Self-Regulated Learning and Science Reading of Middle School Students

Effat Farhana
North Carolina State University
Raleigh, North Carolina, USA
efarhan@ncsu.edu

Teomara Rutherford
University of Delaware
Newark, Delaware, USA
teomara@udel.edu

Collin F. Lynch
North Carolina State University
Raleigh, North Carolina, USA
cflynch@ncsu.edu

ABSTRACT

The role of self-regulated learning (SRL) behaviors for reading scientific texts has been largely recognized by researchers. Unfortunately, not all learners are effectively self-regulating. To provide effective support for SRL activities, it is necessary for us to understand how students adapt their self regulation behaviors during reading. This study investigates students' SRL behaviors in science reading using historical data from a K-12 digital reading platform, Actively learn (AL). We analyze reading related SRL in four contexts, such as, domain-specific sequential pattern, question features, question and content difficulty, and teachers' interaction with the platform. We present findings of our work and seek advice on how the insight that we get from these findings can be used in our proposed methodology.

1. INTRODUCTION

Scientific literacy has been a central goal of international science education reforms for last decades, and researchers consider reading science texts as an integral part of science literacy [9]. Despite the importance of reading comprehension, students in the US lack reading proficiency. According to National Assessment of Educational Progress (NAEP) 2019 report, 37% 8th-graders in the US performed at or above NEAP reading proficiency level [1] and this number is lower than that of 2017. An integral skill for reading is self regulated learning (SRL) [14]. Unfortunately, the typical teacher/student ratio and teachers' priority for topic completion make it difficult for students to learn and practice reading skills and other SRL skills.

Digital reading platforms can provide opportunities to learn and practice SRL strategies in classroom settings. Retrospective analysis of rich data from digital platforms of can provide insight about students' learning pattern to support tailored interventions by instructors.

The present dissertation proposes four research questions (RQs) to investigate students' reading and reading-related SRL behaviors within the AL platform [2].The RQs are described as follows.

- **RQ1. [SRL Patterns and Performance Difference]** *How do students' score connect with their reading and SRL patterns?*
- **RQ2. [SRLs and Question Features]** *How do reading and SRL strategies vary with question features?*
- **RQ3. [SRLs and Content Difficulty]** *How do reading and SRL strategies vary with question and text difficulty?*
- **RQ4. [SRLs and Teachers' Interaction]** *How do teacher interactions with the system connect with students' reading and SRL behaviors?*

In the following subsections we present subquestions, motivations, and possible contributions associated with each RQ.

1.1 RQ1: SRL Patterns and Performance

We split the RQ1 into a subquestion as follows.
RQ1: Which reading and SRL patterns differ between high and low performing student?
RQ1.1 How reading patterns differ for science and social study?
The motivation of RQ1 is to identify reading and SRL behaviors for productive and unproductive students. Additionally, to understand how these behaviors vary for cross domain subjects. Findings of this RQ can be used to develop recommendation system targeted for specific group of students. Also, data driven analysis will be helpful for teacher to make tailored interventions for students.

1.2 RQ2: SRLs and Question Features

To conduct preliminary experiment, we split RQ 2 as follows:
RQ2: How do students' SRL strategies vary with question features?
RQ2.1 Does the association of SRL vary depending on question formats?
RQ2.2 How do other question feature: placement in the text,

[1] https://nces.ed.gov/nationsreportcard/reading/

[2] https://www.activelylearn.com/

Effat Farhana "Self-Regulated Learning and Science Reading of Middle-School Students" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 766 - 769

length of question stem, standard usage are connected with students SRL behaviors?
The motivation of this RQ is twofold. First, to understand which question features prompted what types of SRLs? Second, to understand how question features predict performance in assignment score? While previous SRL researchers [1,5] focused mainly on question formats (i.e., multiple choice question, short answer, ...), our research will examine more fine grained question features to understand students' SRLs.

1.3 RQ3: SRLs and Content Difficulty

We will analyze RQ 3 into two phases as follows: *RQ3.1 How do SRL strategies vary with question difficulty?*
RQ3.2 How do SRL strategies vary with text complexity?
Previous two RQs do not distinguish difficulty level between question formats. RQ3.1 assesses question difficulty from student interaction data at *class* level. We compared our proposed approach with the IRT [10] approach.

1.4 RQ4: SRLs and Teachers' Interaction

The first three RQs analyze students SRL behavior considering their study pattern and question and text features. This RQ focuses on teachers usage the AL system and how it contributes to students SRL usage. We will focus on several teacher-behavior including: how frequently teachers are giving feedback and what question standards are they assigning to questions.

2. METHODS AND CURRENT PROGRESS

Currently, the analysis of the first research question, RQ1 is complete (accepted), a subquestion of the second research question, RQ2.1 is complete (accepted), and a subquestion of the third research question, RQ3.1 is under revision. As RQ3.1 is under submission, we present methodology and results of RQ1 and RQ2.1.

2.1 SRLs in AL

Our scope of this study is evaluating students' SRL usage in middle school science reading within the AL platform. AL reading assignments follow Next Generation Science Standards (NGSS) and have *text embedded* questions. Question formats can be multiple choice (MCQs) and short answer questions (SA) (i.e., fill in the blank and free texts). Questions are graded on a [0-4] scale. The platform's developers claim the platform promotes deep learning by close reading: annotating, highlighting, and engaging with text.
We identify three reading support features of AL as SRL: annotating [8], highlighting [13], and vocabulary lookups, as we believe these features serve as proxies for SRL behaviors. Science text involves concept words and vocabulary terms. Students' reading comprehension and motivation has been decreased due to introduction of concept words [4]. Vocabulary lookups help students to understand concepts when they come across new vocabularies. Annotating requires students comprehend text and write down in their own words [8]. Azevedo described taking notes, summarization, and reading notes in context of SRL strategies for science learning with hypermedia [2]. To select texts for highlighting, students monitor information and connect those to their prior knowledge [13].

2.2 Methodology of RQ1

We describe clustering approach followed by generating sequences, and applying differential sequence mining technique with 12,566 science and 16,240 social study student assignment data.

Clustering Students by Performance Score We calculated four types of scores for each MCQ and SA, resulting in eight performance features. These are: first attempt score, last attempt score, *Norm_last*, and *Long Submission*, *Norm_last* is the multiplication of last score by normalized attempts –the ratio of attempts a student's attempt to all students' attempts on that question in a class. *Long Submission* computes proportion of attempts a student made after the median time for all students on that question in a class. After observing the Silhouette width, we applied K-means clustering with K= 4 on both science and social study data .

Coding Student Actions Student activities in the AL are attempts on question answering and SRL. We codded following question answering first attempts of MCQ (M) and SA (S) and resubmissions of MCQ (m) and SA (s). SRL activities are a reading (R), annotating (A), a highlighting (H), and a vocabulary lookup (V). As the AL system does not record student sessions, we relied on a data-driven approach to identify sessions as described by Kovanovic et al. [7] and Adithya et al. [12]. We plotted histograms of time intervals between consecutive actions to identify last action of any time period. Based upon this analysis we chose a cutoff of 30 minutes as a *session* duration. We split all student activities within a single assignment by *session*. We compacted repeated events by + as done by Kinnenbrew et al. [6].

Frequent Patterns within Clusters Within each cluster, we applied the n-gram sequencing technique and include patterns containing at least one letter from the set {R, A, V, H}. Differential sequence mining algorithm [6], requires two parameters: s-support (frequency of a pattern *within* a cluster) and i-support (frequency of a pattern *within* one *action sequence*). We applied s-support = 0.5 to filter patterns exhibited by at least half of students within that cluster. Next, we applied the Kruskal-Wallis test to identify if a significant difference existed in the mean i-support value within the groups.

2.3 Methodology of RQ2

2.3.1 Methodology of RQ2.1

We used hierarchical linear models (HLMs) to model the relationship between observed behaviors and performance, with assignment at level one, nested within students (level two), nested within classes (level three). We built three models for three different response variables: overall assignment score, MCQ score, and SA score. The fixed-effect variables were the SRL features and number of questions in assignment; these variables were at *Level 1.* Assignment, student, and class were all modeled as random intercepts.

3. RESULTS

In this section we present results of RQ 1 and RQ 2.1.

3.1 Results of RQ1

Four resulting science clusters with student counts (n) were: SA_sc (SA performers in science, n = 4,474), MC_sc (MCQ performers in science, n= 3114), L_sc (low performers in science, n =2,363), and H_sc (high performer in science, n =

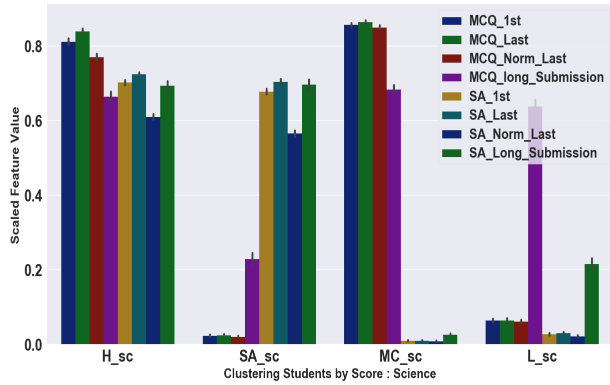

Figure 1: Science Student Clustering by Score

2,636). Similar as in science student clustering, we observe four different groups in social studies: H_ss (n = 8,948) , L_ss (n = 2760), MC_ss (n =2928), and SA_ss(1604). We focused primarily identifying high and low performing student behaviors.

Science Cluster Analysis: Considering H_sc vs L_sc group, two more frequently used patterns describing SA answering behaviors by H_sc students were RS (I-supp_Diff = 0.17, $p < 0.001$) and RS+ (I-supp_Diff = 0.08, $p < 0.001$,). RS and RS+ describe reading prior attempting one (S) or multiple (S+) SAs. Thus, reading prior SA attempt were linked to high performances. H_sc group students also exhibited more annotation behavior than L_sc students (I-supp_Diff = 0.03, $p < 0.001$). Three MCQ attempt related patterns were more exhibited by L_sc group students: RM (I-supp_Diff = -0.16, p <0.001), and V+M (I-supp_Diff = -0.001, $p < 0.05$), and RH+M (I-supp_Diff = -0.002, $p < 0.001$). From Figure 1, we observe L_sc group students have more MCQ_Long_submissions and lower MCQ_Last scores. We conclude L_sc group students struggled in choosing the correct MCQ option.

Social Study Cluster Analysis: Our analysis showed higher-performing students in social study assignments read more frequently before attempting SA and MCQs. Additionally, they looked up more vocabulary. In contrast, low performing students read after attempting SAs. They also had higher resubmission rate of SA questions followed by read event. Our observed patterns explain the way high and low performing students navigated the SA questions. We conclude reading and looking up vocabularies for comprehending the concept prior answering a SA led to score differences for social study subject.

Differential Sequence Mining: Science vs Social Study: We begin with our results for the H_sc vs H_ss comparison. Science students exhibited reading behavior *after* SA submissions compared to social studies: SR (I-sup_Diff = 0.16, p <0.001), S+R (I-sup_Diff = 0.12, $p < 0.001$).
Examining the descriptive statistics, we noticed the mean SA score is higher in social study assignments (SA First = 2.56, SA Last = 2.62) compared to science (SA First = 2.46, SA Last = 2.58) ones. Additionally, mean MCQ scores of science is higher (MCQ first = 2.80, MCQ Last = 2.89) than those of social study (MCQ First = 2.17, MCQ Last = 2.19). Thus, we compared MC_sc vs MC_ss and SA_sc vs SA_ss group. The relatively lower mean SA score in science can be explained by SR (I-sup_Diff = 0.16, $p < 0.001$) and S+R (I-sup_Diff = 0.14, $p < 0.001$). Analyzing MC_sc vs MC_ss group, students with science assignments exhibited more reading behavior before attempting MCQ as described by pattern R+M (I-sup-_Diff = 0.0192, p <0.001). Although the two subject domains are different, our analysis shows reading prior attempting a question associated with higher score in both domains.

3.2 Results of RQ2.1

Table 1: Results from HLM Measuring Association between SRL and Science Score

L1 Level (Assignment)	*β*	***B***	***SE***	***p***
Overall Score				
Intercept		6.533	0.402	<0.001
A	0.055	0.582	0.062	<0.001
H	0.028	0.492	0.072	<0.001
V	0.021	0.275	0.055	<0.001
MCQ Score				
Intercept		5.510	0.369	<0.001
A	0.024	0.206	0.038	<0.001
H	0.016	0.228	0.045	<0.001
V	-0.003	-0.036	0.031	0.259
SA Score				
Intercept		1.699	0.232	<0.001
A	0.040	0.271	0.038	<0.001
H	0.019	0.210	0.043	<0.001
V	0.036	0.289	0.035	<0.001

We report standardized effect size using the formula ß = $(B * SD_x)/SD_y$ (see e.g., [11]). Table 1 presents our findings. All SRL-related variables had positive and statistically significant association with overall science score. Considering question format, the predictive power of note taking was highest (B = 0.271, ß = 0.041, p < 0.001) followed by highlighting (B = 0.210, ß = 0.019, p < 0.001), and vocabulary lookups (B = 0.289, ß = 0.036, p < 0.001). Considering MCQ format, all but the vocabulary lookups continued to be statistically significant positive predictors of MCQ score.

4. FUTURE WORK AND ADVICE SOUGHT

Proposed Methodology of RQ 2.2 We will use multi-task learning to predict common SRL behavior of students for each question (considering question features) and performance on the question. Thus, we will be able to identify students who need help.

Proposed Methodology of RQ 3.2 We will analyze text readability and complexity including lexical, semantic, and argumentation of the text and SRL usage. To analyze readability of science text, we will examine Coh-Metrix [3] and Python's readability package [3]. Additionally, we will examine the argumentation analysis in SA response, particularly questions asking for reasoning, e.g. *Why*, *How*, and *Explain*.

Proposed Methodology of RQ 4 To answer RQ 4, we will perform exploratory analysis to answer the sub questions and calculate association with students' SRL behaviors.

A key limitation of our analysis is, we do not know many confounding variables such as, how teachers used AL assignments (in-class, homework assignment, or extra reading), demographic of students, and how they were using SRLs (i.e., teacher might instruct to take notes). Thus, we seek advice on following aspects:

[3]https://pypi.org/project/readability/

- Is our proposed method of RQ 2.2 generalizable to other context, considering the limitation of our study? The motivation of RQ 2.2 is to provide data-driven recommendation to researchers and educators.

- Beyond my proposed methodology, what other analysis could be more beneficial to understand students' SRL strategies in science reading?

5. REFERENCES

[1] S. Agrawal, G. R. Norman, and K. W. Eva. Influences on medical students' self-regulated learning after test completion. *Medical education*, 46(3):326–335, 2012.

[2] R. Azevedo. The role of self-regulated learning about science with hypermedia. *Recent innovations in educational technology that facilitate student learning*, pages 127–156, 2008.

[3] S. A. Crossley, K. Kyle, and D. S. McNamara. The tool for the automatic analysis of text cohesion (taaco): Automatic assessment of local, global, and text cohesion. *Behavior research methods*, 48(4):1227–1237, 2016.

[4] Y.-S. Hsu, M.-H. Yen, W.-H. Chang, C.-Y. Wang, and S. Chen. Content analysis of 1998–2012 empirical studies in science reading using a self-regulated learning lens. *International Journal of Science and Mathematics Education*, 14(1):1–27, 2016.

[5] S. Jordan. Student engagement with assessment and feedback: some lessons from short-answer free-text e-assessment questions. *Computers & Education*, 58(2):818–834, 2012.

[6] J. S. Kinnebrew and G. Biswas. Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. *International Educational Data Mining Society*, 2012.

[7] V. Kovanović, D. Gašević, S. Dawson, S. Joksimović, R. S. Baker, and M. Hatala. Penetrating the black box of time-on-task estimation. In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 184–193, 2015.

[8] T. Makany, J. Kemp, and I. E. Dror. Optimising the use of note-taking as an external cognitive aid for increasing learning. *British Journal of Educational Technology*, 40(4):619–635, 2009.

[9] S. P. Norris and L. M. Phillips. How literacy in its fundamental sense is central to scientific literacy. *Science education*, 87(2):224–240, 2003.

[10] G. Rasch. Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests. 1960.

[11] T. Rutherford, J. J. Long, and G. Farkas. Teacher value for professional development, self-efficacy, and student outcomes within a digital mathematics intervention. *Contemporary educational psychology*, 51:22–36, 2017.

[12] A. Sheshadri, N. Gitinabard, C. F. Lynch, T. Barnes, and S. Heckman. Predicting student performance based on online study habits: A study of blended courses. *International Educational Data Mining Society*, 2018.

[13] P. H. Winne, J. C. Nesbit, I. Ram, Z. Marzouk, J. Vytasek, D. Samadi, and J. Stewart. Tracing metacognition by highlighting and tagging to predict recall and transfer. *AERA Online Paper Repository*, 2017.

[14] B. J. Zimmerman. Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1):3–17, 1990.