Evaluating the Impact of Research-Based Updates to an Adaptive Learning System

Jeffrey Matayoshi[0000-0003-1321-8159], Eric Cosyn, and Hasan Uzun

McGraw Hill ALEKS, Irvine, CA, USA
{jeffrey.matayoshi,eric.cosyn,hasan.uzun}@aleks.com

Abstract. ALEKS is an adaptive learning system covering subjects such as math, statistics, and chemistry. Several recent studies have looked in detail at various aspects of student knowledge retention and forgetting within the system. Based on these studies, various enhancements were recently made to the ALEKS system with the underlying goal of helping students learn more and advance further. In this work, we describe how the enhancements were informed by these previous research studies, as well as the process of turning the research findings into practical updates to the system. We conclude by analyzing the potential impact of these changes; in particular, after controlling for several variables, we estimate that students using the updated system learned 9% more on average.

Keywords: Adaptive assessment · Forgetting · Retrieval practice

1 Introduction

Since the time of Ebbinghaus and his work on the now famous forgetting curve [2, 5], the study of memory and retention has long been a significant focus of educational research. Over the years, numerous techniques and methods—important examples of which include *spaced practice* [8] and *retrieval practice* [17]—have been shown to help with the long-term retention of knowledge. Within the artificial intelligence in education (AIED) field specifically, learning systems have benefited greatly both by modeling forgetting [4, 16, 25] and using personalized review schedules [10, 15, 21, 22, 27].

The particular system at the center of this work is ALEKS ("**A**ssessment and **LE**arning in **K**nowledge **S**paces") [14], an adaptive learning system covering subjects such as math, statistics, and chemistry. A key feature of ALEKS is its recurring *progress test*, an assessment that is given to a student after a certain amount of learning has occurred. The progress test focuses on the ALEKS problem types a student has recently learned and, among other things, functions as a mechanism for both spaced practice and retrieval practice—as such, it plays a critical role in ensuring students retain their newly acquired knowledge [13]. While the benefits of spaced practice [8, 26] and retrieval practice [3, 9, 17–19] are well-documented, user feedback has shown that students working in the ALEKS system prefer to spend their time learning new material, rather than being assessed by a progress test. Based on these considerations, we began a project with the goal of updating the ALEKS progress test to be shorter and more efficient, thereby giving students additional time to learn and advance in the system—importantly, however, we wanted to do this in such a way as to retain the core benefits of the progress test on knowledge retention.

To that end, we conducted a series of studies [11–13] in an attempt to (a) more completely understand how the retention of knowledge works within the ALEKS system and (b) identify the specific factors that affect this retention. In this work, we discuss how the results from these previous studies were used to make research-based enhancements to ALEKS and the progress test; along the way, we also describe some of the challenges we encountered while implementing these updates. Finally, we conclude with an analysis of the performance of the system after the changes were deployed to production.

2 Previous Research on Forgetting in ALEKS

Our previous research unveiled several findings that informed our updates to the ALEKS system. Using forgetting curves to model knowledge retention in ALEKS, we observed that content on the "edge" of a student's current knowledge decays at a faster rate than content that is "deeper" in the student's knowledge [11]. Building on this result, subsequent work identified other factors affecting knowledge retention in ALEKS—arguably the most significant finding was that the specific characteristics of the problem types have the largest impact on this retention [12]. We also found that a neural network model could effectively use this information to predict the retention of individual problem types [12].

Other results from these studies provided further useful information. For example, we found that students experience a sort of "assessment fatigue" and are less likely to answer a question later in an ALEKS test [11], further highlighting the need to shorten the duration of the progress test. Next, we found evidence that, as a mechanism for retrieval practice, the progress test is more effective when a longer delay exists between the initial learning of a problem type and its appearance on the test [13]. Finally, we observed that further learning of related material in ALEKS can function as a type of retrieval practice. In particular, this act of learning was associated with higher rates of retention compared to the retrieval practice that occurs with a progress test [13], suggesting that learning more and advancing further in the system could be linked to better retention.

3 From Research to Development

Based on these insights, we decided to use the neural network model from [12] to target the problem types students are likely to forget, as these stand to benefit the most from being asked in a progress test. While the previous iteration of the test covered all topics the students recently learned, this targeted approach allowed us to substantially reduce the length of the progress test. Further efficiency was also gained by focusing on problem types learned less recently; as mentioned previously, our research indicated a benefit to delaying the retrieval practice that occurs in progress tests. Finally, other smaller enhancements were made, with the overall goal of helping students learn problem types more efficiently.

The next challenge was to implement these changes within the constraints of a production environment—as these updates would be used by millions of students, the computations needed to (a) run efficiently at scale and (b) be easy to monitor. To address the former concern, we optimized the computational efficiency of the neural network model from [12]. As one example, the original model used a recurrent neural network (RNN) to process the sequential data from students learning in ALEKS. To reduce the resulting computational burden, the RNN was replaced with a set of (non-sequential) features that captured similar information and gave comparable performance. Next, to facilitate the monitoring of the neural network computations, a dedicated database was designed to capture the outputs from the model, along with the relevant input features, making it easy to validate its performance. Additionally, as a side benefit such data would then be readily accessible for any further retraining of the model.

4 Analysis

In this section we use a quasi-experimental design to analyze the impact of the updates on student learning within ALEKS. As the enhancements were pushed to the production servers in July of 2020, we focus on students who started working in the system on or after August 1st, 2020; for these students, we gather data from their activities in ALEKS through the end of the year. To obtain our control group, we find students who started working in the system on August 1st, 2019 or later, and we gather their data through the end of 2019. Additionally, we restrict our search to a selection of seven different math courses that had no version upgrades or changes to their content over this combined time period, as such changes could confound the comparison. These include courses starting as low as fifth grade math and as high as college-level precalculus. Lastly, we require that each student has completed enough work for at least one progress test to be assigned by the system—importantly, the mechanism that assigns this first progress test has not been affected by any of the enhancements to the system.

Year	Students	Average		
		Length	Hours	Number
2019	166,635	19.6	2.3	3.5
2020	154,098	13.3	1.7	3.5

Table 1: Comparison of the 2019 and 2020 student populations.

To compare the behavior of the progress test for these populations, in Table 1 we show the following averages: number of questions on each progress test (length); cumulative hours spent in progress tests per student (hours); and number of progress tests taken per student (number). While the average number of tests is similar between the two populations, the average number of questions and cumulative hours decrease by about 32% and 26%, respectively. The slightly smaller decrease in average hours is likely due to the focus of the updated progress test on problem types students are more likely to forget—these problem types are more challenging and typically take longer for students to answer.

Next, to measure the amount of learning for each student, we compute the difference between the number of problem types the student knows at the beginning of the course—as measured by the *initial test* given in ALEKS—compared to what they know at the end of each study period; we refer to this as the *learning gain* of the student. While we have taken care to try and equalize the student populations from 2019 and 2020, in light of the COVID-19 pandemic it seems unlikely these populations are completely equivalent. Furthermore, from past experience we expect a dependence—or correlation—in the data for students within the same math course, as these students tend to have more similarities.

To address these issues, we fit a multilevel model with the Linear Mixed Effects (LME) class from the `statsmodels` [20] Python library, using the learning gain as our dependent variable and a separate random intercept for each math course. We focus our analysis on an indicator variable that is 1 for students using the updated progress test and 0 otherwise; the coefficient of the variable, β_1, estimates the change in learning gain between the student populations. We also introduce additional independent variables to adjust for differences in the underlying characteristics of the students; these include the time spent learning in the system, amount of learning activity, score on the initial test (as a "ceiling effect" occurs with students who start with more knowledge), and number of problem types in the course (as another ceiling effect occurs with smaller courses). As some of these variables are measured *post-treatment*—i.e., after students have interacted with the progress test—we use the two-step regression procedure outlined in [1] to adjust for post-treatment bias. Specifically, the first step of the procedure is used to adjust for the post-treatment variables, which then allows us to make an unbiased estimate of β_1 in the second step [1, 6, 7, 23, 24]. The resulting estimate for β_1 is 9.8 with a 95% confidence interval of (9.6, 10.0). Thus, holding the other variables constant, students using the updated progress test have an estimated learning gain that is higher by 9.8 problem types on average. As the mean learning gain for students using the original progress test is 104.4, this represents an estimated improvement of approximately 9%.

5 Conclusion

In this work we described a set of research-based enhancements to the ALEKS adaptive learning system, with these enhancements being made to help students learn more and advance further in the system. After adjusting for several variables, a comparison of before and after data indicated that, on average, students using the updated system learned 9% more. In the context of the ongoing COVID-19 pandemic, we find this last result to be encouraging. Given that the pandemic has compounded existing inequities within education, we hope that the improvements made to the ALEKS system can, at least in some part, help students who may otherwise be struggling with their learning.

References

1. Acharya, A., Blackwell, M., Sen, M.: Explaining causal findings without bias: Detecting and assessing direct effects. The American Political Science Review **110**(3), 512 (2016)
2. Averell, L., Heathcote, A.: The form of the forgetting curve and the fate of memories. Journal of Mathematical Psychology **55**, 25–35 (2011)
3. Bae, C.L., Therriault, D.J., Redifer, J.L.: Investigating the testing effect: Retrieval as a characteristic of effective study strategies. Learning and Instruction **60**, 206–214 (2019)
4. Choffin, B., Popineau, F., Bourda, Y., Vie, J.J.: DAS3H: Modeling student learning and forgetting for optimally scheduling distributed practice of skills. In: Proceedings of the 12th International Conference on Educational Data Mining. pp. 29–38 (2019)
5. Ebbinghaus, H.: Memory: A Contribution to Experimental Psychology. Originally published by Teachers College, Columbia University, New York (1885; translated by Henry A Ruger and Clara E Bussenius (1913))
6. Goetgeluk, S., Vansteelandt, S., Goetghebeur, E.: Estimation of controlled direct effects. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **70**(5), 1049–1066 (2008)
7. Joffe, M.M., Greene, T.: Related causal frameworks for surrogate outcomes. Biometrics **65**(2), 530–538 (2009)
8. Kang, S.H.: Spaced repetition promotes efficient and effective learning: Policy implications for instruction. Policy Insights from the Behavioral and Brain Sciences **3**(1), 12–19 (2016)
9. Karpicke, J.D., Roediger, H.L.: The critical importance of retrieval for learning. Science **319**(5865), 966–968 (2008)
10. Lindsey, R.V., Shroyer, J.D., Pashler, H., Mozer, M.C.: Improving students long-term knowledge retention through personalized review. Psychological science **25**(3), 639–647 (2014)
11. Matayoshi, J., Granziol, U., Doble, C., Uzun, H., Cosyn, E.: Forgetting curves and testing effect in an adaptive learning and assessment system. In: Proceedings of the 11th International Conference on Educational Data Mining. pp. 607–612 (2018)
12. Matayoshi, J., Uzun, H., Cosyn, E.: Deep (un)learning: Using neural networks to model retention and forgetting in an adaptive learning system. In: Artificial Intelligence in Education-20th International Conference, AIED 2019. pp. 258–269 (2019)
13. Matayoshi, J., Uzun, H., Cosyn, E.: Studying retrieval practice in an intelligent tutoring system. In: Proceedings of the Seventh ACM Conference on Learning @ Scale. pp. 51–62 (2020)
14. McGraw Hill ALEKS: What is ALEKS? `https://www.aleks.com/about_aleks` (2021)
15. Pavlik, P.I., Anderson, J.R.: Using a model to compute the optimal schedule of practice. Journal of Experimental Psychology: Applied **14**(2), 101 (2008)
16. Qiu, Y., Qi, Y., Lu, H., Pardos, Z.A., Heffernan, N.T.: Does time matter? Modeling the effect of time with Bayesian knowledge tracing. In: Proceedings of the 4th International Conference on Educational Data Mining. pp. 139–148 (2011)
17. Roediger III, H.L., Butler, A.C.: The critical role of retrieval practice in long-term retention. Trends in Cognitive Sciences **15**, 20–27 (2011)

18. Roediger III, H.L., Karpicke, J.D.: The power of testing memory: Basic research and implications for educational practice. Perspectives on Psychological Science **1**(3), 181–210 (2006)
19. Roediger III, H.L., Karpicke, J.D.: Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science **17**(3), 249–255 (2006)
20. Seabold, S., Perktold, J.: Statsmodels: Econometric and statistical modeling with Python. In: 9th Python in Science Conference (2010)
21. Settles, B., Meeder, B.: A trainable spaced repetition model for language learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1848–1858 (2016)
22. Tabibian, B., Upadhyay, U., De, A., Zarezade, A., Schölkopf, B., Gomez-Rodriguez, M.: Enhancing human learning via spaced repetition optimization. Proceedings of the National Academy of Sciences **116**(10), 3988–3993 (2019)
23. Vansteelandt, S.: Estimating direct effects in cohort and case-control studies. Epidemiology pp. 851–860 (2009)
24. Vansteelandt, S., Goetgeluk, S., Lutz, S., Waldman, I., Lyon, H., Schadt, E.E., Weiss, S.T., Lange, C.: On the adjustment for covariates in genetic association analysis: a novel, simple principle to infer direct causal effects. Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society **33**(5), 394–405 (2009)
25. Wang, Y., Heffernan, N.T.: Towards modeling forgetting and relearning in ITS: Preliminary analysis of ARRS data. In: Proceedings of the 4th International Conference on Educational Data Mining. pp. 351–352 (2011)
26. Weinstein, Y., Madan, C.R., Sumeracki, M.A.: Teaching the science of learning. Cognitive Research: Principles and Implications **3**(1), 2 (2018)
27. Xiong, X., Wang, Y., Beck, J.B.: Improving students' long-term retention performance: A study on personalized retention schedules. In: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. pp. 325–329. ACM (2015)