Understanding the Relationships between Student Use of a Virtual Learning Environment for Algebra and Achievement on a High Stakes Test

Anonymous Author(s)

ABSTRACT

There has been a proliferation of virtual learning environments (VLE) available to students for both in-class and at-home use. This study aimed to understand the relationships between student self-regulated learning (SRL) with a VLE across a year and achievement on a high-stakes test. The study used log data from an Algebra VLE used by middle and high school students, and standardized test scores on a high-stakes Algebra 1 test obtained from a large school district. The study used random forests, neural networks, and clustering methods to understand the relationships between time investment, study regularity, and help seeking behaviors in the VLE and Algebra achievement. The results show that student actions related to practice problems were the most important predictors of students passing the high-stakes Algebra 1 assessment mandated by the state. Activities related to viewing videos and participating in a discussion forum were also related to Algebra 1 scores. Time investment was related to achievement, as students with higher frequency of actions performed better in the standardized test. Self-help actions, such as reviewing solution videos and loading the discussion board were important predictors of achievement. However, the importance of regularity of use was not supported by the results, as a surge of use in the fourth quarter was predictive of Algebra 1 scores.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; • **Computing methodologies** → **Online learning settings**.

KEYWORDS

Virtual learning environments, self-regulated learning, algebra, random forests, neural networks, clustering

ACM Reference Format:

Anonymous Author(s). 2024. Understanding the Relationships between Student Use of a Virtual Learning Environment for Algebra and Achievement on a High Stakes Test. In *Proceedings of Learning @ Scale, 2024 (L@S '24).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

L@S '24, July 18–20, 2024, Atlanta, GA

ACM ISBN 978-1-4503-XXXX-X/18/06

https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

There has been a proliferation of virtual learning environments (VLE) available to students for both in-class and at-home use. Some VLE are simply collections of resources for student use, such as videos and practice problems, while other VLE incorporate student knowledge modeling, advanced scaffolding of content, and hints [1]. The effectiveness of VLE in increasing student achievement has been supported by multiple studies [19, 29]. However, there is also evidence that some students benefit greatly from VLE use while others do not. This heterogeneity of effects of VLE has been linked to previous student achievement and self-regulated learning (SRL) strategies [21]. Still, research on student use of VLE over time has been scarce. The current study aims to understand the relationships between student SRL with a VLE across a year and achievement on a high-stakes test. This is critical because understanding the trajectories of student use that lead to the most benefit from VLE can help teachers encourage certain uses of the VLE and optimize its benefits. Because it is common for VLE to have teacher dashboards that allow monitoring of student use trajectories over time, if teachers understand which trajectories of use are related to student achievement, then they can better orchestrate use so that all students follow optimal trajectories of VLE use.

2 BACKGROUND AND RELATED WORK

In this study, we use the theoretical framework of SRL [37] to understand variation in student engagement with a VLE. Previous research has shown that SRL strategies play a crucial role in shaping students' experience with VLE [21], both during class time and outside. Research has shown positive correlations between student self-regulation, academic enjoyment, and performance, as well as between goal setting, self-evaluation, and achievement [8, 9]. Fully 24% of variance in student mathematical problem-solving skills was explained by the SRL dimensions of internal motivation, willingness to complete assigned work, and reflection on work [23]. SRL strategies encompass a student's intrinsic motivation to learn and involve practices like structuring their environment, setting goals, self-rewarding and self-punishing, self-evaluating, seeking social assistance, and reviewing class notes [38]. Indeed, in today's diverse mathematics classrooms, SRL tools, scaffolds, and interventions provide diverse supports for learning [33]. Further, a mix of SRL strategies appears to be significantly more effective for secondary STEM learning than individual strategies [33].

The rise of online learning has highlighted the importance of SRL, as it empowers students to take an active role in their learning process, adjusting effectively to various VLE. With greater agency over when, what, and how to learn, students displaying strong SRL skills are able to optimize their learning experiences both inside and outside the classroom [20, 25]. Social regulation approaches applied in a VLE have made learners more likely to review information related to questions they answered incorrectly [15]. Meta-cognitive questions influence students to make more mathematical inferences based on prior learning as a bridge to new concepts [13]. Goal setting and self-evaluation have shown positive associations with academic achievement in math classrooms. Moreover, students' academic enjoyment and self-perception are associated with SRL and overall achievement. Conversely, deficiencies in SRL skills can hinder students' ability to comprehend complex topics and achieve conceptual understanding [4].

There is strong evidence in the literature that SRL strategies are context specific [5, 31, 39]. To support SRL in the context of VLE, various strategies have been identified, such as prompts, integrated support systems, feedback, or a combination thereof. Of these, prompts are the most employed strategy [32]. Prompts in the form of task definition have been shown to positively mediate the relationship between metacognitive skills and persistence in a VLE by building a student's awareness of expectations in a learning experience [27]. From an instructional standpoint, promoting SRL behaviors in online settings is recommended to enhance students' performance and learning outcomes [38].

In order to promote SRL in VLE, it is necessary for teachers to have an understanding of each students' SRL skills because SRL includes teacher support and students who perceive that they receive insufficient support are at increased risk of poor performance on mathematics assessments [10]. While academic research has measured SRL with self-administered scales such as the Motivated Strategies for Learning Questionnaire (MSLQ) or the Learning and Study Strategies Inventory (LASSI) as an overall measure of SRL [8], these measures are not practical for classroom use at scale. Instead, VLE can provide measures of SRL based on logs of student actions in the VLE. Recently, there have been multiple studies using log data to measure SRL in VLE, with indicators such as time logged in, posts to discussion forums, and videos watched [18, 21, 32]. Analytics methods for measuring SRL in mathematics VLEs have accurately represented student self-reported SRL using VLE activity [36].

The current study addresses the following research question: In what ways do students' SRL with a mathematics VLE over an academic year relate to academic achievement? Understanding SRL from logs of student actions in VLE requires mapping actions available to students to a specific SRL model. Kim et al. [18] proposed a model to understand SRL from log data based on time investment, study regularity, and help seeking behaviors. Learners displaying strong SRL recognize that investing time in course content is crucial for understanding topics. Effective time management strategies involve allocating time for learning activities. VLEs allow students to study both under the guidance of an instructor and independently. Engagement with content is vital for grasping course concepts, but students with high self-regulation optimize their engagement with learning objects within a VLE by focusing on those that address their learning needs most closely. Students adjust their exposure to learning objects based on their perceived importance.

Study regularity is the repeated engagement with the VLE over time [12]. Regularity can be measured at different time intervals, such as daily, weekly, or monthly. A uniform distribution of engagement with a VLE over time would indicate high regularity, while strongly skewed distribution (such using the VLE mostly on the day before a test) would indicate low regularity. Regularity is associated with diligence, punctuality, time management, and grit [31]. Study regularity may be influenced by teachers depending on the extent of classroom use versus home use of the VLE, and specific orchestration strategies employed by teachers [11]. This is because a teacher may employ a VLE regularly for classroom use, thus imposing regularity of student engagement with the VLE.

In a VLE, students may engage in help-seeking behaviors in different ways, such as asking questions, reviewing solutions to questions, and obtaining hints [1]. Some VLE may offer opportunities for students to ask questions within the system through a discussion forum [6] or chatbot [14]. Reviewing solutions to questions may be performed before, during, or after a student engages with practice problems. Some VLEs allow reviewing of solutions before practice by providing example problems. Other VLEs show partial solutions during practice by allowing students to request on-demand hints. VLE may also show complete solutions after the student has submitted the answer to a practice problem. These solutions may be in the form of explanatory text or video.

3 METHODS

3.1 Setting

The current study used log data from the algebra section of a mature VLE that has been available for over 10 years to all students in a southeastern state of the United States. Data collection occurred in the 2021/2022 school year. The Algebra curriculum implemented in the VLE is organized into 10 domains (e.g. linear functions, quadratic functions), each with 7 to 12 topics. For each topic, the VLE has a set of videos where tutors describe solutions to example problems. There are five versions of each video taught by tutors of varying genders and ethnicity. The topics also contains a 3-item quiz with multiple choice and constructed-response items. Each domain has a pool of practice questions and the VLE can generate 10-question quizzes by randomly drawing form the item pool. In the VLE, a quiz is referred to as "Test Yourself" (TYS). Each question also has a solution video. The VLE does not offer hints for the questions. There is a discussion board where students can ask each other questions, which is named "Algebra Wall" in the VLE. There is also a separate area with practice questions for pre-Algebra remediation, which we will refer to as "On the Ramp" tool (ORT).

The study focuses on a single school district that has adopted the VLE as its official algebra curriculum provider. The researchers obtained logs of student use of the VLE by directly querying the VLE's database, and student achievement and demographic variables were obtained from the school district.

3.2 Participants

Participants include 17,912 students from a large district. The school district reported the following demographics for participating students: 51.76% female, 48.24% male, 35.52% Hispanic, 30.87% White, 26.17% Black, 4.09% Asian, 2.75% mixed race, and 0.61% Native American. Just over half (51.18%) were eligible for free or reduced lunch. Students were primarily from 9th grade (32.83%), followed by 8th grade (30.66%), 10th grade (19.85%), 7th grade (9.57%), 11th grade (4.81%), 6th grade (0.99%), and 12th grade (0.46%), with a small percentage in other or unspecified grades (0.82%).

3.3 Measures

The variables used to indicate student engagement with the VLE were constructed based on detailed logs of video views (e.g., video play, pause, rewind), quiz responses (e.g., quiz completed, number correct, incorrect questions reviewed), and discussion board actions (e.g., load discussion board, make a post). We operationalized time investment in the VLE as the number of logs of each student action per quarter. The school district's calendar provided the dates that divided the school year into quarters of approximately equal duration. Quarters are important for instruction because the students' report card is made available to them and their parents at the end of each quarter. We summed the logs of each action by quarter, resulting in four measures of each action (e.g., video views 1, 2, 3, and 4). This resulted in 88 features. We operationalized regularity by examining the differences in frequency of logs across quarters. Help-seeking behaviors were examined with a specific set of actions, which included actions on the discussion board, reviewing solutions to incorrect questions, and reviewing solution videos.

The log data contained outliers due to logging errors of the system. We dealt with outliers by replacing any values above the 97.5th percentile for each variable by missing values, which were handled differently in each analysis method, as detailed below.

The outcomes of the analyses are the student's standardized score in the Algebra 1 End of Course (EOC) assessment, which is a high-stakes test mandated by the state's department of education and required for graduation. A complete list of variables is on Table 1, and a binary indicator of whether students passed the assessment. The state divides the standardized scores into five achievement levels (1, 2, 3, 4 and 5), and students are required to attain a level 3 to pass the assessment. The Algebra 1 EOC assessment is administered in May. Students that do not pass the spring assessment may re-take it in the fall or next spring.

3.4 Analysis

The analysis consisted of two stages. The first stage was to use supervised machine learning methods to identify which of the 22 student actions with the VLE were most predictive of Algebra 1 EOC scores. For this step, we used random forests [7] and neural networks [26, 28, 34]. The second stage was to use unsupervised learning methods to cluster the students according to their VLE actions, and compare the cluster means of Algebra 1 EOC scores.

3.4.1 *Supervised Learning.*

Random Forests. We trained random forest classifiers and feed-forward neural networks to predict students passing the Algebra 1 EOC. We included the variables indicating student usage of the VLE presented in the measures section. We did not include demographic variables in the models because the focus was to identify the student actions in the VLE related to passing the Algebra 1 EOC that could be monitored and encouraged by teachers. Random forest classifiers [7] are an ensemble supervised machine learning algorithm that uses decision trees based on random subsets of the dataset. A neural network is a machine learning program comprised of multiple layers of nodes, each containing an activation function which determines whether they pass data to the next layer. In a classification model, the last layer outputs the final classification decision. To read the data in Python, we used the pandas data analysis library [16].

We split the data into train and test data with a 75:25 split with the scikit-learn library [24]. This library was also used to create and fit the random forest to the dataset. We used default parameters, which are as follows: number of trees (n estimators = 100), measure the quality of a split (criterion = "gini"), minimum number of samples required to split an internal node (min samples split = 2), minimum number of samples required to be at a leaf node (min samples leaf = 1), number of features to consider when looking for the best split (max features = "sqrt"), bootstrap samples are used when building trees (bootstrap = True).

The scikit-learn library also includes model selection methods like GridSearchCV and RandomizedSearchCV [24] to tune the hyperparameters of the classifier such as number of trees, maximum depth of each tree, and the number of samples required for a leaf node. Due to grid search being computationally expensive, we used randomized search with lists of possible values for each hyperparameter. After tuning, the parameters were set to default except that the minimum samples required for a split was 6 and the number of trees was 1000. However, due to the random aspect of the search, the results of the hyperparameter tuning varied. We ran RandomizedSearchCV 10 times to find the optimal parameters. Variable importance was evaluated with permutation importance [3], because impurity-based importances are biased towards high-cardinality and numerical features, while permutation importance is robust.

Neural Networks. The Keras API (Chollet et al., 2015) for TensorFlow (Abadi et al., 2015) was used to build a feed-forward neural network with the goal of classifying whether a given student passed the Algebra I EOC. Before training, we used the train test split module from scikit-learn to create an 80:20 split between training and testing data. The data was then scaled using the StandardScalar class from scikit-learn.

The model's structure was as follows: The input layer contained 128 nodes with the rectified linear unit (ReLU) activation function (Fukushima, 1975). We used 3 internal layers of 256 nodes each, also using ReLU. The output layer was comprised of a single node, using the sigmoid activation function. Our model was trained using the Keras Adam optimizer with a learning rate of .01 over 50 epochs. Loss was calculated with the binary crossentropy function from Keras. Following training, the effectiveness of the neural network was evaluated against the test data with the accuracy, F1 score, and AUC metrics from scikit-learn. Feature importance was ranked by permutating the entries of each column of the test data set. The baseline accuracy was compared to the accuracy obtained with the permutated column, and the difference was calculated as that column's importance (Hardy, 2020). This procedure was repeated 100 times for each feature, resulting in a mean decrease in accuracy.

3.4.2 Unsupervised Learning. Clustering allowed us to contrast different patterns of activity in the VLE and examine differences between clusters regarding Algebra 1 EOC scores. Rather than clustering based on all 88 features built from the logs of student actions in the VLE, we grouped the features into three categories of learning objects: practice questions, instructional videos, and discussion board. Example features in the instructional video category include playing a video and toggling caption. Example features in the practice question category include finishing a quiz and reviewing incorrect questions. Example features in the discussion board category include loading the board and making a post. These features were measured at each of four quarters.

To handle the large set of features, we implemented the BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm [35] using the scikit-learn package. This method is similar to k-means in the sense that the operations are equivalent, however, with the BIRCH algorithm, these operations are applied to distinct subclusters. Subclusters are created by finding the minimum Euclidean distance between datapoints and grouping them together. The number of clusters was determined by the elbow method, the technique plotted the sum of squares for each cluster where an optimal "elbow" point was distinctly identified. Prior to Birch clustering, data was reduced through the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm by giving each datapoint a location in a two or three-dimensional map [30]. BIRCH clustering typically has high sensitivity to data with many dimensions, because of the difficulty calculating Euclidean distances when creating subclusters. Therefore, t-SNE was implemented to not only enhance the quality of each cluster but to enhance the quality of subclusters withing the BIRCH algorithm.

Ordinary least squares (OLS) regression was used to estimate the mean differences between clusters. A separate regression was fit for the clusters of practice questions, instructional videos, and discussion board. The regression models controlled for retaker status, performance on a mathematics high-stakes assessment in the previous academic year, and school membership. The mathematics high-stakes assessment serves as a proxy of a pre-test, because the correlation between the Algebra 1 EOC in 2021/2022 and the mathematics test in 2020/2021 was 0.78. Fixed effects of school membership were controlled by creating dummy-coded indicators of schools and including them in the model. This approach has the advantage of estimating within-school cluster differences and thus controlling for all unobserved confounders at the school level [2]. Estimated marginal means (EMMs) for each cluster provided insight into their average performance, adjusted for covariates. Pairwise comparisons were used to evaluate the significance of differences in EOC performance between clusters adjusting for inflation of Type I error rate with the Bonferroni method.

4 RESULTS

4.1 Supervised Learning

4.1.1 Random Forest. The random forest with default hyperparameters of scikit-learn resulted in an area under the ROC curve (AUC) accuracy of 0.80 for predicting passing the Algebra 1 EOC assessment. The AUC improved by 0.01 to 0.02 after hyperparameter tuning. The permutation importances (see Table 1) indicates that the most important predictor was answering an item on a quiz. The "quiz answer" predictor was significantly more important than other variables, as without it the accuracy would decrease by 9.8%, while all other features were below 3%. The other most important features were loading the discussion page, watching a video, and reviewing solution videos. However, the feature importances varied based on which of the four school quarters the actions were completed in. Figure 1 shows the variable importance by quarter of the 5 most important features. Individually, answering a quiz question, loading a discussion, and watching a video were all most important in the 4th quarter, but the importance of answering practice quizzes increases much more than of the other actions.

Table 1: Permutation Importance by Ranking and Accuracy*

	Random Forests		Neural Networks	
	Rank	Decrease in Accuracy	Rank	Decrease in Accuracy
Quiz answer	1	0.098	1	0.159
Discussion page load	2	0.026	4	0.047
Video play	3	0.018	7	0.034
Quiz review solution video	4	0.017	12	0.023
Video caption	5	0.012	5	0.037
Video watch	6	0.012	13	0.023
Quiz finish	7	0.008	3	0.051
Quiz load	8	0.007	10	0.027
Video pause	9	0.005	2	0.052
ORT load	10	0.004	17	0.007
Video completed	11	0.003	9	0.028
Quiz review incorrect question	12	0.003	11	0.024
Video seek	13	0.002	6	0.035
ORT Video watch	14	0.0007	16	0.009
Quiz review topic video	15	0.000	18	0.006
Discussion search	16	0.000	20	0.002
Discussion load more	17	0.000	19	0.006
Discussion make post	18	0.000	21	0.002
Quiz review correct question	19	0.000	15	0.009
Document view	20	0.000	22	-0.001
Quiz previous	21	-0.001	8	0.030
Leaderboard load	22	-0.001	14	0.015

Note. Decrease in Accuracy indicates by how much the accuracy of the model decreases by permuting the values of the respective features.

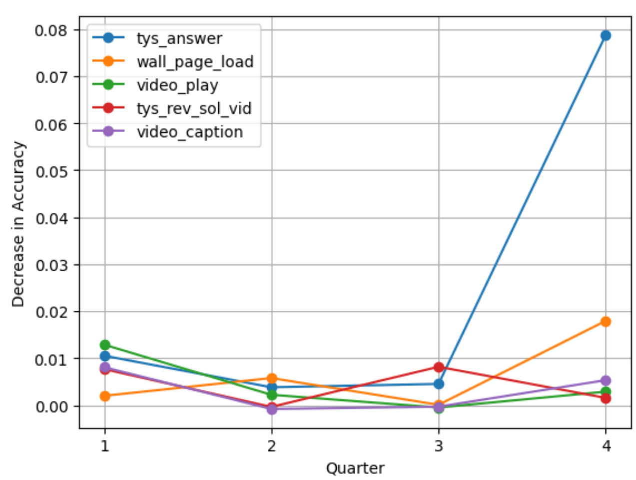

Figure 1: Variable Importance Over Time from Random Forest

Note. tys is a quiz. tys_rev_sol_vid is to review the solution video for a quiz. Wall is the discussion board.

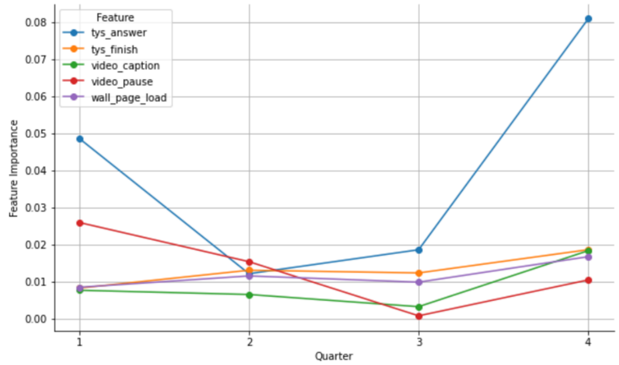

Figure 2: Variable Importance Over Time from Neural Network

Table 2: Cluster Sizes

	0	1	2	3
Practice	12608 (70%)	3111 (17%)	986 (6%)	1207 (7%)
Video	10062 (56%)	4994 (28%)	2856 (16%)	NA
Discussion	7811 (44%)	4716 (26%)	5385 (30%)	NA

4.1.2 Neural Network. The results from neural networks were highly related to the results from random forests. Neural networks provided an AUC of .74. The feature importance for the neural network is shown in Table 1. The correlation between the decrease in accuracy measure of feature importance between random forest and neural network was 0.92.

The action of answering quiz questions was the most important predictor of success on the EOC. This was followed by pausing videos, completing quizzes, loading a discussion board, and turning on video captions. Similar to the random forests, the neural network found that answering questions was by far the most important, with its importance increasing in the 4th quarter (see Figure 2).

4.2 Unsupervised Learning

4.2.1 Cluster Descriptions. The Birch Clustering with dimensionality reduction revealed distinct clusters with each of the three categories of learning objects: practice questions, instructional videos, and discussion board. The cluster sizes and percentages are shown in Table 2.

With practice problems (quizzes), Cluster 0 had the largest proportion of students. Cluster 0 had the lowest engagement with quiz actions, followed by cluster 1. Cluster 1 had higher quiz loading, answering and finishing than Cluster zero, but similar very low engagement with help-seeking behaviors of reviewing incorrect questions and reviewing the topic videos that presented the example for the questions. In cluster 2, students had above average usage of all actions, with reviewing topic video having extraordinarily high usage. Cluster 3 had lower frequency of loading a quiz than Cluster 2, but higher frequency of answering and finishing the quiz. Also, Cluster 3 had much higher frequency of reviewing incorrect questions than Cluster 2. The results show that students in both Cluster 2 and 3 were involved in help-seeking behaviors,

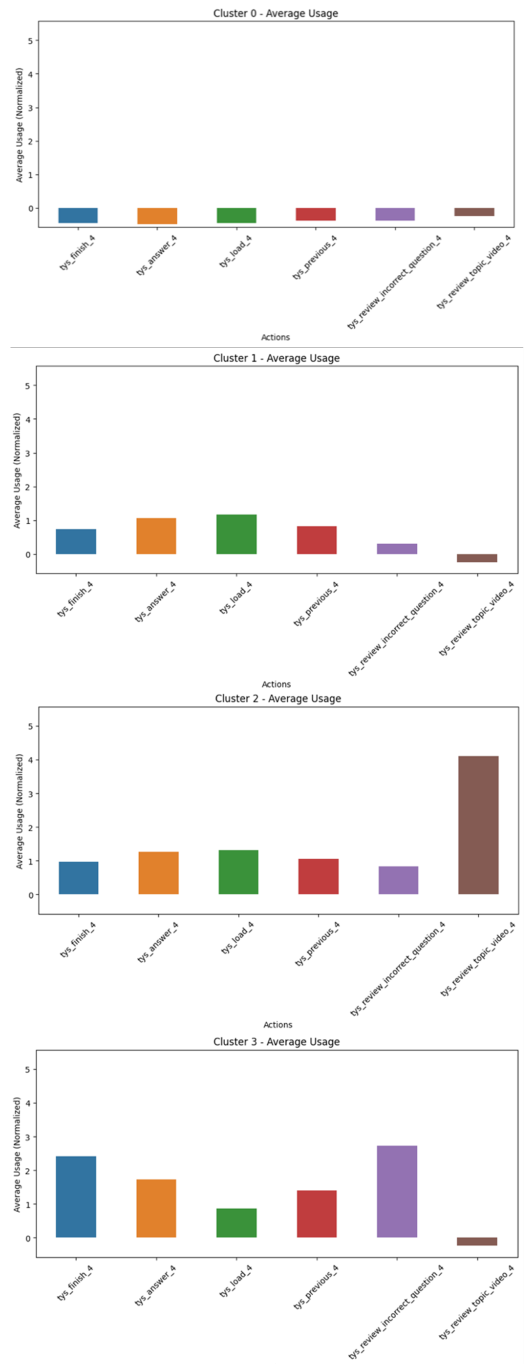

Figure 3: Clusters for features related to practice questions

but while Cluster 2 focused on reviewing videos, Cluster 3 focused on completing practice questions and reviewing incorrect answers.

With Video Activity, Cluster 0 was the largest, followed by Cluster 1 and 2. Members of Cluster 0 and Cluster 1 had low engagement with video, but Cluster 0 had higher video completion than Cluster

1. Cluster 2 was very different than 0 and 1, because it had very high engagement with all available actions related to video.

For clusters based on the discussion board actions, Cluster 0 was the largest, but the differences in cluster sizes were not as pronounced as with the practice question and video clusters. Cluster 2 had the lowest levels of engagement with all actions available in the discussion board. Cluster 0 was similar to Cluster 2, but had higher frequency of loading the discussion board. Cluster 1 had high levels of engagement with the discussion board for all actions.

4.2.2 Regression. For each category of actions, a regression was fit to determine the relationship between cluster membership and Algebra 1 EOC scores, adjusting for covariates. The marginal means for clusters estimated with the regression models are shown in Figure 6.

For practice-related actions, pairwise comparisons showed significant differences between cluster 0 and clusters 1, 2, and 3. Cluster 3, which had the highest frequency of practice-related actions, had the largest marginal mean on EOC scores and largest difference from cluster 0 (β = 0.15, SE = 0.02, t = 6.23, p < 0.0001). Cluster 3 was also had significantly larger marginal mean than cluster 2 (β = 0.06, SE = 0.02, t = 1.8, p < 0.05), but the difference between the means of clusters 2 and 3 was not statistically significant (p).

For video-related actions, pairwise comparisons indicated that the marginal mean of cluster 1 was not significantly different from cluster 0 (p), but cluster 2 was significantly different from clusters 0 and 1. More specifically, members of cluster 2, which had high engagement with video actions, performed better than those in cluster 0 (β = 0.10, SE = 0.02, t = 5.20, p < 0.0001) and cluster 1 (β = 0.11, SE = 0.02, t = 5.25, p < 0.0001) on the Algebra 1 EOC assessment.

For discussion-related actions, the difference between means of cluster 0 and cluster 1 was significant (β =-0.06, SE = 0.02, t = -3.70, p = .0007). This indicates that cluster 1, which had the highest engagement with the discussion board, also had the highest mean Algebra 1 EOC score. Cluster 2 had significant lower mean that Cluster 0 (β = -0.06, SE = 0.02, t = 3.25, p = .0035) and 1 (β = 0.12, SE = 0.02, t = 5.38, p < .0001), suggesting that the low engagement with discussion actions characterized by cluster 2 was least beneficial to EOC performance.

5 DISCUSSION

The results indicate that SRL indicators related to practice (i.e. quizzes), exposure to new content (i.e., videos), and participating in discussion (i.e., Algebra Discussion) were associated with Algebra 1 End-of-Course exam scores. From the set of indicators examined, practice indicators had higher importance than discussion and exposure to new content indicators. The results validate the importance of teacher's assigning practice problems to students and monitoring the indicators of student SRL in the VLE dashboards (Broadbent et al., 2021). Although answering practice problems were the most important feature, the results support a balanced approach of making different types of learning objects available to students in VLEs, as video features and discussion board features also were identified among the most important predictors of passing the Algebra 1 EOC assessment (Xu, Zhao, Zhang, Liew & Kogut, 2023).

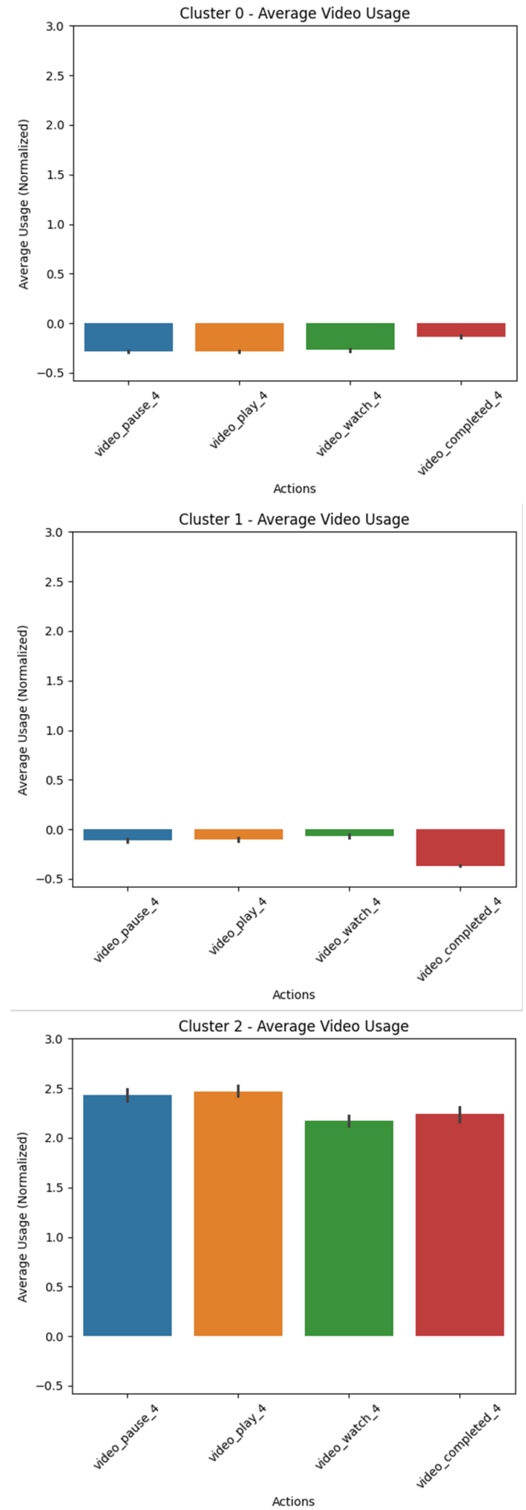

Figure 4: Clusters for Features Related to Videos

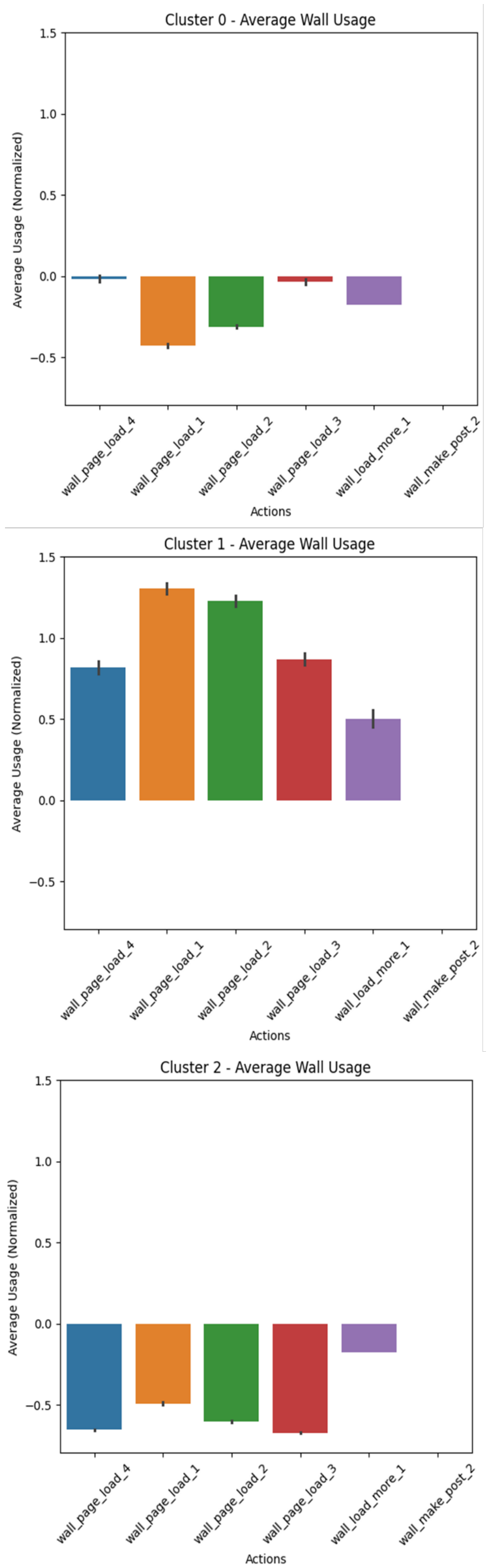

Figure 5: Clusters for Features Related to Discussion Board

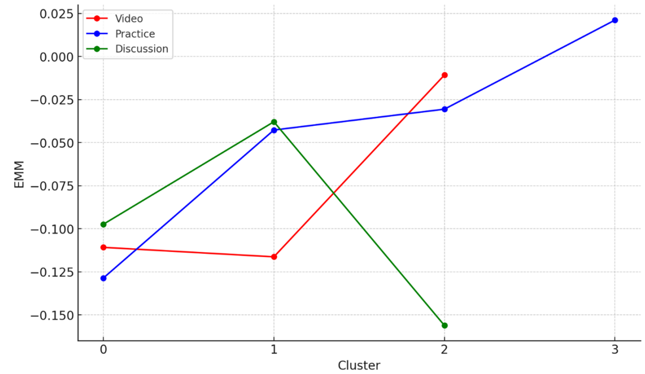

Figure 6: Estimated Marginal Means by Clusters

The results of regression analysis of the relationship between student achievement and clusters have connections to time investment, study regularity, and help seeking behaviors dimensions of SRL [18]. For time investment, the regression results clearly show that students who invested more time in the VLE by engaging in more actions obtained higher standardized scores on the Algebra 1. However, the random forest and neural network prediction results do not support the role of regularity in increasing assessment passing rates, because they show that actions in the fourth quarter were most important. Furthermore, the features most relevant for clustering were actions in the fourth quarter for both practice questions and video categories. For the discussion board, clustering did not select features in the 4th quarter. Interestingly, random forests detected an increased importance of discussion board loading on the 4th quarter, but neural networks did not. Therefore, for VLEs that target preparation for high-stakes testing, increased usage in the weeks prior to the test seems more important than regularity of usage throughout the year. There was evidence of the importance of help-seeking behaviors, as the random forest selected loading the discussion board and reviewing solution videos of practice questions among the 5 most important indicators. With neural networks, the evidence of the importance of help seeking behaviors was not as strong but loading the discussion board still showed among the five most important features (Hwang, Wang & Lai, 2021). The regression results for the mean differences between discussion board clusters also support the importance of help-seeking behaviors, because the cluster with highest usage of the discussion board had significant higher means in the Algebra 1 EOC that the cluster with lowest usage.

6 CONCLUSION

6.1 Limitations

The results are limited by the learning supports available in the VLE that could be used indicate SRL. For example, the VLE does not have on-demand hints or collaboration tools. Also, because the VLE is the adopted curriculum in the school, engagement with it is much higher than it would be if it was a supplemental resource, which may have reduced variability of use and consequently the power to detect group differences.

The study is also limited in that it did not examine the impact of teacher orchestration on student SRL. Previous studies have shown the importance of teacher orchestration of VLE [17, 21, 22], but no data about how the teacher was using the VLE in the classroom was available. Teacher surveys would have enriched the analysis. Also, the VLE has a dashboard that allows teachers to monitor student engagement, but teachers' interactions with the dashboard are not currently being logged.

6.2 Implications for Practice

The results encourage developers to create VLE with a variety of supports because multiple types are important in increasing student achievement. This agrees with meta-analysis evidence that combinations of multiple types of SRL strategies have the largest effect size [33]. Among these supports, those that facilitate practice of problem solving and allow help seeking are particularly important [1]. From the teacher perspective, the results highlight the importance of encouraging students to develop help-seeking strategies. For VLE that are used as the main instructional method, it is easier for teachers to require high frequency and regularity of use, as compared to VLE that are supplemental resources. This is because the teacher can dedicate classroom time to working on the VLE, for activities such as warm up and group work. However, it is more difficult for teachers to require and monitor help-seeking behaviors, and students have more control of whether they look for help, hints, solutions, or examples for the problems they face, or simply move forward to other topics. Therefore, it is critical that teachers model and encourage the full variety of help-seeking behaviors that are available through the VLE. The importance of encouraging student SRL also indicates the need to support development of teacher PD to prepare teachers to better orchestrate the use of VLE in the classroom with activities that encourage SRL and show strongest connections to student achievement.

REFERENCES

[1] Vincent Aleven, Ido Roll, Bruce M. McLaren, and Kenneth R. Koedinger. 2016. Help Helps, But Only So Much: Research on Help Seeking with Intelligent Tutoring Systems. *International Journal of Artificial Intelligence in Education* 26, 1 (2016), 205–223. https://doi.org/10.1007/s40593-015-0089-1

[2] P. D. Allison. 2009. *Fixed effects regression models.* Sage, Los Angeles, CA.

[3] A. Altmann, L. Tolosi, O. Sander, and T. Lengauer. 2010. Permutation importance: a corrected feature importance measure. *Bioinformatics* 26 (2010), 1340–1347.

[4] Roger Azevedo and A. F. Hadwin. 2005. Scaffolding self-regulated learning and metacognition: Implications for the design of computer-based scaffolds. *Instructional Science* 33 (2005), 367–379. Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds..

[5] Lucy Barnard, William Y. Lan, Yen M. To, Valerie Osland Paton, and Shu-Ling Lai. 2009. Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education* 12, 1 (2009), 1–6. https://doi.org/10.1016/j.iheduc.2008.10.005

[6] Fernanda Cesar Bonafini, Chungil Chae, Eunsung Park, and Kathryn Weed Jablokow. 2017. How Much Does Student Engagement with Videos and Forums in a MOOC Affect Their Achievement? *Online Learning* 21, 4 (2017).

[7] L. Breiman. 2001. Random forests. *Machine Learning* 45, 5-32 (2001).

[8] J. Broadbent and W. L. Poon. 2015. Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education* 27 (2015), 1–13.

[9] Jaclyn Broadbent, Stefanie Sharman, Ernesto Panadero, and Matthew Fuller-Tyszkiewicz. 2021. How does self-regulated learning influence formative assessment and summative grade? Comparing online and blended learners. *The Internet and Higher Education* 50 (01 Jun 2021), 100805. https://www.sciencedirect.com/science/article/pii/S1096751621000142

[10] Timothy J. Cleary, Jacqueline Slemp, and Erica R. Pawlo. 2020. Linking student self-regulated learning profiles to achievement and engagement in mathematics. *Psychology in the Schools* 58, 3 (2020), 443–457. https://doi.org/10.1002/pits.22456

[11] Paul Drijvers, Michiel Doorman, Peter Boon, Helen Reed, and Koeno Gravemeijer. 2010. The teacher and the tool: instrumental orchestrations in the technology-rich mathematics classroom. *Educational Studies in Mathematics* 75, 2 (2010), 213–234. https://doi.org/10.1007/s10649-010-9254-5

[12] Tomas Dvorak and Miaoqing Jia. 2016. Online Work Habits and Academic Performance. *Journal of Learning Analytics* 3, 3 (2016), 318–330.

[13] Peidi Gu and Youngjin Lee. 2018. Promoting Students' Motivation and Use of SRL Strategies in the Web-Based Mathematics Learning Environment. *Journal of Educational Technology Systems* 47, 3 (2018), 391–410. https://doi.org/10.1177/0047239518808522

[14] Weijiao Huang, Khe Foon Hew, and Luke K. Fryer. 2021. Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning* 38, 1 (2021), 237–257. https://doi.org/10.1111/jcal.12610

[15] Gwo-Jen Hwang, Sheng-Yuan Wang, and Chiu-Lin Lai. 2021. Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. *Computers & Education* 160 (2021). https://doi.org/10.1016/j.compedu.2020.104031

[16] numFocus Inc. 2024. Pandas. https://pandas.pydata.org/

[17] A. Joyce-Gibbons. 2014. *An exploratory study of teacher orchestration of collaborative Mathematics tasks in relation to learning and interaction in primary schools.* Technical Report. http://etheses.dur.ac.uk/9464/

[18] Dongho Kim, Meehyun Yoon, Il-Hyun Jo, and Robert Maribe Branch. 2018. Learning analytics to support self-regulated learning in asynchronous online courses: A case study at a women's university in South Korea. *Computers & Education* 127 (2018), 233–251. https://doi.org/10.1016/j.compedu.2018.08.023

[19] James A. Kulik and J. D. Fletcher. 2016. Effectiveness of Intelligent Tutoring Systems. *Review of Educational Research* 86, 1 (2016), 42–78. https://doi.org/10.3102/0034654315581420

[20] T. Lehmann, I. Hähnlein, and D. Ifenthaler. 2014. Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in human behavior* 32 (2014), 313–323.

[21] Walter L. Leite, Huan Kuang, Zeyuan Jing, Wanli Xing, Catherine Cavanaugh, and A. Corinne Huggins-Manley. 2022. The relationship between self-regulated student use of a virtual learning environment for algebra and student achievement: An examination of the role of teacher orchestration. *Computers & Education* 191 (2022). https://doi.org/10.1016/j.compedu.2022.104615

[22] Chee-Kit Looi and Yanjie Song. 2013. Orchestration in a networked classroom: Where the teacher's real-time enactment matters. *Computers & Education* 69 (2013), 510–513. https://doi.org/10.1016/j.compedu.2013.04.005

[23] Zeynep Çiğdem Özcan. 2015. The relationship between mathematical problem-solving skills and self-regulated learning through homework behaviours, motivation, and metacognition. *International Journal of Mathematical Education in Science and Technology* 47, 3 (2015), 408–420. https://doi.org/10.1080/0020739x.2015.1080313

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, and O. Grisel. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[25] B. Rienties, D. Tempelaar, Q. Nguyen, and A. Littlejohn. 2019. Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior* 100 (2019), 345–357.

[26] Alberto Rivas, Alfonso González-Briones, Guillermo Hernández, Javier Prieto, and Pablo Chamoso. 2021. Artificial neural network analysis of the academic performance of students in virtual learning environments. *Neurocomputing* 423 (2021), 713–720. https://doi.org/10.1016/j.neucom.2020.02.125

[27] Damian J. Rivers, Michiko Nakamura, and Michael Vallance. 2021. Online Self-Regulated Learning and Achievement in the Era of Change. *Journal of Educational Computing Research* 60, 1 (2021), 104–131. https://doi.org/10.1177/07356331211025108

[28] Carlos Felipe Rodríguez-Hernández, Mariel Musso, Eva Kyndt, and Eduardo Cascallar. 2021. Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence* 2 (2021). https://doi.org/10.1016/j.caeai.2021.100018

[29] Fabienne M. Van der Kleij, Remco C. W. Feskens, and Theo J. H. M. Eggen. 2015. Effects of Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes. *Review of Educational Research* 85, 4 (2015), 475–511. https://doi.org/10.3102/0034654314564881

[30] L. van der Maaten and G. Hinton. 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9 (2008), 2579–2605.

[31] Christopher A. Wolters and Maryam Hussain. 2014. Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning* 10, 3 (2014), 293–311. https://doi.org/10.1007/s11409-014-9128-9

[32] J. Wong, M. Baars, D. Davis, T. Van Der Zee, G. J. Houben, and F. Paas. 2019. Supporting self-regulated learning in online learning environments and MOOCs:

A systematic review. *International Journal of Human–Computer Interaction* 35 (2019), 356–373.

[33] Zhihong Xu, Yingying Zhao, Jeffrey Liew, Xuan Zhou, and Ashlynn Kogut. 2023. Synthesizing research evidence on self-regulated learning and academic achievement in online and blended learning environments: A scoping review. *Educational Research Review* 39 (2023). https://doi.org/10.1016/j.edurev.2023.100510

[34] Nick Z. Zacharis. 2016. Predicting Student Academic Performance in Blended Learning Using Artificial Neural Networks. *International Journal of Artificial Intelligence & Applications* 7, 5 (2016), 17–29. https://doi.org/10.5121/ijaia.2016.7502

[35] T. Zhang, R. Ramakrishnan, and M. Livny. 1996. BIRCH: An Efficient Data Clustering Method for Very Large Databases. , 103-114 pages.

[36] Denis Zhidkikh, Mirka Saarela, and Tommi Kärkkäinen. 2023. Measuring self-regulated learning in a junior high school mathematics classroom: Combining aptitude and event measures in digital learning materials. *Journal of Computer Assisted Learning* 39, 6 (2023), 1834–1851. https://doi.org/10.1111/jcal.12842

[37] B. J. Zimmerman. 1989. A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology* 81 (1989), 329–339.

[38] B. J. Zimmerman. 2008. Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal* 45, 1 (2008), 166–183.

[39] B. J. Zimmerman. Jan. Self-Efficacy: An Essential Motive to Learn. *Contemp Educ Psychol* 25, 1 (Jan), 82–91. https://doi.org/10.1006/ceps.1999.1016 Zimmerman, BJ eng 2000/01/06 Contemp Educ Psychol. 2000 Jan;25(1):82-91. doi: 10.1006/ceps.1999.1016..