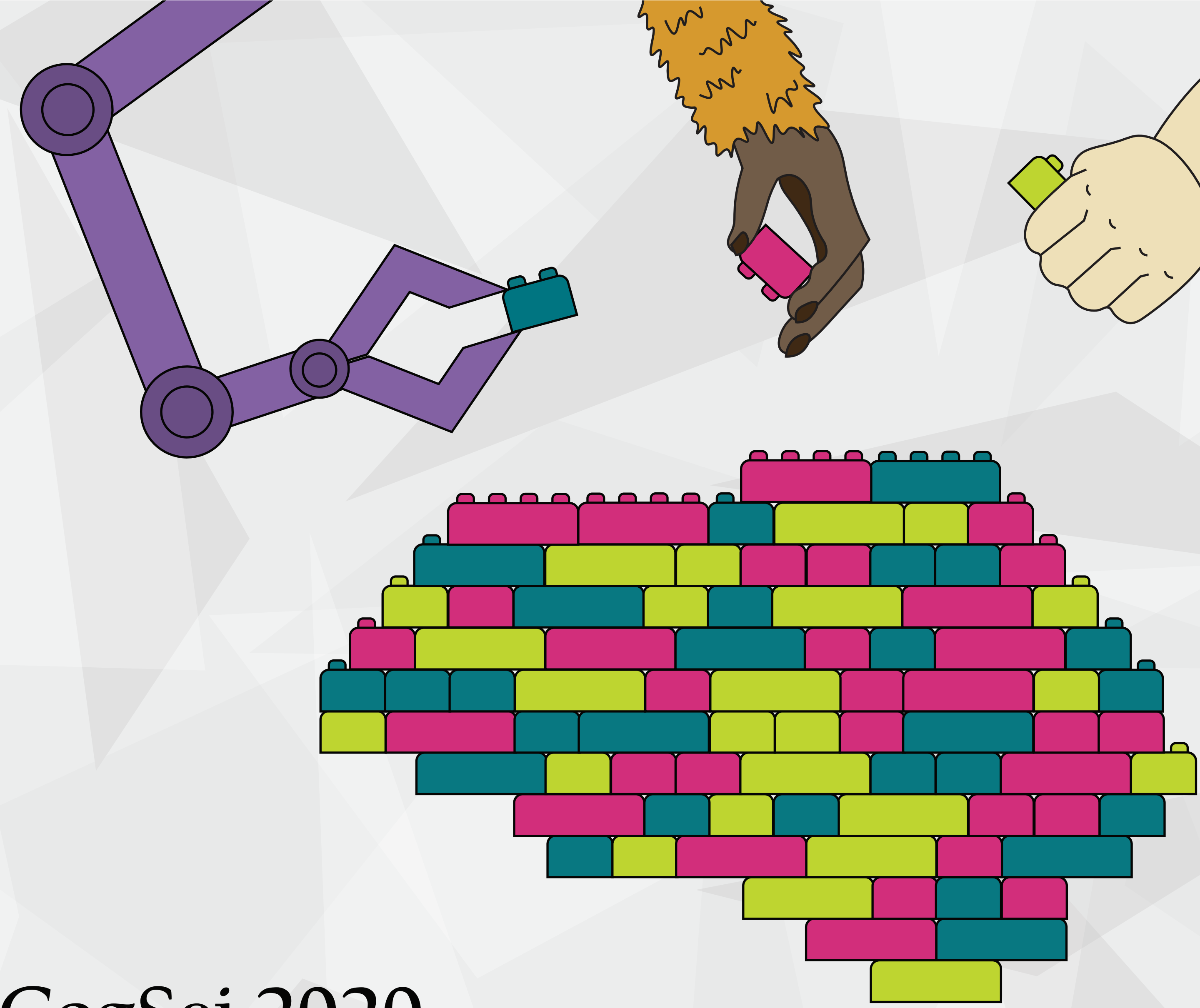

CogSci 2020 Proceedings

Developing a Mind:
Learning in Humans, Animals, and Machines

Invited Speakers
Cecilia Heyes
Geoffrey Hinton
Janet Werker

Invited Panels
Deep Integration of Development and Cognitive Science
Social, Cultural, and Linguistic Constraints on Development
Statistical Learning and Development

Co-Chairs: Stephanie Denison | Michael Mack | Yang Xu | Blair C. Armstrong

Adaptive vs. Fixed Spacing of Learning Items: Evidence from Studies of Learning and Transfer in Chemistry Education

Everett Mettler (mettler@ucla.edu)[1]
Christine M. Massey (cmassey@psych.ucla.edu)[1]
Amina K. El-Ashmawy (ael-ashmawy@collin.edu)[2]
Philip J. Kellman (kellman@cognet.ucla.edu)[1]

[1]Department of Psychology, University of California, Los Angeles
Los Angeles, CA 90095 USA
[2] Department of Chemistry, Collin College
McKinney, TX 75071 USA

Abstract

Spacing presentations of learning items across time improves memory relative to massed schedules of practice – the well-known spacing effect. Spaced practice can be further enhanced by adaptively scheduling the presentation of learning items to deliver customized spacing intervals for individual items and learners. ARTS - Adaptive Response-time-based Sequencing (Mettler, Massey, & Kellman 2016) determines spacing dynamically in relation to each learner's ongoing speed and accuracy in interactive learning trials. We demonstrate the effectiveness of ARTS when applied to chemistry nomenclature in community college chemistry courses by comparing adaptive schedules to fixed schedules consisting of continuously expanding spacing intervals. Adaptive spacing enhanced the efficiency and durability of learning, with learning gains persisting after a two-week delay and generalizing to a standardized assessment of chemistry knowledge after 2-3 months. Two additional experiments confirmed and extended these results in both laboratory and community college settings.

Keywords: adaptive learning, spacing effect, chemistry education, STEM learning

Introduction

Spacing learning opportunities across time improves long-term retention relative to massing material in the short-term - the well known spacing effect (Dempster, 1989; Ebbinghaus, 1913). Spacing improves learning across a variety of materials and learning modes and has the potential to greatly improve educational outcomes, as indicated by experts in an Institute of Education Sciences-sponsored practice guide reviewing scientific evidence (Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel & Metcalfe, 2007). However, experimental studies that involve meaningful content and consequential learning outcomes in real-world settings are uncommon.

Prior laboratory studies have demonstrated that learning gains due to spaced practice can be further enhanced by dynamically generating spacing delays that are appropriate to variations in learners and learning content. ARTS, Adaptive Response-time-based Sequencing, is a method of adaptively scheduling the presentation of learning items to deliver beneficial spacing intervals for individual items and learners as a function of ongoing performance. ARTS determines spacing dynamically from each learner's speed and accuracy in interactive learning trials (Mettler, Massey, & Kellman, 2011, 2016).

Chemistry education poses significant challenges in terms of the amount of material to be learned, the pace of instruction, and the need to achieve sufficient levels of mastery to support subsequent learning. Introductory courses could likely be improved if principles of spacing were applied to the learning of basic information in chemistry. In a series of studies, we assessed the effectiveness of ARTS when applied to the learning of chemistry nomenclature content, where learning items consisted of names and formulas for polyatomic ions and acids. We compared adaptive spacing to fixed schedules of practice that were preset and not adaptive, to investigate the relative benefits of adaptive spacing over fixed spacing. Fixed spacing intervals were 'expanding', that is, spacing delays got continuously larger across the learning session. Expanding intervals have been thought to enhance learning (Bjork & Allen, 1970) compared to fixed spacing schedules with 'equal' interval sizes. Though there are debates about the benefits of expanding spacing (e.g., Karpicke & Roediger, 2007), there is no doubt that fixed expanding schedules are one plausible type of effective spacing schedule that would be useful to compare against adaptive schedules (see Mettler, Massey & Kellman, 2016 for comparisons of schedules with fewer total presentations).

In three studies we examined the effect of adaptive spacing on the learning and fluent use of basic chemical nomenclature. We focused on several questions. First, do adaptive schedules of practice improve learning of chemistry nomenclature? Second, do adaptive schedules outperform fixed schedules when learning is not limited to a fixed number of presentations, but instead proceeds until learners reach objective mastery criteria? Do the advantages of adaptive scheduling replicate between the laboratory and real-world learning scenarios? Finally, does adaptive learning lead to better learning in the classroom and on standardized tests of chemistry knowledge administered at the end of a school semester?

In the first study, we assessed whether learners who learned using adaptive spacing outperformed learners who studied using fixed spacing schedules. We conducted this study with students enrolled in introductory chemistry classes at a community college. In a second study, we replicated the results using undergraduate students who had not taken any college-level chemistry courses. In a third study, also with community college students, we manipulated the type of item retirement that occurred in fixed schedules in order to assess fixed schedules that have capabilities similar to an adaptive scheduling system.

Exp. 1: Adaptive vs. Fixed Expanding Spacing

In order to assess the effectiveness of adaptive spacing in chemistry learning, a set of chemistry nomenclature facts was used as learning items, and the repetition of individual items and the spacing between repetitions was manipulated.

There were two scheduling conditions. In the Adaptive condition, schedules were generated with Adaptive Response-Time-based Sequencing (ARTS). In the Fixed condition, spacing delays were generated using an algorithm that attempted to present items with a fixed, expanding schedule of spacing delays. The fixed spacing algorithm was as follows: Initial intervals were 1 trial, then 5 trials, then 9 trials, then 13, and so on - increasing by 4 trials at each interval. For example, a hypothetical chemistry fact, item A, would be presented on trial number 1, then trial 3, 8, 17, 30, etc. Since, in any given set of learning items there is a maximum possible spacing interval size, which is a function of the total number of items (non-retired items) in the learning set and the number of times items are repeated, the fixed condition continued to present the longest possible delay for each item even if the exact expected spacing delay size was not reached. Similarly, in the event of spacing delay conflict, the algorithm attempted to match as closely as possible the intended spacing interval size for each item at each presentation.

In the Adaptive condition, items were scheduled using ARTS, and each item was subject to the following retirement criteria: four of the last four presentations correct, with each trial's reaction time < 7 seconds. Once an item reached its learning criteria, it was removed from the active learning set. In the Fixed condition, items were not removed, and the experimental session finished when all items had met the learning criteria previously described.

Method

Participants 31 community college students enrolled in an introductory chemistry course (15 in the Adaptive condition and 16 in the Fixed condition) completed study activities as assigned work as part of their course curriculum.

Materials 23 chemistry nomenclature items were used, selected by the course instructor to be most relevant to the learning material in the course. The items included 17 polyatomic ion names and 6 acid names. An individual trial consisted of a presentation of one item in one of two trial type formats, where the user response was followed by

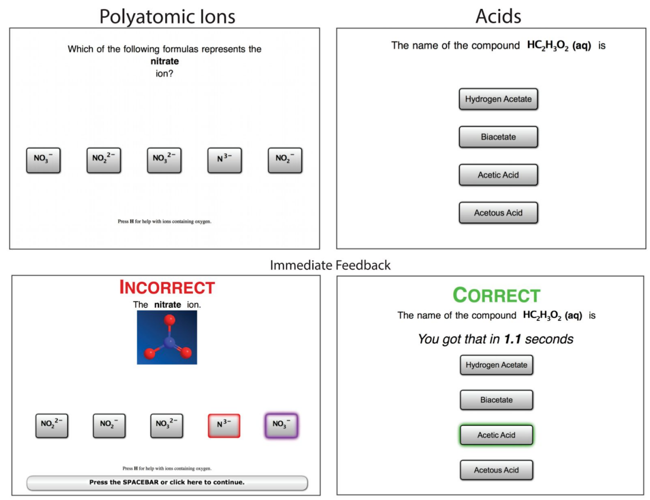

Figure 1: Examples of polyatomic ion naming and acid naming trials with feedback.

feedback indicating the correctness of their answer. Trial types included a mixture of two types of mappings between formulas and names: either a name was presented and learners were asked to generate the corresponding formula by selecting from 4 choices, or a formula was presented and learners were asked to type the corresponding name using the keyboard, in the case of ions, or select the name from four multiple choices in the case of acids. To simplify the design, a random half of the nomenclature items were assigned to one trial type and the other half of items to the other trial type. Trials are illustrated in Figure 1.

Design Two conditions were compared, an Adaptive condition that utilized the ARTS algorithm, and a Fixed scheduling condition, Fixed Continuous Expanding Spacing, where the intervals between presentations of each item were pre-set and where successive spacing intervals grew continuously larger. Participants were randomly assigned to one of the two conditions.

Procedure Participants began with a pretest on 23 items, followed by a training phase. Test and training items were identical except that in training participants received feedback as to the correctness of their answer while test items had no feedback. A delayed posttest was administered 2 weeks after the immediate posttest with the same 23 items.

ARTS - Adaptive Response-time-based Sequencing ARTS uses a priority score system, in which the priority for an item to reappear on each learning trial is computed as a function of accuracy, response time, and trials since the last presentation. The priority score for an item i is given in Equation 1.

$$(1) \qquad P_i = a(N_i - D)[b(1-\alpha_i)Log(RT_i/r) + \alpha_i W]$$

N_i is trials since last presentation; D is an enforced delay constant; α_i is a "switch" that is 1 if the answer is incorrect and 0 if the answer is correct, utilizing RT only in the latter case; W is a priority increment for wrong answers; a, b, and r are constants. Details can be found in other work (e.g., Mettler, Massey & Kellman, 2011). ARTS can also be applied to adaptive category sequencing in perceptual learning where items to be spaced are categories rather than individual facts (e.g., Mettler & Kellman, 2014). ARTS has previously been applied to PL in chemistry education involving perceptual learning of 3d chemical structure (El-Ashmawy et al., 2013).

ARTS implements mastery criteria based on both accuracy and speed. As learning strength increases, as reflected in performance, spacing intervals automatically grow. Because all items compete for presentation on any trial through their priority scores, the system concurrently tends to optimize adaptive spacing for all learning items.

Planned Analyses Our primary measure of learning performance was learning *efficiency*, defined as accuracy gain from pretest to posttest divided by the number of trials invested in learning and multiplied by the number of learning items. Efficiency gives a way of measuring learning retention that incorporates variations in both posttest performance and the number of learning trials required to reach mastery criteria. It may be thought of as a rate measure, indicating performance improvement in an item per learning trial; multiplying by the number of items scales this measure to have a maximum value of 1.0. We also examined raw accuracy change scores between pre and posttests and learning performance at equivalent points during the learning session. Measures of performance at equivalent moments during learning convey the relative rapidity of learning for learners using a given schedule. All measures were assessed using standard parametric statistics such as ANOVA and planned comparisons between conditions. All statistical tests were two-tailed, with a 95% confidence level, all effect sizes d are Cohen's d, and all error bars in graphs show +/- 1 standard error of the mean.

Results

Posttest Efficiency Efficiency scores for the adaptive and fixed spacing conditions at each posttest phase are shown in Figure 2. Scores were highest for the Adaptive condition, both at immediate posttest (Adaptive: M=0.038, SD=0.015; Fixed: M=0.012, SD=0.012) and at a 2 week delayed posttest (Adaptive: M=0.028, SD=0.014; Fixed: M=0.012, SD=0.009). A 2x2 ANOVA was conducted on efficiency scores using posttest phase and scheduling condition as factors. There was a significant effect of scheduling condition ($F(1,29)=24.6$, $p<.001$), a significant effect of posttest phase ($F(1,29)=8.48$, $p=.007$) and a significant interaction between scheduling condition and posttest phase ($F(1,29)=7.02$, $p=.013$). Paired comparisons showed significant differences between Adaptive and Fixed condition efficiencies at immediate posttest ($t(29)=5.32$, $p<.001$, $d=1.92$) and delayed posttest ($t(29)=3.83$, $p<.001$, $d=1.4$). Comparing means across posttest phases, there was a significant difference between immediate and delayed posttests for the Adaptive condition ($t(14)=3.62$, $p=.003$, $d=0.65$), but no significant difference for the Fixed condition ($t(15)=0.272$, $p=.79$, $d=0.06$).

Accuracy Change Scores Accuracy was also analyzed, but owing to differences in the number of trials to reach the same mastery criterion in each condition, accuracy was assumed to be a less informative measure of learning gains than efficiency scores. Accuracy change scores were computed by subtracting each participant's pretest score from their posttest scores. Pretests differed across conditions (Adaptive: M=0.38, SD=0.19; Fixed: M=0.55, SD=0.24 ; $t(29)=2.15$, $p=.04$, $d=0.78$). Change scores were higher for the Adaptive condition both at immediate posttest (Adaptive: M=0.348, SD=0.113; Fixed: M=0.291, SD=0.256) and at a 2 week delayed posttest (Adaptive: 0.258, SD=0.116; Fixed: M=0.253, SD=0.178). A 2X2 ANOVA was conducted on accuracy change scores using posttest phase and scheduling condition as factors. There was no effect of scheduling condition ($F(1,29)=0.28$, $p=.6$), a significant effect of posttest phase ($F(1,29)=7.16$, $p=.012$), and no interaction between scheduling condition and posttest phase ($F(1,29)=1.20$, $p=.281$). Paired comparisons showed no significant differences between Adaptive and Fixed conditions at either immediate posttest ($t(30)=0.8$, $p=.4$, $d=0.31$) or delayed posttest ($t(30)=0.1$, $p=.9$, $d=0.036$). Comparing means across posttest phases, there was a significant difference between immediate and delayed

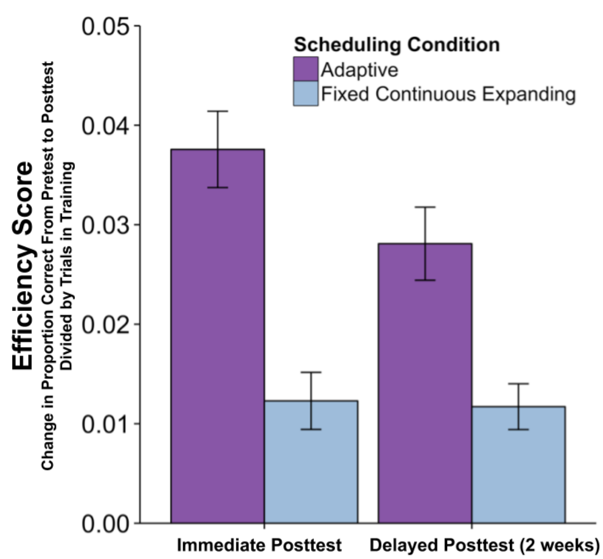

Figure 2: Learning efficiency in immediate and 2-week delayed posttest in Experiment 1.

posttests for the Adaptive condition (t(14)=3.46, p=.004, d=0.787), but not for the Fixed condition (t(15)=0.984, p=.341, d=0.175).

Equivalent Trials Analysis Participants took longer to reach retirement in the Fixed condition (515 trials) than in the Adaptive condition (255 trials), so we compared learning performance at equivalent points during training for the Adaptive and Fixed conditions, specifically for the last 2 presentations of each item before retirement in the adaptive condition and before trial 255 in the Fixed condition. Accuracy was reliably higher in the Adaptive condition (t(29)=5.07, p<.001), with a very large effect size (d=2.62).

Discussion

As demonstrated across dependent measures, adaptive sequencing outperformed predetermined schedules of practice at both an immediate and a delayed test, and on measures of performance taken at equivalent times during learning. In Exp. 2, we sought to replicate these findings with a different learner group in a controlled laboratory rather than a classroom setting.

Experiment 2: Laboratory Replication

Experiment 2 was similar to Experiment 1. Its purpose was to replicate the findings of Exp. 1 in a more controlled laboratory setting. We also used a different learner group, more homogeneous than the community college sample in having no recent exposure to chemistry at the college level.

Method

Participants Participants were 36 UCLA undergraduates who received psychology course credit for participation. Participants were screened for level of chemistry knowledge based on their completion of chemistry courses. Participation was limited to students who had not taken any chemistry courses at the college level.

Materials, Design & Procedure The materials, design and procedure were identical to Experiment 1 except that the delayed posttest was administered after 1 week.

Results

Posttest Efficiency Efficiency scores for the two scheduling conditions and at each posttest phase are shown in Figure 3. Scores were highest for the Adaptive condition, both at an immediate posttest (Adaptive: M=0.042, SD=0.017; Fixed: M=0.025, SD=0.005) and at a 1 week delayed posttest (Adaptive: M=0.024, SD=0.018; Fixed: M=0.015, SD=0.007). A 2X2 ANOVA was conducted on efficiency scores using posttest phase and scheduling condition as factors. There was a significant effect of scheduling condition (F(1,34)=10.72, p=.002), a significant effect of posttest phase (F(1,34)=48.9, p<.001), and a marginally significant scheduling condition by posttest phase interaction (F(1,34)=4.03, p=.053). Paired comparisons showed significant differences between Adaptive and Fixed condition efficiencies at immediate posttest (t(34)=3.86, p<.001, d=1.48), and a marginally significant difference between efficiencies at delayed posttest (t(34)=1.94, p=.061, d=0.704). Comparing means across posttest phases, there was a significant difference between tests for both the Adaptive (t(17)=4.73, p<.001, d=1.05) and the Fixed condition (t(17)=8.17, p<.001, d=1.74).

Pretest Accuracy and Change Scores Pretest accuracy was not different between conditions (Adaptive: M=0.23, SD=0.15; Fixed: M=0.55, SD=0.12 ; t(34)=0.10, p=.92, d=0.03). Change scores were higher for the Fixed condition both at immediate posttest (Adaptive: M=0.413, SD=0.164; Fixed: M=0.609, SD=0.131) and at a 1-week delayed posttest (Adaptive: M=0.229, SD=0.171; Fixed: M=0.357, SD=0.162). A 2X2 ANOVA was conducted on accuracy change scores using posttest phase and scheduling condition as factors. There was a significant effect of scheduling condition (F(1,34)=11.96, p=.001), a significant effect of posttest phase (F(1,34)=81.33, p<.001), and no scheduling condition by posttest phase interaction(F(1,34)=1.97, p=.17). Paired comparisons between scheduling conditions at each posttest showed significant differences between Adaptive and Fixed at immediate posttest (t(34)=3.95, p<.001, d=1.326) and at delayed posttest (t(34)=2.302, p=.028, d=0.768). Comparing means across posttest phases, there were significant differences between immediate and delayed posttest for the Adaptive condition (t(17)=4.95, p<.001,

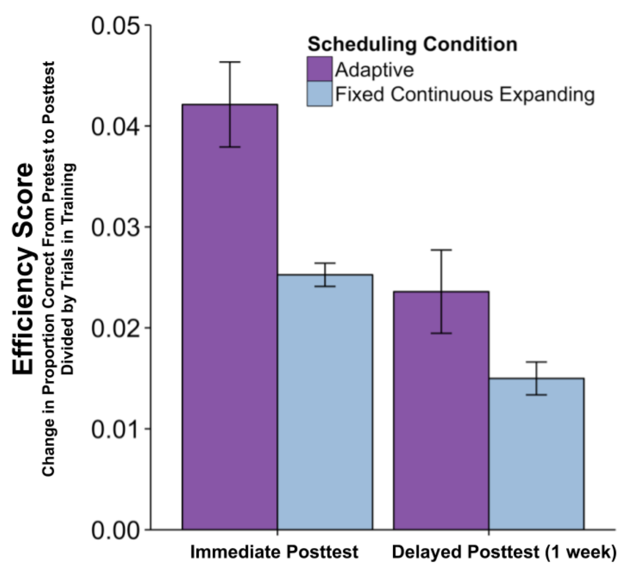

Figure 3: Learning efficiency in immediate and a 1 week delayed posttest in Experiment 2.

d=1.095), and for the Fixed condition (t(17)=8.168, p<.001, d=1.713).

Equivalent Trials Analysis Participants took longer to reach retirement in the Fixed condition than in the Adaptive condition, so we compared learning performance at equivalent points during training for the Adaptive and Fixed conditions. Participants took on average 260 trials to reach retirement in the Adaptive condition and 594 in the Fixed condition. When looking at performance in both conditions at trial 260, accuracy was higher in the Adaptive condition than in the Fixed condition, a significant difference (t(34)=8.75, p<.001, d=4.125).

Discussion

The results of experiment 1 were replicated using university students who were not actively enrolled in chemistry courses. Again, adaptive sequencing showed greater efficiency than predetermined schedules of practice at both immediate and delayed tests, as well as at an equivalent point during the course of learning.

Experiment 3 – Adaptive vs. Fixed Continuous Expanding with Retirement

Experiments 1 and 2 demonstrated that learning could improve in both classroom and laboratory contexts if individual chemistry facts were adaptively scheduled using ongoing learner response speed and accuracy. These differences were present when learning criteria and retirement features were applied to each individual item in the Adaptive condition; Fixed continuous expanding conditions did not include a retirement (dropout) feature. In experiment 3, we included the retirement feature in both Adaptive and Fixed conditions. For the Fixed condition, each item could be retired – removed from the active learning set – when it met retirement criteria. The retirement criteria were thus equivalent across the Adaptive and Fixed conditions and were the same as the retirement criteria for the Adaptive condition in experiments 1 and 2.

Method

Participants Participants were 63 introductory chemistry students at Collin College who participated as part of an introductory chemistry course.

Materials, Design & Procedure The materials, design and procedure were identical to experiments 1 and 2 with the following differences: The procedure was altered so that items in the Fixed condition could be retired after reaching learning criteria. The learning criteria were the same as Experiments 1 and 2 for the Adaptive conditions - four of the last four presentations of an item answered correctly, with each response faster than 7 seconds. The delayed posttest, as in Experiment 1, was administered 2 weeks after the immediate posttest. Participants were assigned to conditions randomly, but due to scheduling errors, there were 33 participants assigned to the Adaptive condition and 30 participants to the Fixed condition.

Results

Posttest Efficiency Efficiency scores for the two scheduling conditions and at each posttest phase are shown in Figure 4. Scores were numerically higher for the Adaptive condition, both at an immediate posttest (Adaptive: M=0.038, SD=0.023; Fixed: M=0.036, SD=0.018) and at a 2 week delayed posttest (Adaptive: M=0.028, SD=0.024; Fixed: M=0.022, SD=0.018). A 2X2 ANOVA was conducted on efficiency scores using posttest phase and scheduling condition as factors. There was no reliable effect of scheduling condition (F(1,61)=0.583, p=0.448), a significant effect of posttest phase (F(1,61)=27.97, p<.001), and no reliable scheduling condition by posttest phase interaction (F(1,61)=0.643, p=0.426).

Paired comparisons showed no significant differences between Adaptive and Fixed condition efficiencies at immediate posttest (t(61)=0.353, p=.72, d=0.09) or at delayed posttest (t(61)=1.022, p=.31, d=0.262). Comparing means across posttest phases, there was a significant difference between tests for both the Adaptive (t(32)=3.074, p=.00, d=0.437) and the Fixed condition (t(29)=4.583, p<.001, d=0.766).

Pretest Accuracy and Change Scores Pretest accuracy was not different between conditions (Adaptive: M=0.33 SD=0.16; Fixed: M=0.33, SD=0.14 ; t(61)=0.11, p=.92, d=0.02). Accuracy change scores were highest for the Fixed condition at immediate posttest (Adaptive: M=0.336, SD=0.173; Fixed: M=0.397, SD=0.154) and the same at a 2-week delayed posttest (Adaptive: M=0.24, SD=0.191; Fixed: M=0.24, SD=0.169). A 2X2 ANOVA was conducted on accuracy change scores using posttest phase and scheduling condition as factors. There was no significant effect of scheduling condition (F(1,61)=0.767, p=.385), a significant effect of posttest phase (F(1,61)=26.3, p<.001), and no scheduling condition by posttest phase interaction (F(1,61)=1.474, p=.229). Paired comparisons between scheduling conditions at each posttest did not show significant differences at immediate posttest (t(61)=1.478, p=.14, d=0.374) or at delayed posttest (t(61)=0.049, p=.96, d=0.012). Comparing means across posttest phases, there was a significant difference between immediate and delayed posttest for the Adaptive condition (t(32)=2.847, p=.01, d=0.529), and a significant difference between immediate and delayed posttests for the Fixed condition (t(29)=4.464, p<.001, d=0.96).

Equivalent Trials Analysis Participants took longer to reach retirement in the Fixed than in the Adaptive condition. Participants took on average 215 trials to reach retirement in

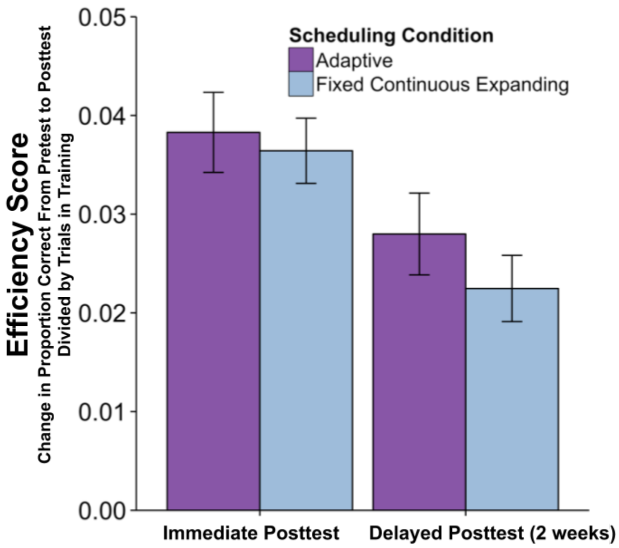

Figure 4: Learning efficiency in immediate and 2-week delayed posttest by condition in Experiment 3.

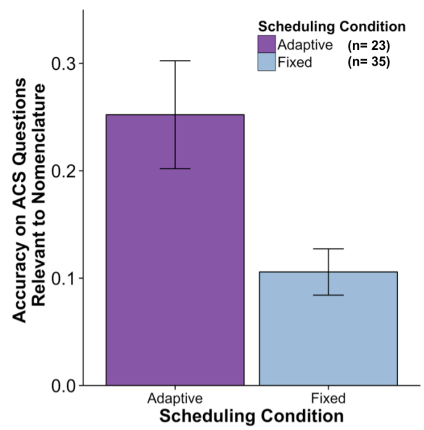

Figure 5: Results by condition from an American Chemical Society standardized examination administered at the end of community college semester.

the Adaptive condition and 275 in the Fixed condition, a significant difference (t(61)=3.53, p<.001, d=0.917). When looking at performance in both conditions at an equivalent point in learning, accuracy was higher in the Adaptive condition than in the Fixed condition, but this was not a significant difference (t(61)=1.05, p=.298, d=0.365).

Discussion

Experiment 3 compared fixed and adaptive schedules when both types of schedule utilized the same mastery criteria and individual items dropped out when a learner reached objective mastery criteria for each item. Equating mastery and retirement criteria tended to equalize performance across the adaptive and fixed expanding schedules, although speed of retirement occurred reliably faster in the adaptive condition. Results with this community college sample did not show the clear advantage of adaptive spacing after equal numbers of trials shown in Experiments 1 and 2 of the present work or in other research (Mettler, Massey & Kellman, 2016; Mettler et al., 2020).

American Chemical Society Standardized Exam

In addition to measures of learning evaluated here, community college participants in our studies took a standardized test developed by American Chemical Society (ACS) exam at the completion of the school semester, as a standard part of their courses. We analyzed 9 questions from the ACS exam that were related to chemistry nomenclature knowledge. The questions were chosen by an instructor who did not have knowledge of the results of the prior studies. Students who participated in the ACS exam had participated in experiments comparing adaptive and fixed schedules, including participants in the Adaptive and Fixed conditions of Experiment 1 above, as well as Adaptive and Fixed participants in an experiment not reported here where the total number of presentations of each item was limited and equated across conditions.

Participants in these two groups, Adaptive and Fixed, were compared in terms of their accuracy on ACS exam questions relevant to chemistry nomenclature. The results of the ACS exam questions are displayed in Figure 5 for the two groups. The accuracy of participants in the Adaptive conditions was higher (M=0.252, SD=0.241, N=23) than in the Fixed conditions (M=0.106, SD=0.128, N=35), a significant difference (t(56)=3.017, p=.004, d=0.795).

Conclusion

Experiments 1 and 2 showed clear benefits of adaptive learning over fixed expanding spacing in parallel studies of chemistry learning in a real-world learning setting and in a controlled laboratory setting. The results suggested that both adaptively generated spacing intervals and use of mastery criteria to retire items produced these benefits, as shown in both efficiency and equivalent trials accuracy analyses. Experiment 3, however, with a different community college sample and mastery criteria applied to both conditions, showed faster learning with adaptive spacing, but generally minimal differences between conditions otherwise. As earlier work showed clear advantages of adaptive spacing apart from use of mastery criteria (Mettler et al., 2016), we believe the differences across studies here reflect the considerable variability of prior and ongoing chemistry learning, as well as less well-controlled conditions in studying community college classes relative to laboratory settings. Most encouraging in these settings, however, is that in the aggregate, students who received adaptive spacing in studies with and without mastery criteria outperformed students who received fixed spacing schedules when tested after a substantial delay on a transfer test: a standardized

ACS test of chemistry learning administered months later at the end of students' courses. Overall, the results of these studies indicated the benefits of adaptive learning methods in a real-world STEM-learning context. An adaptive learning system that guided spacing and mastery based on each individual learner's performance on individual items during the course of learning generally produced immediate, delayed, and transfer test performance that outperformed fixed spacing schedules. The fixed spacing schedules chosen as controls were an evidence-based, non-adaptive alternative where spacing interval sizes continuously increased during learning (expanding spacing intervals).

Taken together, these results confirm and extend earlier results with adaptive systems and support the general ideas that 1) spacing intervals should increase as underlying learning strength increases; 2) that learning strength varies by learners and items and fluctuates during the course of learning; and 3) that a combination of learner accuracy and response time may be used effectively to estimate learning strength, both for determination of favorable spacing and for objective estimation of mastery.

Spacing learning items adaptively based on ongoing measures of learning strength have the potential to improve learning in chemistry education, as well as in other challenging learning domains.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1109228 and 1644916 and by the National Institutes of Health Grant No. 1RC1HD063338-01. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or the National Institutes of Health.

References

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*(5), 567–572.

Dempster, F. N. (1989). Spacing effects and their implications for theory and practice. *Educational Psychology Review, 1*(4), 309-330.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.

El-Ashmawy, A., Russell, A., Mettler, E., Burke, T., Kellman, P., & Massey, C. (2013, June). *Adaptive and perceptual learning of chemical structure and nomenclature.* Poster session presented at the Gordon Research Conference: Chemical Education Research, Newport, RI.

Karpicke, J. D., & Roediger, H. L., 3rd. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*(4), 704–719.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research, 99*, 111–123.

Mettler, E., Massey, C. M., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response-times. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2532-2537). Austin, TX: Cognitive Science Society.

Mettler, E., Massey, C. M., Burke, T., & Kellman, P. J. (2020). Comparing adaptive and random spacing schedules during learning to mastery criteria. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Toronto, Canada: Cognitive Science Society.

Mettler, E., Massey, C. M., & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology. General, 145*(7), 897–917.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. National Center for Education Research.