Unveiling the dynamics of learning behaviors in learning K-12 math: An exploration of an ASSISTments Dataset

Jun Song Huang
National Institute of Education
Nanyang Technological University
Singapore
junsong.huang@nie.edu.sg

Arya Radhakrishnan
Independent Researcher
Singapore
aryaradhakrishnan88@gmail.com

Timothy Lee
National Institute of Education
Nanyang Technological University
Singapore
timothy.lee@nie.edu.sg

Min Lee
National Institute of Education
Nanyang Technological University
Singapore
NIE21.LM1646@e.ntu.edu.sg

Janice Lum
National Institute of Education
Nanyang Technological University
Singapore
janice.lum@nie.edu.sg

Guimei Liu
Institute for Infocomm Research
*A*STAR*
Singapore
liug@i2r.a-star.edu.sg

Jung Jae Kim
Institute for Infocomm Research
*A*STAR*
Singapore
jjkim@i2r.a-star.edu.sg

***Abstract*—This study delves into the dynamics between diverse learning behaviors among K-12 students and their learning gains using a dataset of 508 students learning three math skills in ASSISTments. Employing K-means clustering based on students' initial and final skill mastery alongside their engagement level, three distinct clusters emerged for each skill, revealing varying degrees of learning from ASSISTments. By analyzing decision tree classification models for each skill using affective labels such as boredom and frustration, we hypothesize that students within the same cluster of a skill may exhibit heterogeneous learning patterns that affect their subsequent learning of new skills. Further exploration demonstrates that students who transit between clusters when learning new skills differ significantly in their initial and final mastery of previously learned skills and their affective labels associated with those skills. Regression analysis underscores that students' initial and final mastery of antecedent skills have some influence on their subsequent mastery of new skills. Unraveling the intricate relationship between student learning behaviors and the effectiveness of ASSISTments offers valuable insights into tailoring AI-enhanced educational tools, not only for learning the current skill but also for preparing for the future learning of new skills**.

Keywords—Intelligent Tutoring Systems, K-12 Math Learning, Affect, Student Profiling

I. Introduction

Intelligent Tutoring System (ITS) is one of the most significant research streams in Artificial Intelligence for Education (AIED) that delivers personalized learning. While some meta-analyses [1, 2] have shown that ITSs support learning gains, the findings are not consistent. For example, Steenbergen-Hu and Cooper [3], who performed the only meta-analysis focusing on the effectiveness of ITS for K-12 math learning, found that ITSs had little effect.

Many empirical studies focused on evaluating the effectiveness of ITSs, either as off-the-shelf systems or with some research-based enhancements. However, few studies examined the dynamics between students' diverse learning behaviors and their varied learning gains to inform the understanding of learning in ITSs. A noteworthy study conducted by Muldner, et al. [4] examined students' gaming the system and correlated this learning behavior to explain students' gaps in learning gains. While such an exploration contributes valuable insights into student behaviors in ITS, it is vital to recognize the limitations of profiling students based on their behavior of gaming the system singularly. Thus, a more comprehensive investigation is necessary to uncover how diverse learning behaviors relate to learning gains.

This study delves into examining the dynamics of diverse learning behaviors among students and the differential learning gains, in terms of learning both the current and subsequent skills. Findings could contribute to the understanding of personalized learning in ITS and have implications for the design of ITS to adapt for the needs of different profiles of students. Furthermore, learning is not only cognitive, but also social and emotional [5]. Our study also aims to examine students' affective states when learning in ITS to better understand and optimize learning.

II. Method

A. Dataset

We used the ASSISTments Longitudinal Data Mining Competition 2017 Dataset [6], consisting of data collected from the ASSISTments ITS between 2004-2006. Key variables used in our analysis are Ln, indicating the system's knowledge estimate of a student at each time step, as well as confidence scores for the student affect prediction labels of boredom, concentration, confusion, frustration, off-task behavior, and gaming (the system) at each time step. A more comprehensive description of the method used by ASSISTments to develop the affect and engagement labels can be found in Pardos et al. [7].

This research / project is supported by the Ministry of Education, Singapore, under its Science of Learning Grant (Award ID MOE-MOESOL2021-0006). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00265

The ASSISTments dataset was chosen as it is open access and contains data on epistemic emotions [8] such as boredom and frustration, which allows us to examine students' emotional engagement in learning.

B. Data Preparation

We sought to understand students' learning for the skills in whole number division, fraction multiplication, and fraction division. Hence, we focused on a smaller sample of the ASSISTments dataset containing 508 students who had recorded learning events in all three skills. These three skills were chosen for this study as they are conceptually connected. For example, fraction division is considered analogous to whole number division [9], and students often regard division as opposite to multiplication [10]. Examining these skills helps us understand the extent to which learning of antecedent skills helps students learn the later skills [11].

Data were aggregated for each student within each skill, yielding the following variables for subsequent analysis:

- Learning_Events – the number of learning events recorded for a given student within each skill.
- Initial_Ln – the mean of Ln for the first 3 learning events for a given student within each skill. Using a mean of 3 learning events provides a more stable estimate of students' knowledge state that is not overly influenced by students' performance in a single learning event.
- Final_Ln – the mean of Ln for the last 3 learning events for a given student within each skill.
- Confidence (BORED), confidence (CONCENTRATING), confidence (CONFUSED), confidence (FRUSTRATED), confidence (OFF TASK), confidence (GAMING) – the mean, for a given student within each skill, of the confidence scores for the student's affect prediction labels for boredom, concentration, confusion, frustration, off-task behavior, and gaming respectively.

C. Analysis Method

Data preparation yielded 3 datasets containing the variables above – one each for whole number division, fraction multiplication, and fraction division, based on the learning events of the 508 students learning these skills. The following analysis steps were conducted to understand students' learning of these skills. Findings of this analysis are reported in the next section.

- K-means clustering – datasets were independently scaled and clustered using Learning_Events, Initial_Ln, and Final_Ln. Using the elbow method, 3 clusters were identified as the optimal number for each dataset. For each cluster, Wilcoxon signed rank tests were used to investigate differences between Initial_Ln and Final_Ln. In a subsequent interpretation, cross-skill clusters were identified qualitatively based on similarities between clusters across datasets.
- Decision tree – Decision tree classifiers were trained on each dataset, predicting cluster labels using the confidence scores for students' affect prediction labels. To balance the predictability and interpretability of the machine learning model [12], the model selection pipeline included exhaustive feature selection (EFS) to select the best combination of up to 3 features for a decision tree classifier with a tree depth of 4, followed by hyperparameter tuning using the combination of features identified by EFS through a brute force grid search. EFS and hyperparameter tuning were conducted on a training set of 80% of the original data and evaluated using 5-fold cross validation. The best model with a tree depth of 3 was evaluated on a test set with the remaining original data to ensure robust model performance before being selected for interpretation. Single decision tree classifiers, EFS and a shallow tree depth were employed to improve the explainability of models for qualitative interpretation.
- Cross-tabulation of clusters – A cross-tabulation of clusters was conducted and used to characterize the learning of students across skills and between cross-skill clusters. Simple t-tests compared the confidence scores of students' affect prediction labels between students from the same cluster of antecedent skills but different clusters for later skills (whole number division is antecedent to fraction multiplication, which is antecedent to fraction division). These were conducted to understand whether and how student affect could predict students transiting between cross-skill clusters across skills.
- Linear regression – to understand predictive relationships between antecedent and subsequent skills suggested by the cross-tabulation analysis, Ordinary Least Squares (OLS) regression was conducted to predict the Initial_Ln and Final_Ln values for fraction division, using the same variables from the whole number division and fraction multiplication. To aid understanding of the relative importance of predictor features, a third model used Least Absolute Shrinkage and Selection Operator (LASSO) regularization to identify the more important predictors. For all models, 5-fold cross validation on a training set consisting of 80% of the original data was used to fit the models (and determine optimal hyperparameters for LASSO regularization). The best model was then used to predict the outcome variable of a test set consisting of the remaining 20% of the original data and obtain the R^2 and Mean Square Error (MSE).

III. FINDINGS

A. Learning Gains vs. Behavioral Engagement

In response to the inconsistent findings in the literature on the effectiveness of ITSs, we calculated the students' learning gains (i.e., the difference between Final_Ln and Initial_Ln) for each skill and plotted it against Learning_Events (as a proxy for behavioral engagement). See Fig.1.

Fig. 1 reveals that the learning gains across skills were tightly distributed around zero (mean = 0.04), with only a slight positive skew (skewness = 0.54). Specifically, for whole number division, mean = 0.03, skewness = -0.08; for fraction multiplication, mean = 0.01, skewness = 0.54; for fraction division, mean = 0.08, skewness = 1.60. Students had higher learning gains in fraction division, a more difficult skill.

Most students had learning gains between -0.25 to 0.25 and were well engaged with high Learning_Events. We noted that about half of the students had negative learning gains.

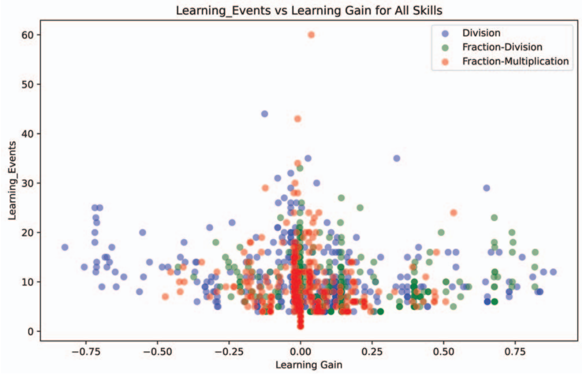

Fig. 1. Delta vs Learning_Events

B. *Clustering Analysis: Learner Profiles*

A K-means clustering using the variables of Initial_Ln, Final_Ln, and Learning_Events, reveals three optimal number of clusters for each skill. Table 1 presents the clustering and the Wilcoxon signed rank test comparing the Initial_Ln and Final_Ln for each cluster to determine the significance of the learning gains.

Although the clustering was conducted independently for each skill, the similarities of the clusters across the skills suggest the appropriateness in profiling students. Learning_Events are also found to be significantly different across the clusters for each skill.

Cluster 0 across the skills (named the 'high achiever' (HA) cluster) were similar with high Initial_Ln and Final_Ln, low Learning_Events and statistically significant learning gains. High achievers mostly answered questions in ITSs correctly. With personalized learning embodied by ITSs, they are generally given fewer questions to answer. Hence, there is no sub-category of highly or lowly engaged high achievers.

Cluster 1 across the skills (named the 'lowly engaged low achiever' (LELA) cluster) were low in Initial_Ln and Final_Ln. Their Learning_Events were relatively low but higher than those of Cluster HA, perhaps due to system behavior (i.e., ASSISTments is designed to give low achievers more questions to answer).

TABLE I. CLUSTERING OF STUDENTS BASED ON SKILLS

	Whole Number Division	**Fraction Multiplication**	**Fraction Division**
Cluster 0 **High Achievers (HA)**	**280 students** *Initial_Ln:* *0.81 (0.10)* *Final_Ln:* *0.86 (0.13)* *W=3328, p=0.000* *Learning_Events:* *5.46 (3.90)*	**131 students** *Initial_Ln:* *0.56 (0.11)* *Final_Ln:* *0.58 (0.16)* *W=199, p=0.035* *Learning_Events:* *3.11 (2.12)*	**140 students** *Initial_Ln:* *0.59 (0.13)* *Final_Ln:* *0.70 (0.14)* *W=59, p=0.000* *Learning_Events:* *3.37 (2.80)*
Cluster 1 **Lowly Engaged Low Achievers (LELA)**	**137 students** *Initial_Ln:* *0 .22 (.17)* *Final_Ln:* *0.33 (.28)* *W=1675, p=0.001* *Learning_Events:* *7.12 (3.28)*	**311 students** *Initial_Ln:* *0.13 (0.09)* *Final_Ln:* *0.14 (0.10)* *W=14,298,p=0.225* *Learning_Events:* *6.88 (2.93)*	**266 students** *Initial_Ln:* *0.19 (0.08)* *Final_Ln:* *0.23 (0.13)* *W=7690, p=0.110* *Learning_Events:* *5.96 (2.64)*
Cluster 2 **Highly Engaged Low Achievers (HELA)**	**91 students** *Initial_Ln:* *0.27 (0.29)* *Final_Ln:* *0.12 (0.16)* *W=1008, p=0.000* *Learning_Events:* *19.00 (5.85)*	**66 students** *Initial_Ln:* *0.10 (0.05)* *Final_Ln:* *0.11 (0.10)* *W=1088, p=0.911* *Learning_Events:* *18.88 (7.56)*	**102 students** *Initial_Ln:* *0 .19 (0.10)* *Final_Ln:* *0.35 (0.26)* *W=1493, p=0.000* *Learning_Events:* *15.24 (4.36)*

Cluster 2 across the skills (named the 'highly engaged low achiever' (HELA) cluster) were low in Initial_Ln and Final_Ln. They had the highest number of learning events. The mean of their learning gains for whole number division was negative.

To determine if students' affect states in learning a skill could predict their cluster labels (i.e., student profiles), we trained and tested decision tree classification models using the confidence scores for student affect to predict the cluster labels. Figures 2-4 present the decision trees for each skill.

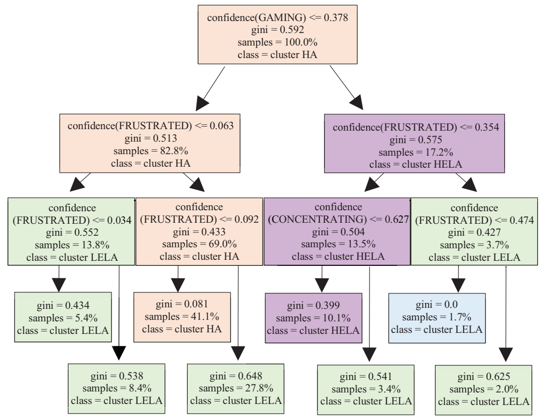

Fig. 2. Decision tree for Whole Number Division clustering.

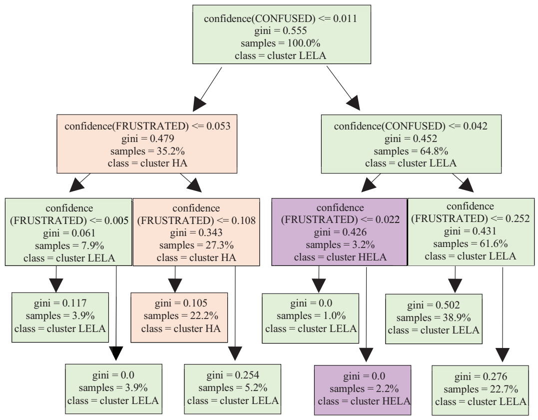

Fig. 3. Decision tree for Fraction Multiplication clustering.

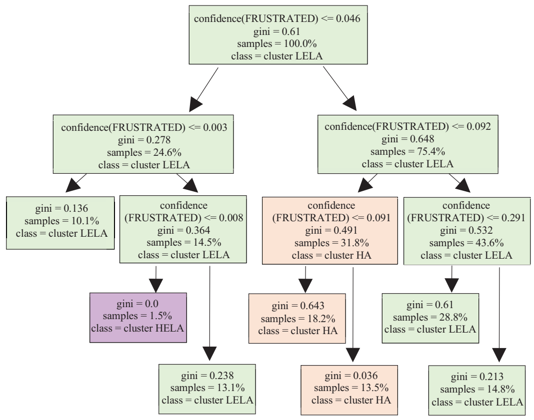

Fig. 4. Decision tree for Fraction Division clustering.

The high gini results suggest that the decision trees balancing predictability and interpretability are not able to use affect states (when learning a skill) to predict their cluster labels (i.e., student profiles) of the skill satisfactorily. It is particularly difficult to differentiate the two low achiever clusters (i.e., lowly- vs highly- engaged low achievers). This may imply some complexity in emotional engagement in learning. A decision tree with more depth and data features is needed to make better predictions but harder to interpret.

We note from the decision trees that the clusters across the skills were relatively sensitive to confidence (FRUSTRATED). It does not mean that students feel very frustrated when using ASSISTments. Rather, it could imply that students of different profiles have subtle differences in frustration, such as being frustrated for different reasons (e.g., work on already mastered questions or repeatedly fail to learn and answer correctly) or at different intensity levels.

We also found that the decision trees could not identify Cluster HELA well across the skills, suggesting that the composition of students in Cluster HELA could be complex. For example, Owen et al. [13] differentiate low achievers with high Learning_Events into unproductive (‘wheel-spinning’ with reduced motivation) versus productive (maintaining action despite failure) persistence. Due to the relatively small sample sizes of Cluster HELA across the skills, we did not examine this further.

C. *Cross-tabulation: Transition of Clusters Across Skills*

To examine the extent to which students transit to different clusters across the skills, cross-tabulations of the clusters between skills are presented in the tables below.

Cross-tabulations reveal that more than half of the students changed clusters when they learnt new skills (e.g., moving from the HA cluster in whole number division to the HELA cluster in fraction multiplication). This transition across the skills could imply two possibilities. First, the learning of previous skills has little influence on the learning of new skills. Second, students’ engagement in learning in ASSISTments does not accurately represent their actual learning and performance. Our subsequent examinations would speak to the possibilities.

TABLE II. CROSS-TABULATION OF WHOLE NUMBER DIVISION AND FRACTION MULTIPLICATION

Cross-tabulation Frequency		Whole Number Division		
		HA	**LELA**	**HELA**
Fraction Multiplication	**HA**	86	30	54
	LELA	171	86	15
	HELA	23	21	22

TABLE III. CROSS-TABULATION OF FRACTION MULTIPLICATION AND FRACTION DIVISION

Cross-tabulation Frequency		Fraction Multiplication		
		HA	**LELA**	**HELA**
Fraction Division	**HA**	61	67	12
	LELA	47	191	28
	HELA	23	53	26

TABLE IV. CROSS-TABULATION OF WHOLE NUMBER DIVISION AND FRACTION DIVISION

Cross-tabulation Frequency		Whole Number Division		
		HA	**LELA**	**HELA**
Fraction Division	**HA**	82	36	22
	LELA	148	78	40
	HELA	50	23	29

We further analyzed Cluster HA of antecedent skill on their transition to different clusters when they learnt descendent skills. It is to understand how their transition can be predicted by the variations in Initial_Ln, Final_Ln, Learning_Events and affect states when they learnt the antecedent skill.

Table 5 (next page) reveals that among the Cluster HA in whole number division, students are more likely to be in Cluster HELA in fraction multiplication if they show emotional disengagement (e.g., bored, confused, and frustrated).

In addition, we interpret Learning_Events as system behavior (i.e., high achievers answer questions in ITS more accurately hence receive less questions from ITS) rather than a predictor of learning gain (i.e., better learning because of behavioral engagement).

Table 6 (next page) shows that students’ emotional engagements (i.e., boredom, confusion, and concentration) when learning whole number division can be early indicators to differentiate student profiling when they learn fraction division.

It is interesting to note in Table 6 that high achievers in fraction multiplication were more likely to transit to the low achiever clusters (Clusters LELA and HELA) in fraction division when they had higher Initial_Ln, Final_Ln, and Learning_Events in fraction multiplication. It suggests that emotional engagements in learning whole number division are better early indicators than skill mastery in predicting the transfer.

In contrast, Table 7 (next page) shows that high achievers in fraction multiplication who have high skill mastery and low frustration are more likely to be higher achievers in fraction division. The difference between Tables 6 and 7 may be because the skill of fraction division is more linked to the skill of fraction multiplication (i.e., fraction division can be solved using an inverse-and-multiply procedure) than whole number division, hence there are transfer of learning from fraction multiplication to fraction division.

We also examined how Cluster LELA students in fraction multiplication transited to different clusters in fraction division. Analyzing this largest cluster across all skills helps us understand conditions in which low achievers in an antecedent skill transit to different clusters in a descendent skill. See Table 8.

TABLE V. DIFFERENCES IN WHOLE NUMBER DIVISION CLUSTER HA PREDICTS THE TRANSITION TO FRACTION MULTIPLICATION CLUSTERS

	Significant differences among Cluster HA Students in Whole Number Division
Cluster HA vs. Cluster LELA in Fraction Multiplication	Learning_Events (t = -2.73, p = 0.007) confidence(BORED) (t = -2.76, p = 0.006) confidence(CONFUSED) (t = -2.50, p = 0.013) confidence(FRUSTRATED) (t = -2.36, p = 0.019)
Cluster HA vs. Cluster HELA in Fraction Multiplication	No significant difference found
Cluster LELA vs. Cluster HELA in Fraction Multiplication	No significant difference found

TABLE VI. DIFFERENCES IN WHOLE NUMBER DIVISION CLUSTER HA PREDICTS THE TRANSITION TO FRACTION DIVISION CLUSTERS

	Significant differences among Cluster HA Students in Whole Number Division
Cluster HA vs. Cluster LELA in Fraction Division	Initial_Ln (t = -2.10, p = 0.037) Final_Ln (t = -2.13, p = 0.034) Learning_Events (t = -3.95, p = 0.000) confidence(BORED) (t = -2.65, p = 0.009) confidence(CONCENTRATING) (t = 2.21, p = 0.028) confidence(CONFUSED) (t = -2.18, p = 0.030)
Cluster HA vs. Cluster HELA in Fraction Division	Initial_Ln (t = -1.97, p = 0.05) Final_Ln (t = -3.92, p = 0.000) Learning_Events (t = -4.29, p = 0.000) confidence(CONFUSED) (t = -2.11, p = 0.037)
Cluster LELA vs. Cluster HELA in Fraction Division	No significant difference found

TABLE VII. DIFFERENCES IN FRACTION MULTIPLICATION CLUSTER HA PREDICTS THE TRANSITION TO FRACTION DIVISION CLUSTERS

	Significant differences among Cluster HA Students in Multiplication
Cluster HA vs. Cluster LELA in Fraction Division	Initial_Ln (t= 2.71 p = 0.008) Final_Ln (t= 2.88 p = 0.005) confidence(FRUSTRATED) (t= -2.60, p = 0.011)
Cluster HA vs. Cluster HELA in Fraction Division	Learning_Events (t= -2.28, p = 0.025)
Cluster LELA vs. Cluster HELA in Fraction Division	No significant difference found

TABLE VIII. DIFFERENCES IN FRACTION MULTIPLICATION CLUSTER LELA PREDICTS THE TRANSITION TO FRACTION DIVISION CLUSTERS

	Significant differences among Cluster LELA Students in Fraction Multiplication
Cluster HA vs. Cluster LELA in Fraction Division	Final_Ln (t = 2.33, p = 0.021) confidence(FRUSTRATED) (t = -2.45, p = 0.015)
Cluster HA vs. Cluster HELA in Fraction Division	confidence(FRUSTRATED) (t = -3.22, p = 0.002)
Cluster LELA vs. Cluster HELA in Fraction Division	Learning_Events (t = -3.49, p = 0.001)

Table 8 reveals that high emotional engagement (i.e., less frustrated) in learning fraction multiplication and good skill mastery (relative among low achievers) can be early indicators to predict lowly engaged low achievers in fraction multiplication who became high achievers in fraction division. The finding suggests ways to help low achievers to catch up, for example using more engaging pedagogy for emotional engagement and providing remedy to improve prior knowledge.

D. Linear Regression Modelling

Lastly, suggested by findings in Tables 6-8, we examined the extent to which the learning of whole number division and fraction multiplication could predict the learning of fraction division using linear regression modelling.

First, we used the Initial_Ln and Final_Ln for whole number division and the Initial_Ln and Final_Ln for fraction multiplication to predict the Initial_Ln of fraction division. The MSE is 0.048 and The R^2 score is 0.0008 which is close to zero, indicating that the model does not explain much of the variability in the Initial_Ln of fraction division based on the predictors chosen. This suggests limited transfer (i.e., students using skills learnt previously to help them solve novel questions of the skills not yet learnt).

Second, we used the Initial_Ln and Final_Ln for whole number division and the Initial_Ln and Final_Ln for fraction multiplication to predict the Final_Ln of fraction division. We did not include Initial_Ln of fraction division as an independent variable because our interested in the learning of antecedent skills predicting the learning of descendant skills.

The MSE is 0.0565 and the R^2 score is 0.1516. While the R^2 is still relatively low, it is much higher than the R^2 value for predicting Initial_Ln of fraction division.

We then performed feature selection using LASSO. The MSE is 0.0578 and R^2 is 0.1315. The coefficients of the predictors (after LASSO regularization) are:

- Initial_Ln of whole number division: 0.00
- Final_Ln of whole number division: 0.00
- Initial_Ln of fraction multiplication: 0.0244
- Final_Ln of fraction multiplication: 0.0545

The results from the LASSO model suggest that students' initial and final mastery of fraction multiplication are more important in predicting their final mastery of fraction division.

IV. Conclusion and Discussion

Putting the findings together, our analyses on the ASSISTments sub-dataset revealed that students generally had slightly more learning gains on fraction division than whole number division and fraction multiplication. The cluster analysis showed three clusters of student profiles, namely 'high achievers', 'highly engaged low achievers' and 'lowly engaged low archivers', across the three skills. In this paper, students with higher Learning_Events are regarded to be highly engaged, behaviorally. We also noted that students transited to different clusters when they learnt new skills, implying dynamics in learning and transfer.

Affective states showed potential to predict student profiling, particularly as early indicators on students transiting to different clusters in future learning of new skills. This finding is congruent with the work by Richey et al. [14] on the importance of epistemic emotions in learning.

We also found that learning and mastery of antecedent skills had some influence on the learning and mastery of subsequent skills.

Rising above from the findings, we highlight key inferences concerning students' learning of K-12 mathematics, their engagements in ASSISTments and the implications. Although our study in the context of ASSISTments dataset and for learning mathematics, the implications we discuss may inform the understanding of learning in broader contexts.

A. Learning and Engagement in ASSISTments

First, the present study found students to be engaged when learning in ASSISTments. Student engagement is a multifaceted construct encompassing various dimensions, namely, behavioral, cognitive, emotional, and agentic [15]. Compared to other ITS, such as Cognitive Tutor, which captures the process data of students problem-solving activities, ASSISTments only records students' activities through their responses, thereby providing limited insights into students' cognitive engagement. Furthermore, like other ITSs, ASSISTments curates questions for students to answer and thus restricts students' agentic engagement when using the platform. Nevertheless, the findings from this study revealed that students demonstrated behavioral and emotional engagement when learning in ASSISTments.

Second, emotional engagement (i.e., less boredom) in ASSISTments is a more reliable predicator of learner profiles and learning gains in ASSISTments. Although our study identified behavioral engagement among students in ASSISTments, it is vital to recognize that the completion of these Learning_Events is predominantly a consequence of system-driven behavior rather than reflective of students' agentic learning behaviors. As a result, these existing behavioural metrics may not be indicative of students' active participation in learning activities within the platform.

In contrast, emotional engagement when learning a skill has a potential to predict learner profiles, not only for the learning of the current skill (e.g., whole number division) but also the future learning of descendent skills (e.g., fraction multiplication and fraction division). The findings underscore the need to engage students emotionally in learning via ITS.

Recognizing the limitations of behavioral metrics (more as system behavior) in capturing student learning engagement, our findings supported the importance of delving into affective dimensions to gain a more holistic understanding of students' learning experiences within the ASSISTments environment.

Third, the findings revealed that students with different learner profiles had different levels of learning gains. To optimize their learning, we can consider the framework of knowledge-learning-instruction (KLI) dependency [16]. Koedinger et al. [16] purports that learning by a particular student from a course can be improved by first identifying the needed concepts and skills and second, selecting the instructional methods that best support the kind of learning needed for these concepts and skills by the profile of student. ASSISTments and other ITSs could further consider how to adapt pedagogical supports to help students of different learner profiles to optimize their learning in ITSs.

B. Learning of Math

We noticed a lot of fluidity across the learning of antecedent and descendant math curriculum topics (i.e., skills referred to in the ITS literature). We observed that about half of the students changed to other learner profiles when they learn new skills, even though the skills are conceptually linked. The linear regression model also showed that learning of antecedent skills has some but limited influence on the learning of subsequent new skills. For example, students' Initial_Ln of fraction division is barely predictable by their knowledge of whole number division and fraction multiplication. Although Initial_Ln and Final_Ln for faction multiplication can predict the Final_Ln of fraction division, the R^2 is still relatively low.

The fluidity across the learning of different math curriculum topics presents promising opportunities for low progress learners to leverage ITS as a level playing field for personalized learning and a remedial resource for academic advancement, particularly when learning difficult curriculum topics such as fraction division. The availability of ITS, therefore, offers a potential avenue to accommodate to the learning needs among different groups of students. This fluidity also cautions students doing well in math, demonstrating the need to refrain from complacency based on prior achievements when learning new curriculum topics. This serves as a reminder that performing well in one math topic does not necessarily translate to ease in mastering others. We acknowledge that general intelligence and math ability could contribute to the learning of new math curriculum topics as well, but these aspects are not measured in ITSs.

C. Enhancing the Design of ITSs

Our clustering of students' learning underscores the importance of a more nuanced way of profiling students in the student model. Student modeling and profiling are critical to ITSs for real-time adaptations, but most student models in ITSs only overlay the students' performance with the knowledge model to identify students' knowledge gaps [17]. They do not have comprehensive student profiling for real-time adaptation. Many education data mining papers have shown ways of profiling students. For example, Bouchet et al. [18] profile students according to their behavioral interactions with ITS. However, these profiling techniques are rarely built in ITSs for real-time profiling and adaptation.

The fluidity across the learning and the KLI dependency also provide a good argument for rich pedagogical approaches to be included and optimized in ITSs to optimize enhance personalized learning. Most ITSs provide direct instruction by showing students correct answers and by providing scaffolded hints. Learning Science research has, over the years, identified many powerful pedagogical approaches and learning strategies, for example cognitive load and worked examples [19], comparing contrasting cases [20], analogical learning [21], etc. Including these pedagogical and learning approaches, particularly those with good emotional engagement, in ITSs not only enhances promotes learning in ITSs through data driven optimization, but also provides an opportunity to further enhance the theorization of these pedagogical and learning approaches [22].

D. Limitations and Future Work

A few limitations are acknowledged to inform the future work. First, to examine the learning of whole number division, fraction multiplication and fraction division, we focused on a smaller sub-sample of the ASSISTments dataset. Because the sample is relatively small, some analysis could not be carried out. This limitation can be addressed in future with larger datasets or complementing it with qualitative case studies. For example, Owen, et al., [13] examine a dataset from a game-based adaptive learning system and find a group of learners with unproductive perseverance (i.e., students spending considerable time on a topic without achieving mastery). They called this behavior 'wheel-spinning'. A qualitative case study following our paper could examine how students in the HELA cluster learn over time to decipher whether students are cognitively engaged or just wheel-spinning. Extending our analysis to larger datasets, including data from other ITS platforms, could also provide a more comprehensive view of learners and their learning processes and outcomes.

Second, we did not critically examine the affect prediction labels (e.g., boredom, confusion, frustration, etc.) in the ASSISTments dataset for their validity corresponding to the respective constructs in the emotion literature. While critically examining these affect prediction labels is beyond the scope of this paper, future study could establish the theorical validity of these emotional constructs in ASSISTments to enable meaningful interpretation of students' emotional engagement in learning in ITS. For example, a more valid understanding of emotional and behavioral engagements could allow us to infer the cognitive engagement of the HELA students and to understand whether they are engaged or just wheel-spinning.

Third, a common limitation of ITS datasets is that these datasets capture only a portion of student learning, hence these ITS datasets provide incomplete, though important, understanding of student learning. As students usually learn and practice on ITS after the usual classroom teaching or as homework, a portion of student learning such as classroom learning is not captured in ITS. Hence, the dynamics between students' diverse learning behaviors and their learning gains in ASSISTments must be understood with this limitation in mind. Further research could consider integrating data from multiple sources, for example by examining how students learn both within and outside ITS, to provide a more holistic understanding of student learning behavior and outcomes.

Acknowledgment

We thank Mr Mrinal Kumar and Ms Simone Ann D Souza for their input in the initial exploration of the datasets.

References

[1] Ma, W., et al., *Intelligent tutoring systems and learning outcomes: A meta-analysis.* Journal of educational psychology, 2014. **106**(4): p. 901.

[2] Kabudi, T., I. Pappas, and D.H. Olsen, *AI-enabled adaptive learning systems: A systematic mapping of the literature.* Computers and Education: Artificial Intelligence, 2021. **2**: p. 100017.

[3] Steenbergen-Hu, S. and H. Cooper, *A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning.* Journal of educational psychology, 2013. **105**(4): p. 970.

[4] Muldner, K., et al., *An analysis of students' gaming behaviors in an intelligent tutoring system: Predictors and impacts.* User modeling and user-adapted interaction, 2011. **21**: p. 99-135.

[5] Schonert-Reichl, K.A., *Social and emotional learning and teachers.* The future of children, 2017: p. 137-155.

[6] Patikorn, T., N.T. Heffernan, and R.S. Baker. *Assistments longitudinal data mining competition 2017: A preface*. in *Proceedings of the Workshop on Scientific Findings from the ASSISTments Longitudinal Data Competition, International Conference on Educational Data Mining*. 2018.

[7] Pardos, Z.A., et al., *Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes.* Journal of Learning Analytics, 2014. **1**(1): p. 107-128.

[8] Pekrun, R. and E.J. Stephens, *Academic emotions*, in *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* 2012, American Psychological Association. p. 3-31.

[9] Sidney, P.G. and M.W. Alibali, *Creating a context for learning: Activating children's whole number knowledge prepares them to understand fraction division.* 2017.

[10] Mauro, D.G., J.-A. LeFevre, and J. Morris, *Effects of problem format on division and multiplication performance: division facts are mediated via multiplication-based representations.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 2003. **29**(2): p. 163.

[11] Huang, J.S. and S.A. D Souza, *Creative transfer and domain-specific knowledge: The effect of prior exposure*, in *Proceedings of the 15th International Conference of the Learning Sciences (ICLS)*, E. de Vries, Y. Hod, and J. Ahn, Editors. 2021, International Society of the Learning Sciences: Bochum, Germany.

[12] Rudin, C., *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead.* Nature machine intelligence, 2019. **1**(5): p. 206-215.

[13] Owen, V.E., et al., *Detecting Wheel-Spinning and Productive Persistence in Educational Games.* International educational data mining society, 2019.

[14] Richey, J.E., et al., *More confusion and frustration, better learning: The impact of erroneous examples.* Computers & Education, 2019. **139**: p. 173-190.

[15] Reeve, J. and C.-M. Tseng, *Agency as a fourth aspect of students' engagement during learning activities.* Contemporary educational psychology, 2011. **36**(4): p. 257-267.

[16] Koedinger, K.R. and E.A. McLaughlin, *The knowledge-learning-instruction (KLI) dependency: how the domain-specific and domain-general interact in STEM learning*, in *Integrating cognitive science with innovative teaching in STEM disciplines*, M.A. McDaniel, et al., Editors. 2014, Washington University Libraries: St. Louis, Missouri. p. 53-73.

[17] Sottilare, R.A., et al., *Design recommendations for intelligent tutoring systems: Volume 1-learner modeling*. Vol. 1. 2013: US Army Research Laboratory.

[18] Bouchet, F., et al., *Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning.* Journal of Educational Data Mining, 2013. **5**(1): p. 104-146.

[19] Kirschner, P.A., *Cognitive load theory: Implications of cognitive load theory on the design of learning.* Learning and instruction, 2002. **12**(1): p. 1-10.

[20] Schwartz, D.L., et al., *Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer.* Journal of Educational Psychology, 2011. **103**(4): p. 759.

[21] Novick, L.R. and K.J. Holyoak, *Mathematical problem solving by analogy.* Journal of Experimental Psychology: Learning, Memory, and Cognition, 1991. **17**(3): p. 398-415.

[22] Huang, J.S., et al., *Building Learning Theories with AI-Enhanced Conjectural Mapping*, in *20th Biennial EARLI Conference*. 2023, The European Association for Research on Learning and Instruction (EARLI): Thessaloniki, Greece.