087	0.08
Theory Problems: %Correct (SD)	-0.127	0.01		**-0.166**	**0.00**	*	0.064	0.2
Wrong after 6sec: Prop. Of Actions (Avg)	**-0.163**	**0**	*	**-0.245**	**0.00**	*	-0.117	0.02

One pattern that emerges in this data is the relationship between features based on the percentage of correctly worked problems and the standard deviation (SD) of the same value. For A-level, B-level, Speed Games and Theory Problems, the percentage of correct problems (Avg) is positively correlated with post-year self-concept scores, but the SD is negatively correlated. This suggests that while average performance is associated with self-concept, students who show less consistency in their performance are less likely to have a strong self-concept. Trends for this pattern are also seen in the pre-year values for self-concept, which is positively correlated with percent correct for three of these problems types (A-level, B-level, and Speed Game Problems) and negatively correlated with one of the SD measures (for B-Level Problems).

Another important set of patterns that emerge is self-concept's relationship to performance on C-level Problems (the most advanced problem type) and known skills. Six features associated with C-level Problems are positively correlated with post-year self-concept scores. Generally speaking, learners who chose to spend time and actions on C-level Problems and performed well on them reported higher post-year self-concept. Conversely, features related to known skills (Guess and Slip) are both negatively correlated to post-year scores of self-concept.

Finally, we see relationships between self-concept scores and speed measures. The feature, "Wrong After 6 Sec" is negatively correlated with self-concept scores, as is the Average Completion Time (slower pace) in the Speed Game. These trends suggest that students rapidly guessing on normal problems,or not answering questions quickly in speed games are also lower in math self-concept.

4.3 Correlations with Value

Only 5 features are significantly correlated to *value* scores: 4 to post-year scores and 2 to pre-year scores (see Table 3). Average performance (percent correct) is positively correlated to *value*. For post-year results, only 2 features are significant, and for pre-year results, only performance on A-Level problems is significant. Likewise, the SD of average performance is negatively (but not significantly) correlated to both pre-year and post-year *value*. Features related to C-level Problems are not significantly correlated to *value*, but those related to performance on Known Skills (Slip and Guess) are negatively correlated.

Results for features related to speed also mimic the trends for *self-concept*. Speed Game Completion Time is negatively correlated to *value* scores, as are frequent incorrect attempts during regular activities (*Wrong After 6 Sec*). Though the former is not significant, it is notable in that it follows the trend seen for *self-concept*; the latter is significant, but only for post-year scores.

Table 3

Correlations between student interaction features and value; asterisks mark sig. p-values (after FDR corrections). Sig. negative correlations are in light gray; positive correlations are in dark gray.

-	Pre-year			Post-year			Change	
	rho	p		rho	p		rho	P
A-level Problems: %Correct (Avg)	**0.171**	**0.00**	*	**0.233**	**0.00**	*	0.096	0.06
B-level Problems: %Correct (Avg)	0.126	0.01		**0.168**	**0.00**	*	0.076	0.14
Known Skills: Cont. Guess (Avg)	-0.163	0.00		**-0.197**	**0.00**	*	-0.061	0.22
Known Skills: Cont. Slip (Avg)	**-0.173**	**0.00**	*	**-0.203**	**0.00**	*	-0.041	0.42

Wrong after 6sec: Prop. Of Actions (Avg)	-0.143	0.00		**-0.218**	**0.00**	*	-0.087	0.09

4.4 Correlations with Interest

Like *value*, *interest* scores correlate to a total of five features. As shown in Table 4, only one feature is associated with pre-year scores for *interest*, and the general pattern of correlation coefficients mirrors self-concept and value. Average performance (percent correct) is positively correlated with higher scores, and this finding is significant for A and B-level problems. The SD of B-level problem performance is negatively correlated with both pre- and post-year scores of *interest*, though not with normalized gains. Wrong after 6 seconds is again significant for post-year scores.

Though not significant for self-concept or value, students who became faster at solving speed game problems over the course of a year show higher post-year *interest*. This suggests that students interested in mathematics also become more skilled at fast mental-math style operations.

Table 4

Correlations between student interaction features and interest; asterisks mark sig. p-values (after FDR corrections). Sig. negative correlations are in light gray; positive correlations are in dark gray.

-	**Pre-year**			**Post-year**			**Change**	
	rho	**p**		**rho**	**P**		**rho**	**P**
A-level Problems: %Correct (Avg)	0.074	0.14		**0.181**	**0.00**	*	0.101	0.05
B-level Problems: %Correct (Avg)	0.134	0.01		**0.235**	**0.00**	*	0.119	0.02
B-level Problems: %Correct (SD)	**-0.197**	**0.00**	*	**-0.235**	**0.00**	*	-0.033	0.53
Speed Game: Completion Time (Slope)	-0.043	0.39		**-0.207**	**0.00**	*	-0.176	0.00
Wrong after 6sec: Prop. Of Actions (Avg)	-0.09	0.08		**-0.183**	**0.00**	*	-0.093	0.07

5. Conclusions and Future Research

This study examines how Reasoning Mind student interactions correlate with math identity, which is operationalized as *math self-concept*, *value*, and *interest.* Notably, there are not large decreases in these scores, despite these children being at an age when research suggests that their self-concept should drop. The features examined include both those which were derived directly from performance (e.g. Percent Correct) and those more likely to demonstrate student choices (e.g., proportion of time on C-level problems). This variation in feature design was deliberate, as the literature has been careful to define self-concept as distinct from simple performance measures.

Two trends emerge from our data that provide further opportunities to study math identity. First, while average performance on A, B, and C-level problems is positively correlated with all three survey scales, standard deviations for these measures are negatively correlated. If students perform better on these problems, they also score higher for *self-concept*, *value*, and *interest.* However, relatively minor changes in performance (as reflected by higher standard deviations) may affect students' identity despite high overall performance. This fits with the literature on the developmental patterns associated with this age group, even though most of the students in this sample are not experiencing declines. Second, low post-year *self-concept* and *value* scores are associated with more guess and slip (Baker et al., 2008); the higher the probability that a student's actions are random guesses or careless errors, the lower their post-year scores for *self-concept* and *value.*

There are multiple avenues for future research, but analyses examining potential thresholds for improved math identity (e.g., a specific slope of performance over the course of the year that correlates with improvements) seem particularly important. Identifying the point at which students' self-concept typically starts to increase could help us to better understand identity development and to support students who might otherwise not notice their improvements because they are continuing to receive challenging material rather than being allowed to plateau. In doing so, we have tried to address an area of research that has not been thoroughly explored by ITS researcers..

Acknowledgements

Our thanks to the NSF (Cyberlearning Award #1623730) for their funding and to Penn's Graduate School of Education Research Apprenticeship Course for feedback on earlier drafts of this research.

References

Baker, R.S.J.d., Corbett, A., Aleven. V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian Knowledge Tracing. *Proc. 9th Intern'l Conf. on ITS*, pp. 406-15.

Benjamini, Y., Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289-300.

Bong, M., Skaalvik, E. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psych Review. 15*(1), pp. 1-40.

Boriack, A., Wright, K., Stillisano, J., Waxman, H. (2015). Independent Evaluation of Reasoning Mind Programming. College Station, TX: Education Research Center, Texas A&M University.

Eccles J.S., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J., Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (Ed.), *Achievement and achievement motivation* (pp. 75–146). San Francisco, CA: W. H. Freeman.

Frenzel, A. C., Goetz, T., Pekrun, R., Watt, H. M. (2010). Development of mathematics interest in adolescence: Influences of gender, family, and school context. *Journal of Research on Adolescence*, *20*(2), 507-537.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior high school students. *Journal of Educational Psychology, 77*(6), pp. 631-645.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology, 82*(3), pp. 525.

Guay, F., Marsh, H., Boivin, M. (2003). Academic self-concept and academic achievement: developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*(1), pp. 124.

Khachatryan, G., Romashov, A.V., Khachatryan, A.R., Gaudino, S.J., Khachatryan, J.M., Guarian, K. Yufa, N. V. (2014). Reasoning Mind Genie 2: An Intelligent Tutoring System as a Vehicle for International Transfer of Instructional Methods in Mathematics. *Intern'l J. of Artificial Intelligence in Ed. 24*, 333-382.

Kulik, J., Fletcher, J. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of Educational Research*, *86*(1), pp. 42-78.

Lent, R., Brown, S. Hackett, G. (2002). Social cognitive career theory. In *Career Choice & Development*. John Wiley & Sons, San Francisco, pp. 255-311.

Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, *106*(4), pp. 901.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology, 89(1)*, pp. 41.

Ocumpaugh, J. (2015). Effects of Reasoning Mind on Student Engagement in DallasISD Mathematics Classrooms. NY: NY. Teachers College, Columbia U.

Ocumpaugh, J., San Pedro, O., Lai, H., Baker, S. (2016). Middle school engagement with mathematics software and later interest and self-efficacy for STEM careers. *J. of Sci. Ed. & Tech.*, 25(6), pp. 877-887.

Ryan, K., Ryan, A. (2005). Psychological processes underlying stereotype threat and standardized math test performance. *Ed. Psychologist*, *40*(1), pp. 53-63.

Spinath, B., Spinath, F., Harlaar, N., Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, *34*(4), pp. 363-374.

Steinmayr, R., Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning & Individual Differences*, *19*(1), pp. 80-90.

Waxman, H., Houston, W. (2012). Evaluation of the 2010-11 Reasoning Mind program in Beaumont ISD. College Station, TX: Texas A&M University.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, *25*(1), 68-81.