Feedback and Self-Regulated Learning in Science Reading

Effat Farhana[1], Andrew Potter[2], Teomara Rutherford[2], Collin F. Lynch[1]
[1]North Carolina State University, [2]University of Delaware
efarhan@ncsu.edu, ahpotter@udel.edu, teomara@udel.edu, cflynch@ncsu.edu

ABSTRACT
How do students respond to feedback in a reading platform? In this study we examined students' (n = 670) reading and SRL behaviors after receiving feedback from their teachers. First, we examined the extent in which students revised their responses after receiving feedback. Second, we examined the association of reading and SRL behaviors with student scores after feedback. Third, we examined relationships between the type of feedback received (i.e. teacher comments) and subsequent student behaviors. We found that students who revised their answers more had greater score improvements. Teacher feedback in writing conventions was shown to produce fewer reading and SRL behaviors when compared to other types of feedback. The number of reading events was correlated with improved scores, although the effect size was small. These findings suggest that teacher feedback can help students employ reading and SRL behaviors and improve their reading comprehension under the right conditions. We discuss recommendations and possible design implications for online reading platforms.

Keywords
Self-Regulated Learning, Feedback, Sequence mining, Reading comprehension, Natural Language Processing

1. INTRODUCTION
Feedback can improve students' performance [38] and Self-Regulated Learning (SRL) behaviors [10]. However, students must understand feedback in order to apply it [48]. Feedback gaps occur when students receive but do not act upon feedback [24], and may be caused by lack of clarity [9], students' misunderstanding of feedback application [55], and the feedback paradox [61] (i.e., students do not address feedback despite understanding its importance). Researchers have recently emphasized the actionability of feedback as one factor to change students' actions and behaviors [12]. This concept remains largely underexplored [34].

To address the feedback actionability gap, researchers have analyzed how students act upon receiving feedback by examining students' perceptions [37, 50] and analyzing student behavior, including timely response to feedback [34], the effect of different types of feedback on the same question [27], and students' learning strategies usage [43].

We examine students' feedback response behavior in science reading. Science reading skills are of critical importance, but challenging for students to master [63]. Science reading can be enhanced through the application of SRL skill [15, 47]. To investigate SRL and science reading, we conducted our analysis on middle school science readings and questions from an online learning platform, Actively Learn (AL). We answered three research questions:

RQ1. How do students' scores vary after receiving feedback?
- **RQ1.1**. To what extent do students change their answers after receiving feedback?

RQ2. How does students' reading and SRL usage vary upon receiving feedback?
RQ3. Is feedback type associated with subsequent reading and SRL behaviors?

2. RELATED WORK
2.1 SRL and Reading
SRL refers to four regulatory processes during learning: goal setting, self-monitoring, self-evaluating, and applying strategies [65]. Self-regulated learners use self-monitoring skills to monitor their tasks [69] and can judge their learning outcomes in light of their goals [68]. Self-regulation is associated with academic performance [49, 66]. SRL researchers have proposed theories and models (e.g., Pintrich's SRL framework [49], Zimmerman's Cyclic model [66]) to explain learners' SRL behaviors. In this study, we adopt Winne and Hadwin's model [56, 58] to measure students SRLs from students' log trace data within AL, as it has proven a useful framework for similar research [4].

SRL-based reading interventions have been effective in improving middle school reading in experimental studies [54]. Computer-based learning environments (CBLEs) can integrate SRL instruction via features to help students foster SRL skills. Examples of CBLEs that are rooted in models of SRL and have been shown to support reading comprehension and SRL behaviors for reading include iSTART [44-45], nSTUDY [4], and ReaderBench [18-19]. In this study, we examine the web-based platform Actively Learn (AL), which uses platform features that promote SRL (Section 3.1).

2.2 Sequence Mining
Sequence mining techniques can identify students' learning behaviors [2, 29]. For example, *n*-gram sequencing techniques have been applied to a game-based learning platform to identify students' problem-solving behavior [2] and to study associations between students' academic performance and transition behavior among multiple platforms [29].
In this study, we are focused on what SRL activities students engage in on the AL platform *after* they receive feedback on a prior submission. In this analysis we applied an approach used by Sheshadri et al. [52] to examine sequence behaviors across platforms. In this approach we aggregated distinct SRL and question submission actions within AL and then examined the frequency and sequence of the activities prior to a resubmission.

2.3 Feedback

Providing feedback and opportunities for students to respond to feedback is one way teachers can assist their students in reading to learn tasks in STEM domains [42]. However, feedback quality can influence students' responses [53]. Hattie and Timperley [31] characterized feedback at four levels: the task (i.e., how well the student accomplished a task), processing (i.e., the processes required to complete the task), self-regulation (i.e., how students choose and implement self-regulatory strategies to accomplish a task); and the self-level (i.e., personal evaluations). Feedback effectiveness is also moderated by the amount of information provided; different types of feedback should be considered as separate constructs [60]. Prior studies have also shown that timely engagement with personalized feedback was associated with academic success for undergraduate students [34] and may also prompt more engaged learning activities when compared to general feedback [43].

Corrective and self-regulatory feedback given to students after answering comprehension questions in response to texts in a digital environment can enhance SRL behaviors and performance [39, 40]. In an experimental study, students who received self-regulatory feedback made more text searches and included more textual info in their responses when compared to students who received less informative or no feedback [39]. A follow-up study replicated these findings and also demonstrated that requiring students to select relevant text information before re-submitting answers led to improved SRL behaviors [40]. Taken together, these studies suggest that corrective and self-regulatory feedback can improve SRL behaviors and reading performance when students are tasked with re-submitting answers to comprehension questions with digital texts.

However, it can be challenging for teachers to provide timely and informative feedback at scale [31, 53]. A prior study on feedback comments of science assignments [8] indicated that feedback that did not provide a correct answer was only helpful if students knew where to find the correct answer; more informative feedback was required when students lacked background knowledge Prior research also suggests that timely engagement with feedback, particularly personalized feedback, was associated with academic success for undergraduate students [34]. Written comments can provide an effective means for providing feedback on science content [8] and in digital reading comprehension tasks [39, 40]. In this study we examined how teacher feedback comments within the context of a digital science reading comprehension related to students' SRL behaviors.

3. Actively Learn (AL) Platform

AL is an online K-12 reading platform for multiple disciplinary subjects. AL catalogs curriculum-integrated readings that teachers can assign as in-class or homework assignments. Teachers can also add their own content as assignments. AL assignments contain text-embedded questions that can be multiple choice (MCQ) and short-answer (SA), including open ended questions and fill-in-the blanks. Teachers can give feedback on students' answers to questions by scoring questions on a scale from 0-4 and writing comments.

We adopted Winne and Hadwin's SRL model in our study. Winne and Hadwin's model has four phases: task defining (**Phase 1**), goal setting (**Phase 2**), enacting tactics and strategies (**Phase 3**), and metacognitively adopting strategies (**Phase 4**). We primarily focus on students' usage of SRL tactics/strategies within AL (**Phase 3)** and adapting reading and SRL (**Phase 4)** upon receiving feedback. Our study is grounded in the Winne and Hadwin model, as its focus on the events underlying SRL [57] fits the retrospective analysis of student interaction data within our study. Furthermore, we focus on three types of SRL events that are consistent with prior literature situated in Winne and Hadwin's model: annotating [3, 41], highlighting [59], and vocabulary lookups [5].

3.1 Dataset Preparation

The present study was conducted with middle school physical science data collected from AL in 2018. The initial dataset contained 17,886 student records from 1,033 classes. First, after data cleaning, we included classes containing 10-60 students (n = 14,925 students). Second, we identified student submissions on which they received feedback. This reduced dataset included 1,819 unique students, 3,867 questions, and 5,373 submissions. Third, we applied the following filtering criteria: 1) a student submitted a question multiple times, 2) received at least one instance of feedback, and 3) re-submitted after receiving feedback. The trimmed dataset, which included student empty submissions, contained 670 unique students in 113 classes, 58 teachers, 156 assignments, 1,072 questions, and 2,502 submissions. All questions in our dataset are SA questions.

4. METHODOLOGY

We describe our methodology for each RQ in this section.

4.1 RQ1 Methodology

To answer RQ1 we measured students' score differences by calculating the difference between the first and last submission scores when students made multiple attempts after receiving feedback. We observed three categories of submissions: score increased in the last submission, score decreased in the last submission, and score was unchanged in the last submission.

To assess whether students were addressing teachers' feedback, we calculated similarity between subsequent answer submissions of a question. We hypothesized that changes in submitted answers would result in a greater score difference in a question. Thus, we measured the cosine similarity between subsequent submissions of a question. Specifically, we calculated cosine similarities between ith and (i-1) th submissions, for i => 2 attempts and took the average. We took all submissions because we wanted to assess how students' changed their answers upon receiving feedback, and how those changes impacted their final scores.

To encode students' responses into vector representations, we used the Universal Sentence Encoder (USE) [15]. The USE can take a word, a sentence, or a paragraph as inputs and encodes into a fixed length vector of 512 values. We then used a Deep Averaging Network (DAN) model with USE to encode questions and question-dependent texts into vectors [15]. DAN averages unigrams and bi-grams of word embedding to construct sentence embedding. Moreover, to evaluate how the answer modifications were connected to score differences, we calculated Spearman correlation between mean cosine similarities and score difference.

4.2 RQ2 Methodology

To answer RQ2, we coded student actions within the AL system as either an answer submission, reading event (R), or SRL event, such as annotating (A), highlighting (H), or vocabulary-lookup (V). The AL system does not define explicit student sessions.

Therefore, we adopted a data-driven approach from prior research to define session [36, 52]. First, we aggregated students' assignment actions and timestamps into a unified transaction log. We then plotted histograms of two consecutive action sequences to estimate the intervals between consecutive actions within an assignment. Based on our exploratory analysis we selected 30 minutes as a "session" cutoff. Any student actions exceeding 30 minutes were defined as a new session. After defining session cutoffs, we split all student actions within an assignment by session. Next, we counted SRL events ***before*** a student's resubmission of the question received feedback.

We then applied a four-level hierarchical linear model (HLM) to predict the last score of a question. HLM is commonly used in educational research [24, 50] to account for nested data [62]. Our HLM model included questions at level-one, assignment ID at level-two, student ID at level-three, and teacher ID at level-four. Fixed effect variables included students' first score on questions and features of SRL usage during attempting questions. All grouping variables were modelled as random intercepts.

4.3 RQ3 Methodology

To answer RQ3, we categorized teacher feedback comments using deductive analysis, which is a method for analyzing content using a predefined model based on prior research [23]. Our deductive analysis categories were adapted from Hattie and Timperley's [31] feedback categories, which have been used in prior research [1, 30]. Our model was also influenced by Shute's [53] review of formative feedback, and Bruno and Santos' [8] combined inductive and deductive coding scheme of teacher written comments in a science classroom context for task and processing-level feedback. We established five a-priori feedback categories using these models. The feedback categories included feedback on the task and processing [31] that (i) asked a student to either provide a correction to a response or to (ii) provide an explanation of a response [8], (iii) self-level feedback, (iv) and SRL behaviors. We also created a category for feedback that only addressed (v) conventions (e.g., spelling, grammar).

The SRL behavior category included teacher comments that referred to the SRL reading behaviors described in the previous section. SRL feedback has been defined as high-information feedback about task performance and suggestions for employing self-regulation strategies to monitor cognitive processes, self-evaluate performance, and strategy development to improve performance [31, 60]. We defined SRL feedback more broadly to include teacher comments that provided feedback on referring to the text to make revisions to an answer, annotating or highlighting the text, or to look up a vocabulary term. This definition is more appropriate within the context of AL, in which teachers leave brief comments on comprehension questions. Prior research has also defined SRL feedback in this context as feedback that includes knowledge about when to refer to the text [39] and which text information is relevant for completing the task [40].

Two members of the research team trained on coding comments using a sample. All differences in training were resolved by discussion. One researcher then coded all teacher feedback comments (n = 1,441). A second researcher independently coded 23% of this sample. Inter-rater reliability (IRR) was calculated using Cohen's kappa and was found to be acceptable ($\kappa = 0.74$, $p < 0.001$). We then applied a nonparametric Kruskal-Wallis test to identify if reading and SRL behaviors varied significantly among feedback categories.

5. RESULTS

In the following subsections we discuss our results for each RQ.

5.1 RQ1 Results

We calculated the average cosine similarities between subsequent submissions (sim_score) and score difference (d) with and without empty student submissions. A higher sim_score indicates that the submitted answers are more similar to each other. The frequencies of six different score difference categories and question counts (n) are: **-2** (n = 4), **-1** (n = 53), **0** (n = 187), **1** (n = 474), **2** (n = 252), **3** (n = 87), and **4** (n = 15). Total questions = 1,072. The Spearman correlation test between score difference (d) and mean cosine similarities (sim_score) was (coefficient = -0.315, $p < 0.001$). The negative coefficient indicates when the mean cosine similarity score decreases, the score difference increases. In other words, the more changes are present in students' subsequent answers, the greater the score difference.

Score Increased Descriptive statistics in this category are: 828 unique questions, 1,963 submissions by 543 students. First attempt score ranged from 0 to 3 with a mean 1.41. Last attempt score varied from 1 to 4 with a mean 2.98. We found that the positive score change groups have increased by 1, 2, 3, and 4 points. In these four groups, sim_score has a lower median value compared to the rest. This observation indicates that students with greater score increases had submissions that differed more than their original answers, as represented by a lower sim_score. We examined student submissions with identical responses (sim_score = 1) but an increase in final score (n = 40) submissions.

Score Decreased. This group includes 57 unique questions with 124 submissions by 49 students. First attempt scores ranged from 1 to 4 with a mean 1.58. Last attempt score varied from 0 to 3 with a mean of 0.51.

Score Unchanged. Descriptive statistics for this category are: 187 unique questions, 415 submissions by 165 students. First and last attempt scores have the same statistics in this category. First and last attempt scores ranged from 0 to 4 with a mean 1.69.

5.2 RQ2 Results

Standardized effect sizes were calculated using the formula, $\beta = (B*SDx)/(SDy)$ [50]. First attempt score had the highest predictive power ($B = 0.32$, $\beta = 0.28$, $p < 0.001$). Only reading was a statistically significant positive predictor. Highlighting behavior was negatively associated with the last score.

5.3 RQ3 Results

Feedback comments were coded as requiring either a correction (n = 654), explanation (n = 565), a SRL behavior (n = 134), addressing conventions (n = 77), and self-level feedback (n = 11). SRL events after receiving feedback and before resubmission were identified. Kruskal-Wallis test results indicated statistically significant differences across the five feedback categories for reading ($p < 0.001$), highlighting ($p = 0.007$), and vocabulary lookup events ($p = 0.006$). Annotating text was not significant.

We conducted post-hoc analyses using Dunn's pairwise tests with Benjamini-Hochberg correction for features with statistically significant results. Effect size (r) is reported using a nonparametric test, Cliff's-Delta. Results indicated that students were less likely to engage in reading after conventions feedback when compared to SRL behavior feedback ($p < 0.001$, r = 0.25), corrective ($p < 0.001$, r = 0.28), and explanation feedback ($p < 0.001$, r = 0.29). The Kruskal-Wallis test assessed whether the group with non-zero entry (i.e., SRL Behavior) was statistically different from the ones with all zero entries. We found statistically significant differences between SRL Behavior and corrective feedback (p = 0.002, r = 0.009) and explanation feedback (p = 0.001, r = 0.009). Students were more likely to look up vocabulary words after corrective over explanation feedback (p = 0.003, r = 0.021).

6. DISCUSSION and CONTRIBUTIONS

Scholarly Implication: Student Response to Feedback

RQ1 results show that students' who modified their answers had greater score differences, which is consistent with prior findings on automated feedback [39, 40, 64]. We found that teachers at times scored revised responses lower than students' initial score. When examining students' responses, we found students sometimes submitted the identical answer or an empty answer ("No response") despite the teacher asking for explanation or suggesting additional correction. This phenomenon in which students do not address teacher feedback is known as the ***"feedback gap"*** [24]. Students might not respond to feedback if they find it difficult to decipher [9], lack study habits [20], or erroneously believe it does not apply to them [28]. One limitation of the present study is that it is not equipped to determine the reason for lack of student response.

Our HLM analysis from RQ2 shows that reading events and initial scores were statistically significant predictors of last scores. However, SRL variables such as annotation, highlighting, and vocabulary lookups were not statistically significant predictors. We also found that highlighting was underutilized by students and that self-level feedback was not commonly employed by teachers.

Feedback comments categorized as focusing on correction, explanation, and SRL behaviors were associated with more reading events during student revisions when compared to feedback about conventions. We expected SRL feedback to produce more reading events and SRL behaviors than other categories based on prior research with automated feedback, because these comments directed students to revisit the text to revise their answers [39, 40]. However, SRL feedback did not produce statistically significant differences in student behaviors compared to correction and explanation feedback. One reason for this finding might be that these feedback categories had similar amounts of information; the level of feedback informativeness may have a greater impact on student performance and behavior [60]. Corrective (e.g., "Protons cannot be gained or lost") and explanation feedback provided did not explicitly direct students to revisit the text or use an AL feature, but perhaps these behaviors were implied perhaps these behaviors were implied during a task-oriented reading assignment with explanation feedback comments such as: "Great definitions but you need to explain why phase changes are considered physical changes.". This might explain why vocabulary look-ups were more common in corrective and explanation comments when compared to SRL (e.g., "Go back and reread paragraph 9 and reanswer. Might help you to plug some numbers into the equation to see how the inverse relationship works.") and conventions feedback ("Capitalize the first word in a sentence."). Conversely, perhaps the SRL feedback could more effectively influence reading events and SRL behaviors if teachers provided more explicit information that helped students decide when and how to revisit the text to revise an answer [39] or required students to select relevant information from the text to support their answer [40]. SRL feedback may have directed students to relevant portions of the text based on relatively greater highlighting behavior after SRL feedback, but this effect size was small, and highlighting was not positively related to score change, calling into question the value of this behavior.

For Teachers: Feedback Quality

Our analysis also showed that teacher comments were generally short and contained limited information. It may be possible for teachers to improve the quality and effectiveness of their feedback by providing more SRL feedback [60, 31], and by avoiding self-level feedback and comments about conventions, which were shown to not support student performance in the present study. To optimize feedback from teacher comments and increase student feedback uptake, teachers should support students in understanding feedback comments and evaluation criteria [13], which may require greater elaboration within comments and potentially instruction outside of AL.

Design Implications: Automated Feedback Affordances

Feedback can improve performance [38], but poor feedback can hinder student learning [31]. Middle school teachers may not have enough time to provide quality and timely feedback to all students, particularly when providing feedback to open-ended questions that require source-based explanations [6]. Although automated feedback may assist teachers in providing quality and timely feedback without increasing their workload [6, 14], challenges remain in building platforms to provide such feedback within the context of source-based science questions. For example, AL science questions are often constructed responses that require connecting information from different paragraphs. The state-of-the-art NLP research to automatically infer information from paragraphs in reading comprehension is still in early stage [22, 35]. One possible design for automated support would be to collect other teachers' feedback on the same question in the AL platform and provide suggestions to the teacher.

Design Implication: Supporting Feedback Actionability

Some students were not responsive to feedback as indicated by their submission of an empty answer or re-submission the same answer. One solution to increase actionability could be pointing to additional learning materials in an automated feedback setting [33]. For example, Broos and colleagues [7] designed a button **"*Okay, what now?*"** in a dashboard to provide actionable feedback. Students could click the button to view extra reading content. Similarly, a nudge can be implemented in AL—***"Are you sure you want to submit that empty answer?"***

7. CONCLUSIONS

This study has two main contributions to reading and SRL research: (i) empirically evaluating students' response changes to short answer questions upon receiving feedback and (ii) measuring the association of students' reading and SRL with five feedback categories. Our findings show that students who revised their answers demonstrated statistically significant differences in their scores. We also observed that teachers mainly provided corrective feedback followed by explanatory and feedback related to SRL behavior. Students exhibited more reading behavior upon receiving these types of feedback than convention-related feedback. These results may aid educators in writing feedback comments to students for maximal impact.

8. ACKNOWLEDGEMENTS

This research was supported by NSF #1821475 ``Concert: Coordinating Educational Interactions for Student Engagement'' Collin F. Lynch, Tiffany Barnes, and Sarah Heckman (Co-PIs).

9. REFERENCES

[1] Rola Ajjawi and David Boud. 2017. Researching feedback dialogue: An interactional analysis approach. Assessment & Evaluation in Higher Education, 42, 2, 252-265. DOI:https://doi.org/10.1080/02602938.2015.1102863

[2] Bita Akram, Wookhee Min, Eric Wiebe, Bradford Mott, Kristy Elizabeth Boyer, and James Lester. 2018. Improving Stealth Assessment in Game-Based Learning with LSTM-Based Analytics. International Educational Data Mining Society.

[3] Roger Azevedo. 2008. The role of self-regulated learning about science with hypermedia. Recent innovations in educational technology that facilitate student learning (2008): 127-156.

[4] L. P. Beaudoin, and P. Winne. 2009. nStudy: An Internet tool to support learning, collaboration and researching learning strategies. In Canadian e-Learning Conference

[5] Andrew Biemiller, and Naomi Slonim. 2001. Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. Journal of educational psychology 93, 3, 498.

[6] Anthony F. Botelho and Neil T. Heffernan. 2019. CROWDSOURCING CHAPTER 11–FEEDBACK TO SUPPORT TEACHERS AND STUDENTS. Design Recommendations for Intelligent Tutoring Systems: Volume 7-Self-Improving Systems, 101.

[7] Tom Broos, Laurie Peeters, Katrien Verbert, Carolien Van Soom, Greet Langie, and Tinne De Laet. 2017. Dashboard for Actionable Feedback on Learning Skills: Scalability and Usefulness. In Learning and Collaboration Technologies. Technology in Education (Lecture Notes in Computer Science), Panayiotis Zaphiris and AndriIoannou (Eds.). Springer International Publishing, Vancouver, Canada, 229–241.

[8] Inês Bruno and Leonor Santos. 2010. Written comments as a form of feedback. Studies in Educational Evaluation, 36, 3, 111–120. DOI:http://dx.doi.org/10.1016/j.stueduc.2010.12.001

[9] Deirdre Burke. 2009. Strategies for using feedback students bring to higher education. Assessment & Evaluation in Higher Education 34, 1, 41-50.

[10] Deborah L. Butler, and Philip H. Winne. 1995. Feedback and self-regulated learning: A theoretical synthesis. Review of educational research 65, 3,: 245-281.

[11] Therese Bouffard-Bouchard, Sophie Parent, and Serge Larivee. 1991. Influence of self-efficacy on self-regulation and performance among junior and senior high-school age students. International journal of behavioral development 14, 2, 153-164.

[12] David Carless and David Boud. 2018. The development of student feedback literacy:enabling uptake of feedback. Assessment & Evaluation in Higher Education 43, 8 (2018), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

[13] David Carless and Naomi Winstone. 2020. Teacher feedback literacy and its interplay with student feedback literacy. Teaching in Higher Education (2020): 1-14.

[14] Anderson Pinheiro Cavalcanti, Arthur Diego, Rafael Ferreira Mello, Katerina Mangaroska, André Nascimento, Fred Freitas, and Dragan Gašević. 2020. How good is my feedback? a content analysis of written feedback. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, 428-437.

[15] Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant et al. 2018. Universal Sentence Encoder for English. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 169-174. 2018

[16] Richard Correnti, Lindsay Clare Matsumura, Laura Hamilton, and Elaine Wang. 2013. Assessing students' skills at writing analytically in response to texts. The Elementary School Journal, 114, 2, 142-177.

[17] Jennifer G. Cromley and Roger Azevedo. 2007. Testing and refining the direct and inferential mediation model of reading comprehension. Journal of Educational Psychology 99, 2, 311.

[18] Mihai Dascalu, Philippe Dessus, Ştefan Trausan-Matu, Maryse Bianco, and Aurélie Nardy. 2031. ReaderBench, an environment for analyzing text complexity and reading strategies. International Conference on Artificial Intelligence in Education, pp. 379-388. Springer, Berlin, Heidelberg, 2013.

[19] Mihai Dascalu, Larise L. Stavarache, Stefan Trausan-Matu, Philippe Dessus, Maryse Bianco, and Danielle S. McNamara. 2015. ReaderBench: An integrated tool supporting both individual and collaborative learning." In Proceedings of the fifth international conference on learning analytics and knowledge, pp. 436-437. 2015.

[20] Phillip Dawson, Michael Henderson, Tracii Ryan, Paige Mahoney, David Boud, Michael Phillips, and Elizabeth Molloy. 2018. Technology and feedback design. Learning, design, and technology.

[21] Project DRIVER_SEAT https://www.neilheffernan.net/projects/funded-projects/driver seat

[22] Dheeru Dua, Yizhong Wang, Pradeep Dasigi, Gabriel Stanovsky, Sameer Singh, and Matt Gardner. 2019. DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1 (Long and Short Papers), pp. 2368-2378.

[23] Satu Elo and Helvi Kyngäs. 2008. Qualitative Research and Content Analysis. Journal of Advanced Nursing, 62, 107-115.

[24] Carol Evans. 2013. Making sense of assessment feedback in higher education. Review of educational research 83, 1, 70-120.

[25] Effat Farhana, Teomara Rutherford, and Collin F. Lynch. 2020. Associations Between Self-Regulated Learning Strategies and Science Assignment Score in a Digital Literacy Platform. International Society of the Learning Sciences (ISLS) .

[26] Actively Learn Feedback https://help.activelylearn.com/hc/en-us/articles/115000590154-Revise-answers

[27] Tomer Gal, and Arnon Hershkovitz. 2019. Different Types of Response-Based Feedback in Mathematics: The case of textual and symbolic messages. In Proceedings of the 9th

International Conference on Learning Analytics & Knowledge, pp. 265-269. 2019.

[28] Teresa Garcia. 1995. The Role of Motivational Strategies in Self-Regulated Learning. New Directions for Teaching and Learning 63, 29-42.

[29] Niki Gitinabard, Tiffany Barnes, Sarah Heckman, and Collin F. Lynch. 2019. What Will You Do Next? A Sequence Analysis on the Student Transitions Between Online Platforms in Blended Courses. International Educational Data Mining Society (2019).

[30] Marjan J.B. Govaerts , Margje W.J. van de Wiel , Cees P.M. van der Vleuten. 2013. Quality of Feedback following Performance Assessments: Does assessor expertise matter? European Journal of Training and Development, 37, 105-125. DOI: https://doi.org/10.1108/03090591311293310

[31] John Hattie and Helen Timperley. 2007. The power of feedback. Review of Educational Research, 77, 81-112. DOI:10.3102/00346543029848

[32] Anton Havnes, Kari Smith, Olga Dysthe, and Kristine Ludvigsen. 2012. Formative assessment and feedback: Making learning visible. Studies in Educational Evaluation 38, 21–27. DOI:http://dx.doi.org/10.1016/j.stueduc.2012.04.001

[33] Christothea Herodotou, Sarah Heiser, and Bart Rienties. 2017. Implementing randomised control trials in open and distance learning: a feasibility study. OpenLearning: The Journal of Open, Distance and e-Learning 32, 2, 147–162.

[34] Hamideh Iraj, Anthea Fudge, Margaret Faulkner, Abelardo Pardo, and Vitomir Kovanović. 2020. Understanding students' engagement with personalised feedback messages. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, pp. 438-447. 2020.

[35] Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. Spanbert: Improving pre-training by representing and predicting spans. Transactions of the Association for Computational Linguistics 8 (2020): 64-77.

[36] Vitomir Kovanović, Dragan Gašević, Shane Dawson, Srećko Joksimović, Ryan S. Baker, and Marek Hatala. 2015. Penetrating the black box of time-on-task estimation. In Proceedings of the fifth international conference on learning analytics and knowledge, pp. 184-193. 2015.

[37] Franki YH Kung, and Abigail A. Scholer. 2018. Message Framing Influences Perceptions of Feedback (In) directness. Social Cognition 36, 6, 626-670.

[38] Philip Langer. 2011. The Use of Feedback in Education: A Complex Instructional Strategy. Psychological Reports 109, 3, (December 2011): 775–84.

[39] A. C. Llorens, R. Cerdán and Eduardo Vidal-Abarca. 2014. Adaptive formative feedback to improve strategic search decisions in task-oriented reading. Journal of Computer Assisted Learning, 30, 3, 233–251. DOI:http://dx.doi.org/10.1111/jcal.12050

[40] A. C. Llorens, Eduardo Vidal-Abarca and R. Cerdán. 2016. Formative feedback to transfer self-regulation of task-oriented reading strategies. Journal of Computer Assisted Learning, 32, 4, 314–331. DOI:http://dx.doi.org/10.1111/jcal.12134

[41] Tamas Makany, Jonathan Kemp, and Itiel E. Dror. 2009. Optimising the use of note‐taking as an external cognitive aid for increasing learning. British Journal of Educational Technology 40, 4, 619-635.

[42] Linda H. Mason and Laura R. Hedin. 2011. Reading science text: Challenges for students with learning disabilities and considerations for teachers. Learning Disabilities Research & Practice, 26, 4, 214–222. DOI:http://dx.doi.org/10.1111/j.1540-5826.2011.00342.x

[43] Wannisa Matcha, Dragan Gašević, Nora'Ayu Ahmad Uzir, Jelena Jovanović, and Abelardo Pardo. 2019. Analytics of learning strategies: Associations with academic performance and feedback. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge, pp. 461-470.

[44] Danielle S. McNamara. 2004. SERT: Self-explanation reading training. Discourse processes 38, 1 , 1-30.

[45] Danielle S. McNamara., Irwin B. Levinstein, and Chutima Boonthum. 2004. START: Interactive strategy training for active reading and thinking. Behavior Research Methods, Instruments, & Computers 36, 2, 222-233.

[46] Danielle S. McNamara, Tenaha P. O'Reilly, Rachel M. Best, and Yasuhiro Ozuru. 2006. Improving adolescent students' reading comprehension with iSTART. Journal of Educational Computing Research 34, 2, 147-171.

[47] Danielle S. McNamara. 2007. Reading comprehension strategies: Theories, interventions, and technologies. Psychology Press., 2007.

[48] David Nicol. 2010. From monologue to dialogue: improving written feedback processes in mass higher education. Assessment & Evaluation in Higher Education 35, no. 5 (2010): 501-517.

[49] Paul R. Pintrich. 2000. The role of goal orientation in self-regulated learning. 2000. Handbook of self-regulation, 451-502. Academic Press, 2000

[50] Anna D. Rowe, and Leigh N. Wood. 2008. Student perceptions and preferences for feedback. Asian Social Science 4, 3,: 78-88.

[51] Teomara Rutherford. 2017. Within and between person associations of calibration and achievement. Contemporary Educational Psychology 49, 226-237.

[52] Adithya Sheshadri, Niki Gitinabard, Collin F. Lynch, Tiffany Barnes, and Sarah Heckman. 2018. Predicting student performance based on online study habits: a study of blended courses.International Educational Data Mining Society (2018).

[53] Valerie J. Shute. 2008. Focus on formative feedback. Review of Educational Research, 78, 153–189. DOI:http://dx.doi.org/10.3102/0034654307313795

[54] Tuncay Türkben. 2019. The Effect of Self-Regulation Based Strategic Reading Education on Comprehension, Motivation, and Self-Regulation Skills. International Journal of Progressive Education 15, 4,: 27-46.

[55] Melanie R Weaver. 2006. Do students value feedback? Student perceptions of tutors' written responses. Assessment & Evaluation in Higher Education 31, no. 3 (2006): 379-394.

[56] Philip H.Winne and Allyson F. Hadwin. 1998. Studying as self-regulated learning. The educational psychology series. Metacognition in educational theory and practice (277–304). Lawrence Erlbaum Associates Publishers.

[57] Philip H. Winne and Nancy E. Perry. 2000. Measuring

self-regulated learning. Handbook of self-regulation, 531-566. Academic Press, 2000.

[58] Philip H. Winne and Allyson F. Hadwin. 2012. The weave of motivation and self-regulated learning. 2012. Motivation and self-regulated learning , pp. 309-326. Routledge, 2012.

[59] Philip H. Winne , John Cale Nesbit, Ilana Ram, Zahia Marzouk, Jovita Vytasek, Donya Samadi, and Jason Stewart. 2017. Tracing Metacognition by Highlighting and Tagging to Predict Recall and Transfer. AERA Online Paper Repository (2017).

[60] Benedikt Wisniewski, Klaus Zierer, and John Hattie. 2020. The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10, 1-14. DOI:http://dx.doi.org/10.3389/fpsyg.2019.03087

[61] Carol Withey. 2013. Feedback engagement: forcing feed-forward amongst law students. The Law Teacher 47, no. 3 (2013): 319-344.

[62] Heather Woltman, Andrea Feldstain, J. Christine MacKay, and Meredith Rocchi. 2012. An introduction to hierarchical linear modeling. Tutorials in quantitative methods for psychology 8, 1, 52-69.

[63] Larry D Yore. 2012. Science literacy for all: More than a slogan, logo, or rally flag!. In Issues and challenges in science education research, pp. 5-23. Springer, Dordrecht, 2012.

[64] Mengxiao Zhu, Hee-Sun Lee, Ting Wang, Ou Lydia Liu, Vinetha Belur, and Amy Pallant. 2017. Investigating the impact of automated feedback on students' scientific argumentation. International Journal of Science Education 39, 12, 1648-1668

[65] Barry J. Zimmerman, and Albert Bandura. 1994. Impact of self-regulatory influences on writing course attainment. American educational research journal 31, 4 (1994), 845-862.

[66] Barry J. Zimmermann. 1989. Models of self-regulated learning and academic achievement. Self-Regulated Learning and Academic Achievement: Theory, Research and Practice, 1-26.

[67] Barry J. Zimmermann. 1989. A social cognitive view of self-regulated academic learning. Journal of educational psychology, 81, 3 ,329.

[68] Barry J. Zimmerman. 2000. Self-efficacy: An essential motive to learn. Contemporary educational psychology, 25, 1, 82–91

[69] Barry J. Zimmerman. 2000. Attaining self-regulation: A social cognitive perspective. Handbook of self-regulation, 13-39. Academic Press, 2000.