Distributing Practice: Challenges and Opportunities for Inquiry Learning

Vanessa Svihla, University of New Mexico, Hokona Hall, Albuquerque, NM 87131, vsvihla@unm.edu
Marcia Linn, University of California at Berkeley, Tolman Hall, Berkeley CA 94720, mclinn@berkeley.edu

Abstract: This study investigates the impact of differing lag times on distributed practice compared to the traditional clustered approach. High school students studied the Web-based Inquiry Science Environment unit, Global Climate Change in one of three conditions. We compare student learning outcomes. We find that traditional spacing resulted in larger gains, in apparent contrast to the findings of many studies of distributing practice. We further investigate this finding to understand how spacing might have impacted revisiting and restudying of earlier work. Specifically, we use log files to determine how students revisited earlier work, finding that condition moderated the effect of revisiting on outcomes. Students in the distributed condition tended to revisit proximal steps, whereas those in the clustered condition tended to revisit steps while writing explanations at the end of the unit. We discuss implications and directions for further study.

Introduction

Helping students develop coherent understanding of science concepts is challenging, particularly given the fragmented curriculum in place in many schools. Research focusing on how to help students integrate their ideas has employed the Knowledge Integration (KI) framework to investigate shorter units focused on specific concepts (Linn, 2006), and recently has shifted to integrating core ideas across science disciplines, such as energy transfer and transformation (Svihla et al., 2010). One approach to helping students develop such integrated understanding is to distribute practice. Experimental laboratory studies have demonstrated benefits of distributed practice for learning a new concept or category. For instance, in one study, participants who studied paintings in a distributed manner, with other artist's paintings interleaved, performed better than those who studied in a massed manner, viewing the paintings consecutively. Interestingly, the participants in the latter condition rated it as a better learning condition, even though their performance was lower (Kornell & Bjork, 2008). This finding held for repetition tests as well (Kornell, Castel, Eich, & Bjork, 2010).

In an experiment simulating classroom practice with eight hours of video-delivered statistics lectures, a significant advantage was found for distributed practice over massed practice on a retention test given five days after the training period (Smith & Rothkopf, 1984). Smith and Rothkopf explain this finding as demonstrating that distributed practice supports learning, whereas massed practice primarily supports retrieval, making a distributed approach the more efficient approach to instruction (Smith & Rothkopf, 1984).

In a series of studies ranging from the laboratory to the classroom, Seabrook and colleagues found an advantage for distributed practice (Seabrook, Brown, & Solity, 2005). In their laboratory experiments, they found that spacing was beneficial across age groups, and introduced a clustered condition, intermediate to massed and distributed conditions and better reflecting teaching, finding that clustering was statistically similar to massed practice, and both conditions performed significantly lower compared to the distributed condition. In their classroom-based study, which involved young children learning phonics, children in the distributed condition also outperformed those in the clustered condition.

Though spacing has been recommended for the classroom (Pashler et al., 2007), few studies have sought to examine the spacing effect in real world settings. Sobel and colleagues (Sobel, Cepeda, & Kapler, 2010) conducted a study in a fifth grade classroom, in which students were taught the definitions of unfamiliar words. Students studied some words in a distributed manner, with two 15-min learning sessions separated by seven days, and studied other words in a massed manner, with two 15-min consecutive learning sessions. A vocabulary test given five weeks after the second learning session revealed a significant advantage for the words learned in the distributed condition, though in both conditions, retention was poor (Sobel et al., 2010). The classroom and classroom-like studies have not investigated the effect of distributing practice on helping students integrate their understanding of complex science content. This comparison study, set in intact classrooms, investigates the impact of different spacing—traditional or longer lag – between sessions on how well students integrated their understanding of a core science concept.

Methodological Approach

Materials

Students studied a previously-tested Web-based Inquiry Science Environment (WISE) (Slotta & Linn, 2009) unit called Global Climate Change (Svihla & Linn, 2011). This unit teaches students about the greenhouse effect and global warming. Students investigate NetLogo visualizations (Wilensky & Reisman, 2006) representing the earth and atmosphere, and variables involved in climate change. The Global Climate Change project includes a focus on energy transfer and energy transformation in global climate change processes. WISE units scaffold students using an inquiry map (Figure 1) to supports them in developing coherent understanding of science ideas (Linn, 2006). We sought to examine the impact of different lag times on both what students learned, and how they studied the CLEAR unit on Global Climate Change.

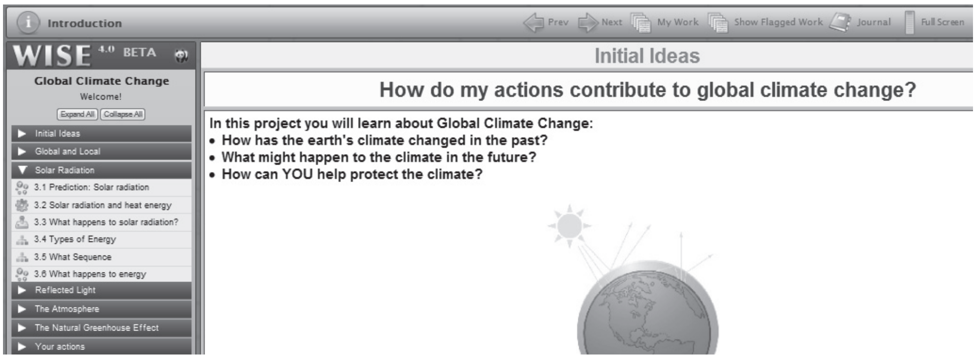

Figure 1. Inquiry map for the WISE 4 environment.

Participants

The participants of this study were students in intact sections of Chemistry and Earth Science at a suburban high school serving a diverse community. Students in the chemistry class were primarily in their junior year whereas those in Earth Science were primarily in their first year of high school. Across classes, the teacher hosted guest speakers each Wednesday, and these ranged in topic. On the following class meeting (which due to block scheduling may not have been the following day) the teacher began with a starter activity in which students first completed a worksheet with review questions about the speaker, followed by whole class discussion in which he reviewed the questions and asked students what they learned. Several of the invited speakers discussed climate change. Thus, all conditions had the opportunity to learn about climate change outside of the WISE unit during the semester. Students studied the WISE unit in dyads, but completed assessments as individuals.

Conditions and Study Design

To examine how distributed practice might be used in classroom settings with complex instruction, we compared three conditions that varied in terms of lag (N=291). In contrast to other classroom and laboratory research focusing on restudy of identical items specifically, we considered the distribution of a WISE unit in one of the following ways:

Clustered conditions. The WISE unit was used for the majority of instruction time, across consecutive class meeting times. In this context, the lag period between each class session with instructional time is two or three days, either because of intervening weekend days or block scheduling. Within this condition, two timing conditions were also designed: an *early* condition (n=31) and a *late* condition (n=91). The teacher found it challenging to schedule the early condition separately from his other courses; thus, as implemented, the early condition ended up being a mid-point between the clustered and distributed conditions, with students completing activities in the unit at their own pace, but with variable lag time between days spent on the unit (Figure 2).

Distributed condition. The WISE unit was used for a portion of instruction time each week, with sufficient time for students (n=169) to complete a given activity within the unit (Figure 2). More class periods were assigned to this condition because of numerous findings of a benefit of distributing practice.

Three individually-completed tests were given to track students' progress. The unit began with a pre-test given in week one, and ended with a post test as soon as the dyads completed the activities in the unit. A delayed post test was given at the end of the semester; thus the students in the clustered, late condition completed the delayed post four weeks after finishing the unit, the distributed condition did so five weeks after, and the clustered early did so six weeks after. Pre-, post- and delayed post items, as well as dyad responses to embedded assessments were scored using previously developed KI rubrics (Svihla & Linn, 2011). Log files of dyad progress were examined for total time spent, number of unique visits to steps within the unit, and unique visits to steps across days.

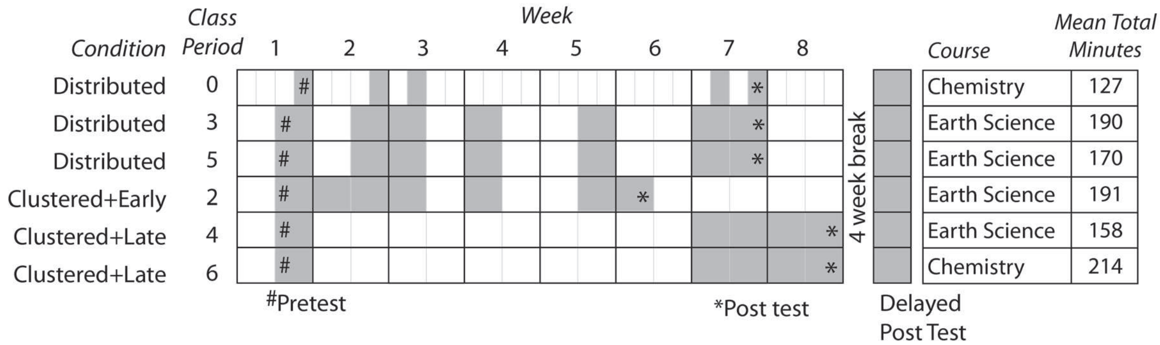

Figure 2. Implementation of conditions, with grey showing times the unit was studied

We consider two competing hypotheses in this study. Implicit in both of these hypotheses is the notion that restudying is valuable to learning. We operationalize restudying as revisiting earlier steps in the unit, but recognize the limitations this presents. There are many reasons a student might decide to revisit a step apart from deliberate restudy; however, we view any revisiting as affording an *opportunity* to restudy, and partially address the limitations by considering how and when students revisited steps. Our hypotheses are:

- The distributed condition should outperform other conditions because they have more opportunities to restudy.
- The clustered conditions should outperform the distributed condition because-- though the distributed condition provides apparent opportunities for students to restudy-- the traditional classroom context tends to fragment learning activity, such that students in the distributed condition are less likely to actually restudy.

Analysis and Results

Because this study took place in intact classrooms, we first sought to determine whether the natural clustering of class periods or the conditions explained variance in our measures. We used regression analysis to model the total scores for each time point using contrast codes; these allowed us to model variance in scores as a function of course type (chemistry or earth science), class period, and condition. When controlling for pretest, course type did not explain significant variance, t=0.99, p=.33.By comparing the models, we find that condition accounts for significant variance in most cases (Figure 3). In this figure, which summarizes a sequence of regression models, we use grey lines to depict significant differences by condition, and red lines to show significant differences by class period.

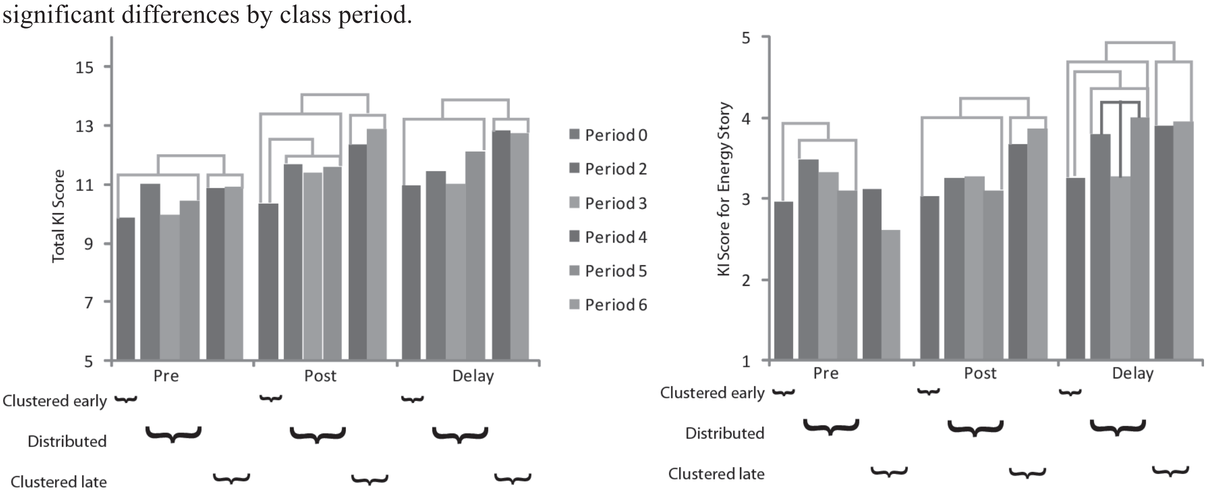

Figure 3. Scores on pre-, post- and delayed post-tests by class period, grouped within condition. Grey lines represent differences by condition; red lines by class period. In general, condition accounted for variance.

Through this modeling approach, we found that significant variance could be explained between the clustered and late condition and all other conditions for the total scores at each time point. Additionally, at the post-test, significant variance is explained by including a contrast code for clustered early and distributed conditions. No further variance is explained using class period contrast codes. For the Energy Story, contrast codes for clustered early and distributed conditions explain significant variance in KI level at the pretest and delayed post test, and contrast codes for clustered late compared to all other conditions explain significant variance in KI level at the post and delayed post time points. The only additional significant variance explained by class period within a condition was found for the delayed post test Energy Story score, between Period 3 and all other periods within the distributed condition.

We therefore primarily focus our analyses on differences by condition, rather than by class period. At all time points, condition explained significant variance in scores. We therefore continued using regression modeling to seek best-fit models using contrast codes for condition and pre- and post- test scores as we sought to explain variance in delayed post scores.

First, we modeled variance in the total KI scores from the delayed post test as a linear combination of scores from the pre-test and post-test, as well as contrast codes comparing the clustered early condition to the distributed condition, and comparing the clustered late condition to all other conditions. This explained a significant amount of variation in delayed post-test scores, $F(4, 99) = 9.26$, $p < .01$. A more parsimonious model, including only the post-test scores and the contrast codes for clustered late condition compared to all other conditions was sought while still explaining a significant amount of variation, $F(2, 101) = 17.88$, $p < .01$ (Table 1). This latter model may be interpreted as follows: In general, higher scores on the post test predict higher scores on the delayed post-test. For the mean post test score of 11.93, the delayed post test score would be predicted to be 12.61 for the clustered late condition, and 11.74 for the other conditions.

Table 1: Regression modeling of delayed post test scores

	Unstandardized Coefficients		Standardized Coefficients	
	B	Std. Error	β	*t*
Model 1: All contrast condition codes and prior scores				
Intercept	5.88	1.46		4.01**
Pre test scores	0.13	0.11	0.11	1.21
Post test scores	0.41	0.09	0.40	4.33**
Contrast: M+E v. D	-0.01	0.26	0.00	-0.05
Contrast: M+L v. others	0.26	0.14	0.18	1.88
Model 2: clustered early versus distributed omitted				
Intercept	5.88	1.46		4.03**
Pre test scores	0.13	0.11	0.11	1.21
Post test scores	0.41	0.09	0.40	4.39**
Contrast: M+L v. others	0.26	0.13	0.18	1.95
Model 3: Parsimonious model				
Intercept	7.02	1.11		6.32**
Post test scores	0.42	0.09	0.41	4.59**
Contrast: M+L v. others	0.29	0.13	0.19	2.14**

Model 1 $r^2 = .24$, r^2 change**; Model 2 $r^2 = .25$, r^2 change NS; Model 3 $r^2 = .25$, r^2 change NS. * Significant at $p<.05$; ** Significant at $p<.01$

We also modeled variance in the KI levels for Energy Stories from the delayed post test as a linear combination of Energy Story KI scores from the pre-test and post-test, as well as contrast codes comparing the clustered early condition to the distributed condition, and comparing the clustered late condition to all other conditions. This explained a significant amount of variation in delayed post-test Energy Story scores, $F(4, 79) = 3.34$, $p < .05$. A more parsimonious model, including only the pre-test scores and the contrast codes was sought while still explaining a significant amount of variation, $F(3, 80) = 4.03$, $p < .01$ (Table 2).

This latter model may be interpreted as follows: In general, higher scores on the pre test Energy Story predict higher scores on the delayed post-test. For the mean pre test score of 3.12, the delayed post test Energy Story score would be predicted to be 3.91 for the clustered late condition, 3.27 for the clustered early condition, and 3.68 for the distributed condition.

Overall, we find that the clustered late condition achieved the highest scores and the clustered early condition the lowest scores (Tables 3 & 4). The KI scores are significantly different across times, $F(2, 204) = 34.21$, $p <.01$.

For the Energy story, students generally began with a score of a three, meaning that they gave answers such as "The energy comes from the sun and goes to Earth where it warms the planet." Only the clustered late condition made gains by the post test, shifting to answers that additionally included details about how the energy was transferred to Earth (e.g., "The energy from the sun travels through space as electromagnetic waves and reaches Earth").

These findings partially support our second hypothesis, that the clustered conditions should outperform the distributed condition because though the distributed condition provides apparent opportunities for students to restudy, the classroom context tends to fragment learning activity, such that students in the distributed condition are less likely to actually restudy.

Table 2: Models of delayed post scores for Energy Stories

	Unstandardized Coefficients		Standardized Coefficients	
	B	Std. Error	β	*t*
Model 1: All contrast condition codes and prior scores				
Intercept	2.69	0.41		6.61**
Pre test scores	0.21	0.11	0.22	2.00*
Post test scores	0.09	0.08	0.12	1.11
Contrast: CE v. D	-0.24	0.12	-0.22	-2.02*
Contrast: CL v. others	0.11	0.07	0.19	1.69
Model 2: Parsimonious model with test scores omitted				
Intercept	2.94	0.34		8.80**
Pre test scores	0.22	0.11	0.23	2.13*
Contrast: CE v. D	-0.23	0.12	-0.22	-1.99*
Contrast: CL v. others	0.13	0.06	0.23	2.10*

Model 1 $r^2 = .10$, r^2 change*; Model 2 $r^2 = .10$, r^2 change NS. * Significant at $p<.05$; ** Significant at $p<.01$

Table 3: Test scores for each condition

	Pretest		Post test		Delayed Post test	
	Mean	SD	Mean	SD	Mean	SD
clustered early	9.86	1.66	10.36	2.31	10.96	2.06
distributed	10.43	1.84	11.51	1.85	11.58	2.12
clustered late	10.88	1.60	12.65	2.24	12.79	2.11

Table 4. KI scores for Energy Stories

	Pretest		Post test		Delayed Post test	
	Mean	SD	Mean	SD	Mean	SD
clustered early	2.97	0.76	3.04	0.96	3.26	0.81
distributed	3.28	0.79	3.20	1.31	3.77	0.82
clustered late	2.89	0.97	3.78	1.42	3.93	0.81

We therefore next how explore students progressed through the unit. We assumed an advantage for revisiting prior steps in the unit, and predicted students in the distributed condition would revisit steps more frequently. We also wanted to know whether students spent the same amount of time on the unit, across conditions. For these investigations, we examined log files of dyad interactions with the unit.

We computed the average number of visits to steps as the average number of unique visits to a step, provide the visit lasted five seconds or longer. This eliminated situations in which students either clicked the submit button repeatedly (each time an answer is submitted, it is recorded as a separate visit in the log file) or when the student clicked the "Next" or "Back" button in rapid succession. The latter behavior was observed during classroom observations, as when students would click "Back" repeatedly to reach a desired step, rather than navigating to it through the interface, or when playing and simply clicking back and forth through the steps. We also computed scores for the number of revisits to steps across days, predicting an advantage when students revisited a step they had studied on a previous day. Using the same criteria as described for the visits, we further calculated the number of days on which unique visits occurred, then summed the total number of steps revisited by each dyad on more than one day.

We computed the total amount of time each dyad spent logged into the unit. We used regression analysis to model these three variables of interest: the average of the total number of unique visits a dyad made to steps; the total amount of time spent by each dyad, and the number of visits to steps across days. Using the contrast codes we modeled variance in each as a function of class period, and as a function of condition. By comparing the models, we find that condition accounts for significant variance in most cases (Figure 4). In this figure, which summarizes a sequence of regression models, we use grey lines to depict significant differences by condition, and red lines to show significant differences by class period.

Both clustered conditions spent more time on the unit, and revisited steps more frequently, both across the unit in general and across days specifically. This also supports our second hypothesis, that the clustered conditions should outperform the distributed condition because though the distributed condition provides apparent opportunities for students to restudy, the classroom context tends to discourage making connections between activities over time, such that students in the distributed condition are less likely to actually restudy.

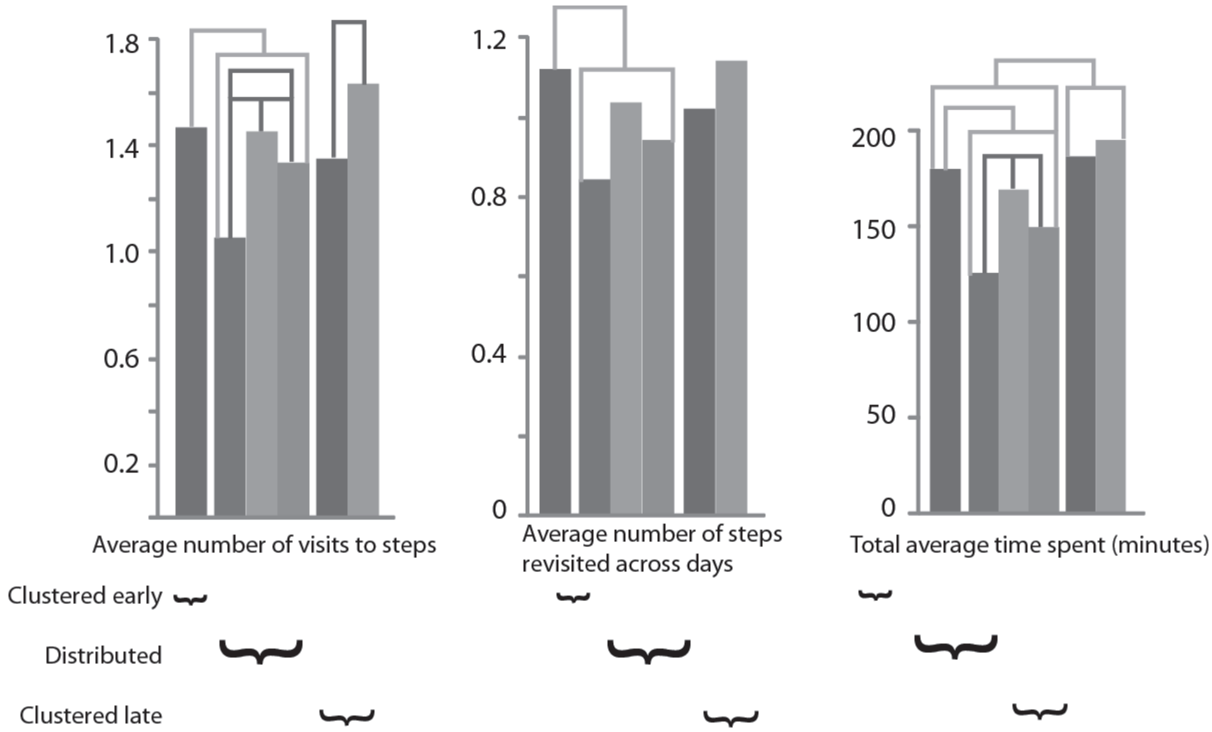

Figure 4. Amount of time and revisiting behavior of dyads during the unit. Grey lines represent differences by condition; red lines by class period. Variability in number of visits to steps is not well accounted for by condition.

To further explore revisiting of steps in the unit, we combined the steps containing visualizations. These steps also contained the majority of content, and therefore are of interest in terms of how and when students revisited them. By condition, there was no significant difference in the overall number of times students visited steps containing visualizations. However, we wished to better understand the relationship between revisiting and integrating ideas. We modeled variance in the delayed post test scores as a linear combination of pre-test and post-test scores and included revisiting across days. We focused on revisits to steps that included visualizations, because these were the steps that tend to be most difficult for students to initially understand. We further wanted to test the idea that condition moderated the revisiting activity across days. We created contrast codes and applied them to the clearest cases, the distributed condition (-1) and the clustered late (1) condition (note the semi-distributed profile of the clustered early condition, Figure 1). We then generated an interaction variable using these codes and the revisiting variable.

The full model explained a significant amount of variation in delayed post-test scores, $F(3, 81) = 10.77$, $p < .01$ (Table 4). This model may be interpreted as follows: In general, higher scores on the pre test and post test predict higher scores on the delayed post-test, with the post explaining significant variance in delayed post scores. For the distributed condition, (coded -1), less frequent revisiting of visualizations predicts higher scores; whereas for the clustered late condition (coded 1), more frequent revisiting of visualizations predicts higher scores, but this was not statistically significant. A parsimonious model omitted pre-test scores, $F(3, 81) = 10.77$, $p < .01$., and revisiting steps across days approached significance ($p<.10$).

$$\text{Delayed post test} = 6.305 + 0.49\ (\text{post test score}) + 0.48\ (\text{contrast code})(\text{revisiting score})$$

where the distributed condition is coded as -1 and the clustered late condition is coded as 1.

Table 4: Model of delayed post scores, using revisiting across days and time

	Unstandardized Coefficients		*Standardized Coefficients*	
	B	*Std. Error*	*β*	*t*
Model 1: Delayed post scores with interaction between revisiting across days and condition				
Intercept	5.519	1.669		3.306**
Pre test scores	0.09	0.12	0.07	0.76
Post test scores	0.48	0.11	0.45	4.51**
Revisit x condition	0.27	0.18	0.15	0.15
Model 2: Parsimonious model				
Intercept	6.305	1.301		4.846**
Post test scores	0.49	0.11	0.46	4.66**
Revisit x condition	0.30	0.18	0.16	1.67

Model 1 $r^2 = .28$, r^2 change*; Model 2 $r^2 = .28$, r^2 change NS. * Significant at $p<.05$; ** Significant at $p<.01$

This model suggests that though there was no significant difference between conditions in the frequency of revisiting steps containing visualizations across days, there is likely a difference in why or how they revisited steps. This would explain why revisiting may have been beneficial for one condition. It is possible that in the clustered condition, when students began a new session, they revisited steps more deliberately, because they remembered questions or confusions from the previous session; whereas in the distributed condition, they may have revisited less deliberately, with a goal to remind themselves of the unit, but not in a way that encouraged knowledge integration. While we cannot determine this with certainty, we can determine whether there are differences in the patterns of revisiting by condition. For instance, we might hypothesize students in the clustered condition would tend to revisit steps upon reaching a point of confusion, whereas students in the distributed condition would primarily revisit in the first minutes of a new session.

Preliminary results suggest that revisits across days tended to be proximal jumps within one-two steps for the distributed condition, and to occur in the beginning of a new session (Figure 5).

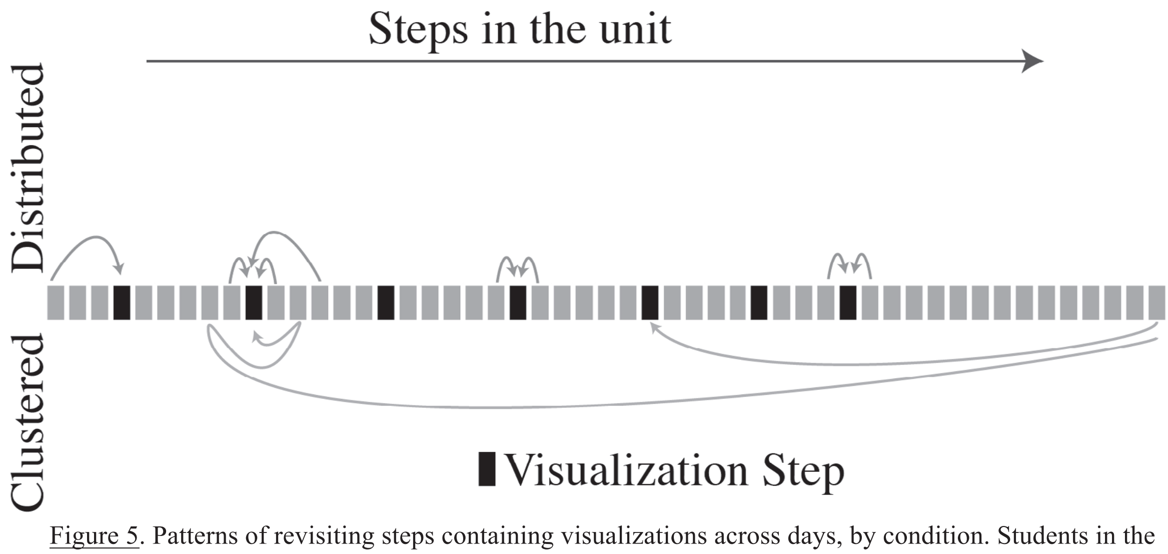

Figure 5. Patterns of revisiting steps containing visualizations across days, by condition. Students in the clustered condition revisited visualizations towards the end of the unit, but students in the distributed condition revisited proximal steps.

Students in the clustered condition tended to return to earlier steps while composing their Energy Stories or other later steps, which would have provided clearer opportunities for restudy, better supporting students to integrate their understanding.

Conclusions and Implications

In WISE, clustered practice supported the development of an integrated understanding of the role of energy in climate processes, whereas distributed practice reinforced the idea that each activity is self-contained. We interpret the further advantage for the clustered late condition as evidence that the students were also able to integrate ideas learned during the course, but outside of the unit. Ultimately, this study is not focused merely on clustered or distributed practice, but rather how they occur when there is not an emphasis on cumulative learning in classrooms.

In the distributed condition, more revisiting was associated with lower scores on the later assessments, but in the clustered condition, more revisiting was associated with higher scores on later assessments. This is explained through the analysis of log files, showing that revisiting patterns differed by condition, affording restudying that would be expected to lead to different opportunities for knowledge integration. In the distributed condition, students revisited proximal, related steps, meaning that they had the opportunity to develop a better understanding of particular ideas. In the clustered late condition, students revisited various earlier steps while writing explanations late in the unit, affording opportunities to integrate their ideas across the unit. Further study is indicated, as this work was conducted in the classrooms of a single school and using a particular curricular unit. Based on these findings, further research is warranted on possible scaffolds to support the types of revisiting observed in the clustered condition, should distributed practice be used.

References

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories. *Psychological Science, 19*(6), 585.

Kornell, N., Castel, A., Eich, T., & Bjork, R. A. (2010). Spacing as the Friend of Both Memory and Induction in Young and Older Adults. *Psychology and Aging, 25*(2), 498.

Linn, M. C. (2006). The knowledge integration perspective on learning and instruction. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 243-264). New York: Cambridge University Press.

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*, 63.

Seabrook, R., Brown, G., & Solity, J. (2005). Distributed and massed practice: From laboratory to classroom. *Applied Cognitive Psychology, 19*(1), 107-122.

Slotta, J., & Linn, M. C. (2009). *WISE Science: Web-Based Inquiry in the Classroom*. New York: Teachers College Press.

Smith, S., & Rothkopf, E. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*(3), 341-358.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2010). Spacing effects in real world classroom vocabulary learning. *Applied Cognitive Psychology*.

Svihla, V., Gerard, L., Ryoo, K., Sato, E., Visintainer, T., Swanson, H., . . . Dorsey, C. (2010, June 29 - July 2). *Energy across the Curriculum: Cumulative Learning Using Embedded Assessment Results*. Paper presented at the International Conference of the Learning Sciences, Chicago, IL.

Svihla, V., & Linn, M. C. (2011). A Design-based Approach to Fostering Understanding of Global Climate Change. *International Journal of Science Education*, 1-26. doi: 10.1080/09500693.2011.597453

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction, 24*(2), 171-209.

Acknowledgments

Though funded by the NSF (Grant #0822388), the views expressed here are not necessarily those of the NSF. We would like to acknowledge the teachers and students who participated in this research, and express thanks to other members of the research group for their feedback.