WIP: Analyzing Students' Practices Behaviors in an Introductory Computer Science Course and Monitoring Their Practices Behaviors in a Subsequent Class

Marcia C. Moraes
Colorado State University
Computer Science Department
Fort Collins, USA
marcia.moraes@colostate.edu

James E. Folkestad
Colorado State University
School of Education
Fort Collins, USA
james.folkestad@colostate.edu

Abstract*—This innovative practice WIP aims to analyze students' practice behavior and its impacts on students' performance and retention. Studies have shown that memories weaken over time, and the most significant drop in retention happens just a day after you learn something new. Ebbinghaus's forgetting curve suggests that people tend to forget newly learned materials in days or weeks when there is no attempt to review the learned materials. To prevent forgetting, we need to reinforce what we learn by using techniques such as spaced retrieval practice and interleaving. We did that in an Introductory Computer Science Course (CS1). We used quizzes as low-stakes retrieval practice activities (RPAs) that students could take as many times as they want during the semester, along with the U-Behavior teaching and learning method. The U-Behavior method is composed of an application that generates personalized visualizations of student's study behaviors and self-reflections to help students improve their study behaviors to reinforce long-term learning. In a previous study, with data from a Fall 2021 CS1 course, we found a statistically significant increase in final exam scores, final coding exam, and final course grades for students who practice desirable behaviors, spacing and mixing their practice over the semester, compared with students who did not. This study presents a work in progress that uses data from Fall 2023 CS1 course to analyze if the increase in grades persists in a different cohort of students for the same course as well as to extend the previous study by monitoring students practice behaviors in a subsequent class, CS2, that is being taught this Spring 2024 semester. Of 94 students registered in CS1 Fall 2023, a total of 56 consented to participate in the study. Of the 56 students, 18 demonstrated desirable behaviors by spacing and interleaving their practice during the semester, and 38 did not. Initial analysis of Fall 2023 CS1 data showed a statistically significant increase on final exam and final course grades for students who practice desirable study behaviors. Contrary to Fall 2021 CS1 data, no statistical significance was found regarding final coding exam grades for Fall 2023 CS1 students between students who practice desirable behaviors and those who didn't. We analyzed if students who demonstrated desirable practice behaviors are continuing practice in a spaced and interleaved in the subsequent class, CS2. Initial findings and future directions are presented.

***Keywords*—higher education, undergraduate, first year experience, knowledge retention, CS1**

I. Introduction

There are three recognized stages in learning including: encoding, storage, and retrieval. Encoding transforms information into a form that can be stored in memory, creating a memory trace. Storage maintains the encoded information in memory, and retrieval re-access previously encoded information, reactivating the memory traces for effective learning [1].

Studies have shown that memories weaken over time, and the most significant drop in retention happens just a day after you learn something new [2][3][4]. Ebbinghaus's forgetting curve [2] suggests that people tend to forget newly learned materials in days or weeks when there is no attempt to review the learned materials. Besides of that, Rivera-Lares et al. [4] pointed out that the rate of forgetting is independent of the initial degree of learning considering the number and type of presentations of the material.

To improve retention, we need to reinforce what we learn by using techniques such as spaced retrieval practice and interleaving, which may be implemented as low-stake quizzes and used as learning tools [5]. We did that in an Introductory Computer Science Course (CS1). We used quizzes as low-stakes retrieval practice activities (RPAs) that students could take as many times as they wanted during the semester. The innovative part of this work is that we partner the use of RPAs with the U-Behavior method, which is composed of an application integrated into Canvas that generates personalized visualizations of students' study behaviors and self-reflections to help students to be more aware of their study behaviors and improve those behaviors; reinforcing long-term learning.

In a previous study [6], with data from a Fall 2021 CS1 course, we found a statistically significant increase in final exam scores, final coding exam scores, and final course grades for students who practice using the desirable behaviors, spacing, and mixing their practice over the semester, compared with students who did not. This current study presents a work in progress that uses data from Fall 2023 CS1 course to analyze if the increase in grades persists in a different cohort of students for the same course as well as to extend the previous study by monitoring students practice behaviors in a subsequent class, CS2, that is being taught this Spring 2024 semester.

II. Related Works

Spaced retrieval practice has improved performance and long-term retention in laboratory and classroom research. These findings have been repeated in several areas, such as Biology, Math, Education, Engineering, Nursing, History, and Computer Science [7-14].

Khademi et al. [15] developed an open question bank with over 1,500 multiple-choice questions, Parsons problems, and programming exercises for CS1. The authors ran a pilot study and observed a strong engagement from the students. O'Malley and Aggarwal [16] analyzed how students with no prior programming experience (PPE) and with PPE engaged with ungraded quizzes that were designed and deployed to encourage retrieval practice in an introductory programming course. They found that students, independent of previous experience, who engaged with quizzes had a better performance on final exams. Similarly, Aggarwal et al. [17] examined how students with PPE and students without PPE engaged with weekly optional practice quizzes containing multiple-choice and free-response questions. They found that while gender was not significant, students who do not have PPE were more likely to use optional practice than those with PPE.

Interleaved practice has been demonstrated to improve learning compared with mass practice (repeating the practice repeatedly on the same day) in several different contexts [18]. YeckehZaare et al. [12] developed a learner-centered retrieval practice and interleaving tool for an introductory Python course. The authors found that the use of the tool correlated with higher exam grades. In another work, YeckehZaare et al. [13] proposed using a "retrieval-based teaching" strategy where instructors asked ungraded rapid-fire questions. Each question moves from one student to another until the correct answer is reached. The instructor provided feedback for every incorrect answer and presented the correct answer if students could not find it collaboratively. They found that students who were taught with this strategy earned an average of 2.36 percentage points higher in course grades than those who were not exposed to it. Zhang [19] proposed the use of low stake quizzes as part of a lecture. According to the author, the quizzes should contain topics from various contents covered in previous weeks to create a spaced and interleaved practice that is known to improve learning outcomes [20].

U-Behavior [21, 22] is an innovative approach to teaching and learning that reconfigures online quizzes into Retrieval Practice Activities (RPAs) to enhance durable learning. These RPAs are designed to help students master the techniques of distributed retrieval and interleaved practice. Through U-Behavior, students receive online instruction on the importance of practicing these methods. The approach is reinforced with personalized visual learning analytics, allowing students to monitor their practice behavior. These visualizations are complemented with ongoing coaching for learners guiding them to change their practice behavior.

Unlike previous work [12, 13, 16, 17, 19], which applied spacing and interleaving in CS courses, this proposal uses the U-Behavior teaching and learning method by incorporating RPA reflection assignments in an introductory computer science curriculum that uses formative quizzes [7, 18]. The RPA assignments make use of spacing and interleaving strategies that teach students about effective study behaviors by applying self-reflection.

III. METHODS

This study was conducted in a CS1: Java for prior programming experience group as part of a Computer Science Undergraduate program at a Research 1, land-grant university, in the Fall 2023 semester. A convenience sample of 94 students self-registered for the course, with 56 consenting to be included in the study. The course was taught in-person, implemented in 16 weeks, and organized in the form of a spiral curriculum proposed by [23] with three lectures and two labs per week. The course instructor is part of the research team and already implemented U-Behavior in other courses. Canvas was the Learning Management System (LMS) used.

Demographic data was obtained from the institutional research for participants that consented with the study. For the total of 56 consented participants:

- 15 (27%) were female and 41 (73%) were male;
- 16 (28%) were first-generation and 40 (72%) were non-first-generation students;
- 19 (34%) were minoritized and 37 (66%) were non-minoritized students;

Non-minoritized groups in our data were Hispanic/Latino, Asian, Black, and Multi-racial.

The U-Behavior teaching and learning method was incorporated into the course in three steps: a module containing the explanation of the method; RPAs that students have available to do during the entire semester as low stake quizzes; and three reflection assignments. Students start working with U-Behavior by going through the explanation consisting of (1) a short animated interactive video describing the benefits of retrieval practice activities for long-term learning and student performance and the U-Behavior App, (2) a quick survey about the video, and (3) an RPA guide document. The guide explained the four levels of practice behaviors that students would use to self-reflect upon their practice.

- Level 4 - Highly Effective Practice Behavior: Students at this level practiced at least 70% of the total RPAs available in the course on three different days, and also interleaved at least 40% of their practice between RPAs.
- Level 3 - Effective Practice Behavior: Students at this level practiced at least 70% of the total RPAs available in the course on three different days, but interleaved practice between fewer than 40% of the RPAs.
- Level 2 - Less Desirable Practice Behavior: Students at this level practiced between 40-69% of the total of the RPAs available in the course on three different days, interleaved practice is not considered at this level.
- Level 1 - Low Practice Behavior: Students at this level practiced less than 40% of the total of the RPAs available in the course on three different days, interleaved practice is not considered at this level.

Students are asked to do a reflection assignment at three distinct points during the semester (4th week, 8th week and 15th week). The first two reflections were done in the week of the mid-terms, the last reflection was done in the week before final exam. In the first two reflections, students generate their RPA graph visualization using the U-Behavior App (see Fig. 1), compare their graph with four possible graphs (one for each of the previous levels presented), and answer the following reflection question: "After reviewing the feedback on your RPA practice behavior (in Step 2), do you plan to change how you use the RPAs in the upcoming weeks (please describe why and how)?". In the third reflection, students also generate their RPA graph, compare their graph with a desirable study behavior where a student spaced and interleaved their practice during the entire semester, explain if they demonstrate desirable behaviors, and to classify themselves in terms of the percentage of effective practice behaviors, according to the four levels previous presented.

Students had 11 RPAs to practice during the semester, one RPA per week when new concepts were introduced (exams weeks did not have RPAs). The RPAs are part of the required assignments for the course, could be done several times during the semester and the highest grade was kept. The reflections were also required assignments for the course. The U-Behavior App was available as a link in the Canvas navigation bar of the course and could be accessed during the entire semester. Students did not receive any other instruction regarding study habits besides the initial explanation module and these self-reflection assignments.

A. Data Collection

Quantitative data was collected using the Canvas course grade book, and qualitative data was collected from the RPA reflection assignments for all 56 consented students. Of the 56 students, 32 registered for CS2. For those 32, we also collect their Canvas course grade book data and their reflections for the subsequent CS2 course. All data was anonymized after collection.

B. Data Analysis

For the analysis of CS1 data, we will use the last reflection assignment and the following grades: final exam score, final coding exam score, and final course grade. This allowed us to compare the results of the Fall 2021 students with the results of the Fall 2023 students regarding their study behaviors and grades.

To analyze if CS1 students kept the same study behaviors in the subsequent CS2 course, we considered their final reflection assignment in both CS1 Fall 2023 and CS2 Spring 2024.

Students were considered to have a desirable behavior if they reached Level 3 or Level 4, and not desirable behavior if they were in Level 1 or Level 2 of the U-Behavior levels of practice previously presented.

IV. INITIAL RESULTS

A. Fall 2023 CS1

Of the 56 participants, 18 were classified in Level 4 as demonstrating desirable behaviors. Six participants were females, and 12 were males. Three were first-generation students, and 15 were not. Five students were from minoritized groups, and 13 were not. No student was classified in Level 3. Fig. 1 shows the RPA graph of a student in Level 4. As we can observe, this example student practice 90.91% of the total of the available RPAs on three different days, spacing practice sessions with an interval of at least 24 hours, and interleaved/mixed all the RPAs. As we can also observe, student concentrated their attempts around the weeks of the mid-term exams, which is not an optimal behavior. That was the behavior of the majority of the students in this level.

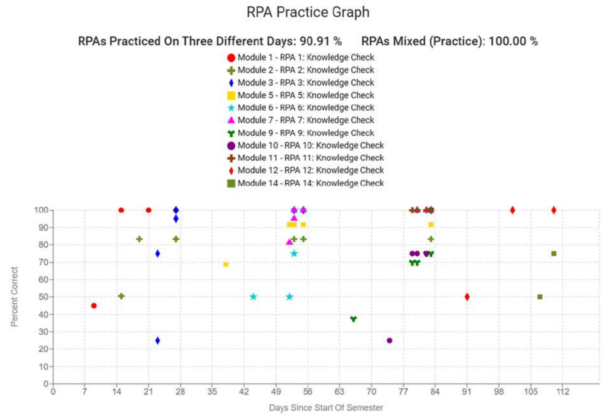

Fig. 1. Example of a student classified in Level 4

Thirty-eight (38) students from the 56 were classified as having undesirable behaviors, 13 were in Level 2, and 25 were in Level 1. Of the 13 in Level 2, three were female, three were first-generation, and three were non-minority. Of the 25 in Level 1, six were female, ten were first-generation, and 11 were from a minoritized group.

Fig. 2 shows an example of a student classified in Level 2. As we can observe from Fig. 2, the student practiced spacing and interleaving until the middle of the semester, and then they stopped that behavior. That was the kind of behavior of all 13 students in that level.

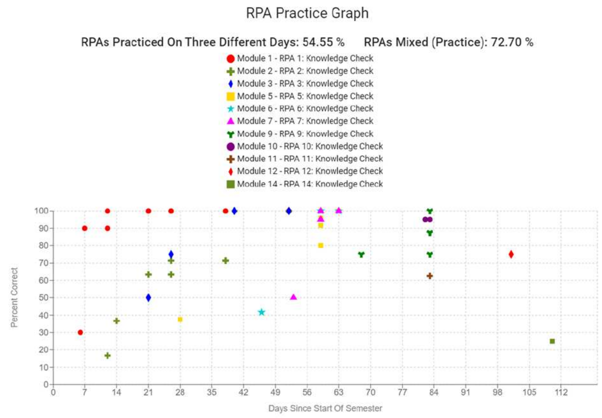

Fig. 2. Example of a student classified in Level 2

Fig. 3 shows an example of a student in Level 1. As we can observe, students massed their practice for each RPA, and after attaining the maximum grade, they never retested the RPA again. That was the behavior of the majority of the students in that level. A few students, after failing to achieve the highest grade on their first attempt at the RPA, returned on another day. Once they achieved the maximum grade, they did not return to that RPA again.

To compare the performance of the 2023 Fall cohort with the 2021 Fall cohort, we divided the students into two groups, following the same directions as our previous work [6]. Group 1 contains students from Level 4 (desirable behaviors), and Group 2 includes students from Levels 1 and 2 (no desirable behaviors). Considering that our dependent variables were not normally distributed, we ran a non-parametric Wilcoxon rank-sum to compare the means of final exam scores, final coding exam scores, and final grade of the course for Group 1 and Group 2.

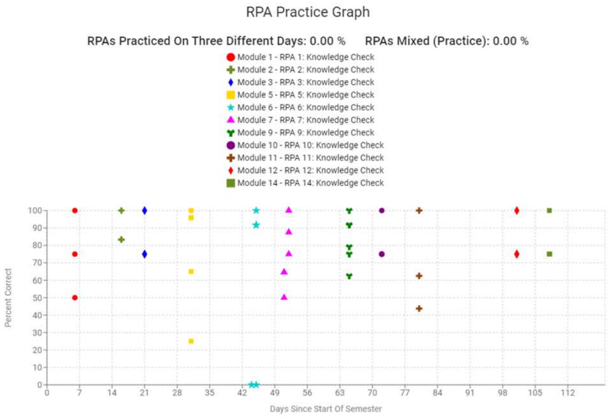

Fig. 3. Example of a student classified in Level 1

We found that Group 1 (N=18, M=18.572, SD=3.916) had a statistically higher final exam score than Group 2 (N=38, M=15.021, SD=5.911), p=.01321, r = 0.298, with a small effect size. We found that there is no statistical difference between Group 1 final coding exam (N=18, M= 6.963, SD=2.610) and Group 2 final coding exam (N=38, M= 6.807, SD=2.820). We found that Group 1 (N=18, M= 86.119, SD= 13.464) had a statistically higher final course grade than Group 2 (N=38, M=75.692, SD=18.219), p= 0.005, r= 0.347, with a medium effect size.

B. *CS2 Study Behaviors*

Of the 56 students in CS1, 32 are registered in CS2 this Spring. Of those 32, 11 are classified as Level 4 in CS1 in the previous semester and five are classified as Level 2, and 16 are classified as Level 1.

Of the 11 students in Level 4, five were female, three were first-generation, and four were from a minoritized group. Of the five students in Level 2, 1 was female, two were first-generation, and one was from a minoritized group. From the 16 in Level 1, four were female, five were first-generation, and eight were from a minoritized group.

We analyzed students' last reflections for CS2 and used the same U-Behavior levels of practice that had been previously presented to classify students' behavior for CS2 and compare those behaviors with their previous behaviors at the end of CS1.

Of the 11 students in Level 4 for CS1, only three maintained the same level of desirable behaviors at the end of CS2. Of those, two students were female, two were first-generation, and two were from a minoritized group (Hispanic/Latino). One student went from Level 4 to Level 2. The remaining seven went to Level 1. Of these seven, three were female, one was first-generation, and two were from the minoritized group.

Of the five students in Level 2 for CS1 only one student kept the same level, 4 students went to Level 1. Of the 16 students in Level 1, one student improved their study behavior and went to Level 2. The remainder of the 15 students maintained their study behavior at Level 1.

V. DISCUSSION AND FUTURE WORKS

Initial analysis of Fall 2023 CS1 data showed a statistically significant increase in final exam scores and final course grades for students who practice desirable study behaviors. Unlike the Fall 2021 CS1 data, the Fall 2023 CS1 data showed no statistically significant difference in final coding exam grades between students who practiced desirable behaviors and those who did not.

Students in the desirable study behaviors, Level 4, for Fall 2023 CS1 had a different behavior than those from Fall 2021. In Fall 2021, students spaced out their attempts across the entire semester. On the contrary, students in Fall 2023 spaced their practice but concentrated their attempts more around the mid-term exam dates, which is not optimal for durable long-term learning. Further analysis on the qualitative data from the reflections should be done to understand the difference between these behaviors.

Students in Level 2 for Fall 2023 seemed to space their practice until the middle of the semester and fall into non-desirable behavior after that point. We need to analyze the qualitative data from the second reflection to understand why that may happened.

Students in Level 1 for Fall 2023 and Fall 2021 exhibited similar study behaviors. After achieving the maximum grade for one RPA, they did not return to that RPA.

Future analysis should consider demographic data such as gender, first-generation and minoritized status, and how those relate to study behaviors and performance.

Behavioral change is complex and difficult to sustain [24]. As we observed, from the 11 students demonstrating desirable study behaviors at the end of Fall 2023 CS1, only three kept the desirable study behaviors in the subsequent class, Spring 2024 CS2. We had just one exception of a student who improved their study behavior from Level 1 to Level 2.

It is important to note that CS1 and CS2 courses had a different organization structure for the RPAs. CS1 has one RPA available at the end of each week, except for the midterm exam and finals week. CS2 has one RPA per unit, and some units are extended in two weeks. In CS1 students have infinite attempts for each RPAs. In CS2, students were limited to 15 attempts per RPA. That limitation had the intention to make students to space their practice. However, some students apparently massed their practice of specific RPAs around the specific mid-terms and ran out of attempts, while others tried to “save” their attempts for finals weeks. We need further analysis to evaluate if having the same organization, one RPA per week, and infinite attempts would help students to continue practicing desirable behaviors, transferring those behaviors between semesters and courses.

We also need to analyze if students use the U-Behavior app to keep track of their study behaviors or if they only use it when they have their reflection assignments. Although this work in progress has a limited number of participants, and the results can’t be generalized to other contexts, it does use an innovative combination of study behavior visualizations, RPAs, and self-reflection that improved performance for students who practice desirable study behaviors.

REFERENCES

[1] A. Savarimuthu, and R. J. Ponniah, “Receive, Retain and Retrieve: Psychological and Neurobiological Perspectives on Memory Retrieval,” Integrative Psychological and Behavioral Science, vol. 58, pp. 303–318, March 2024.

[2] H. Ebbinghaus, “Memory: A Contribution to Experimental Psychology”. 1913th ed., Columbia University. 1855.

[3] J. M. J. Murre, and J. Dros, “ Replication and analysis of Ebbinghaus’ forgetting curve,” PLoS ONE, vol. 10, ed. 7, e0120644, July 2015.

[4] K. Rivera-Lares, R. Logie, A. Baddeley, and S. Della Sala, “Rate of forgetting is independent of initial degree of learning,” Memory&Cognition, vol. 50, pp. 1706–1718, January 2022.

[5] D. H. Murphy, J. L. Little, and E. L. Bjork, “The Value of Using Tests in Education as Tools for Learning—Not Just for Assessment”, Educational Psychology Review, pp. 35–89, September 2023.

[6] Anonymized for blind-review.

[7] R. Ariel, and J. D. Karpicke, “ Improving self-regulated learning with a retrieval practice intervention,” Journal of Experimental Psychology: Applied, vol. 24, no. 1, pp. 43–56, March 2018. https://doi.org/10.1037/xap0000133

[8] R. A Bjork, J. Dunlosky, and N. Kornell, “Self-regulated learning: Beliefs, techniques, and illusions,” Annual review of psychology, vol. 64, pp. 417–444, September 2013.

[9] M. K. Hartwig, and E. D. Malain, “Do students space their course study? Those who do earn higher grades,” Learning and Instruction, vol. 77, 101538, February 2022.

[10] P. Harindranathan, and James Folkestad, “Learning Analytics to Inform the Learning Design: Supporting Instructors’ Inquiry into Student Learning in Unsupervised Technology-Enhanced Platforms,” Online Learning, vol. 23, no. 3, pp. 34–55, September 2019.

[11] K. B. Lyle, C. R. Bego, R. F. Hopkins, J. L. Hieb, and P. A. S. Ralston, “How the amount and spacing of retrieval practice affect the short-and long-term retention of mathematics knowledge,” Educational Psychology Review, vol. 32, no. 1, pp. 277–295, March 2020.

[12] I. YeckehZaare, P. Resnick, and B. Ericson, “A Spaced, Interleaved Retrieval Practice Tool That is Motivating and Effective,” in Proceedings of the 2019 ACM Conference on International Computing Education Research (Toronto ON, Canada) (ICER ’19). Association for Computing Machinery, New York, NY, USA, pp. 71–79. https://doi.org/10.1145/3291279.3339411

[13] I. YeckehZaare, C. Aronoff, and G. Grot, “Retrieval-based Teaching Incentivizes Spacing and Improves Grades in Computer Science Education,” in Proceedings of the 53rd ACM Technical Symposium on Computer Science Education, vol. 1, pp. 892–898, February 2022.

[14] C. Yang, L. Luo, M. A. Vadillo, R. Yu, and D. R. Shanks, “Testing (Quizzing) Boosts Classroom Learning: A Systematic and Meta-Analytic Review,” Psychological Bulletin, vol. 147, No. 4, pp. 399–435, March 2021.

[15] K. Khademi, M. de Vin, A. Adiraju, L. Lin, O. Adeyemi, and B. Hui, “An Open CS1 Learning Platform to Promote and Incentivize Deliberate Practice,” in 2023 IEEE International Conference on Advanced Learning Technologies (ICALT), Orem, UT, USA, 2023, pp. 9-13.

[16] C. O’Malley, and A. Aggarwal, “Evaluating the Use and Effectiveness of Ungraded Practice Problems in an Introductory Programming Course,” in Proceedings of the Twenty-Second Australasian Computing Education Conference (Melbourne, VIC, Australia) (ACE’20). Association for Computing Machinery, New York, NY, USA, pp. 177–184, February 2020. https://doi.org/10.1145/3373165.3373185

[17] A. Aggarwal, N. Puthanveetil, and C. Gardner-Mccune, “Who Attempts Optional Practice Problems in a CS1 Course? Exploring Learner Agency to Foster Mastery Learning,” in Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1 (SIGCSE 2023). Association for Computing Machinery, New York, NY, USA, pp. 1055–1061, March 2023. https://doi.org/10.1145/3545945.3569854

[18] H. L Roediger III, and M. A Pyc, “Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice,” Journal of Applied Research in Memory and Cognition, vol. 1, no. 4, pp. 242–248, December 2012.

[19] L. Y. Zhang, “Tips for Using Gamified Real-Time Polling Quizzes as a No-Stakes Engagement Tool for Computing Courses,” in Proceedings of the 25th Western Canadian Conference on Computing Education (WCCCE '23). Association for Computing Machinery, New York, NY, USA, Article 14, pp. 1–2, July 2023. https://doi.org/10.1145/3593342.3593357

[20] P. C. Brown, H. L. Roediger III, and M. A. McDaniel. Make it stick: The science of successful learning. Harvard University Press, 2014.

[21] K. McKenna, B. Pouska, M. C. Moraes, and J. E. Folkestad, “Visual-form learning analytics: a tool for critical reflection and feedback,” Contemporary Educational Technology, vol. 10, no. 3, pp. 214–228, July 2019.

[22] K. McKenna, J. Folkestad, and M. Moraes, “U-Behavior: Visual-Form Learning Analytics to Enhance Teaching and Learning,” in 6th International Conference on Higher Education Advances. Editorial Universitat Politècnica de València, pp. 697–705, July 2020.

[23] A. Lionelle, S. Ghosh, B. Say, and J. Ross Beveridge, “Increase Performance in CS 2 via a Spiral Redesign of CS 1,” in Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1 (Providence, RI, USA) (SIGCSE 2022). Association for Computing Machinery, New York, NY, USA, pp. 502–508. March 2022. https://doi.org/10.1145/3478431.3499339

[24] M. E. Bouton, “Why behavior change is difficult to sustain.” Prev Med, vol. 68, pp. 29–36, November 2014.