Is there an explicit learning bias?
Students beliefs, behaviors and learning outcomes

Paulo F. Carvalho (pcarvalh@andrew.cmu.edu)
Elizabeth A. McLaughlin (mimim@cs.cmu.edu)
Kenneth R. Koedinger (koedinger@cmu.edu)
Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave
Pittsburgh, PA 15213 USA

Abstract

Learning by doing refers to learning practices that involve completing activities as opposed to explicit learning (e.g., reading). Although the benefits of learning by doing have been described before, it is still relatively uncommon in instructional practice. We investigated how much students employ learning by doing in online courses, and whether it is associated with improved learning outcomes. Spending more time completing activities had a larger impact on learning outcomes than spending more time reading, even in the case of mostly declarative content, such as in a Psychology course. Moreover, learning by doing is more efficient: grade improvements of 1 standard-deviation require 10-20% less time in learning by doing than reading. Finally, we contrast this evidence with students' *a priori* intuitions on best study strategies for their online course. Students overestimate the value of explicit learning through reading, and underestimate the value of active learning.

Keywords: learning by doing; retrieval practice; self-regulated learning; doer effect

Introduction

A lot of instruction is focused on explicit learning (for example, through textbook reading, classroom lectures, and online videos). The underlying assumptions often are (a) that most knowledge we expect students to acquire in our courses is declarative in nature and, (b) perhaps, that even procedural knowledge can initially be acquired this way. Consistent with these beliefs, much emphasis has been devoted to the creation of video-based Massive Open Online Courses (MOOCs) and text-based online courses.

The emphasis on explicit learning is in stark contrast to established phenomena in cognitive psychology, advocating for the use of testing (Roediger & Karpicke, 2006) and active learning (Wieman, 2014) as better learning tools. The testing effect describes the positive effects of engaging in self-testing, instead of additional passive study (for a review see Roediger & Karpicke, 2006). This effect has been repeatedly shown in laboratory settings with diverse materials, including word pairs and text passages (e.g., Karpicke & Blunt, 2011; Karpicke & Roediger, 2008). The success of the testing effect in the laboratory led to some in-classroom studies looking at its extensibility as a tool to promote students' learning, also with positive outcomes (e.g., McDaniel et al., 2007).

Why might active practice not be used in the classroom? One possibility is that the effect is limited to controlled laboratory contexts in which other aspects of real-world instruction do not vary. There is currently a lack of evidence from large-scale classroom studies demonstrating the benefits of testing over reading outside the lab. Another possibility is that the advantage of learning by doing is specific to some types of materials (e.g., procedural knowledge), thereby limiting its use by instructors and students. Indeed, there is some evidence showing that, under some circumstances, additional passive reading practice, compared to doing activities, might result in better learning (Sweller & Cooper, 1985). In the KLI framework, Koedinger and colleagues (2012) postulate that the learning goals and the nature of the materials being studied are the determining forces behind whether reading or doing are better for improving learning.

In sum, conceivably, learning by doing is not used because it is not effective in real-world contexts or across a wide range of knowledge types. Moreover, there is an underlying assumption that when the focus of learning is declarative knowledge, the emphasis should be on reading activities that would foster the formation of connections between concepts and the creation of robust declarative knowledge (Anderson & Schunn, 2000). Nonetheless, learning by doing is important because most of human expertise involves tacit knowledge of the cues and conditions for deciding when, where, and what knowledge to bring to bear in complex situations (Zhu et al., 1996). In this view, there might be no verbal shortcut to acquiring expertise; it might be best acquired by repeated practice.

In our research, we explore whether learning by doing is a better way of learning across different types of knowledge (i.e., declarative and procedural), and whether it is more efficient. We compare learning outcomes of students enrolled in two online courses as a function of frequency and time spent completing practice activities (doing) vs. reading.

Students' study behavior and their beliefs about the best study strategy

Even if a class is designed to encourage students to learn by doing, including extensive self-testing and guided practice activities but minimal text, it is an open question whether students a) realize its potential, b) use it, and c) whether self-directed learning by doing in the classroom is as effective as its guided counterparts in the laboratory. These questions are theoretically and practically important because previous research on other cognitive approaches to improve learning have repeatedly shown a difference in outcomes between when students are in control of their study and when they are not (Carvalho et al., 2016; Ciccone & Brelsford, 1976), as well as a lack of awareness by the students on how to best organize their study (Karpicke et al., 2009).

Koedinger et al. (2015, 2016) illustrated the power of learning by in the context of online courses used in real classrooms. The Open Learning Initiative (OLI) at Carnegie Mellon University (CMU) is a learning environment that includes several courses each focusing on rich and interactive learn-by-doing activities, aligned with student-centered learning outcomes, and designed around science-based learner models. By analyzing student self-regulated study behavior in online classes taught at different universities using OLI materials, Koedinger et al. (2015, 2016) identified a "doer effect" – completing more practice activities is a stronger predictor of student performance than completing more reading activities.

The present work

The present work builds on early evidence of the "doer effect" and extends it. One explanation for why completing more doing activities has a larger impact on learning than completing more reading activities is that completing doing activities may be more time intensive. If students devote more time to studying, regardless of how they do it, they are more likely to learn more. In other words, more learning results from the time devoted to an activity (e.g., reading or doing), not from the activity itself. Conversely, if learning by doing is more beneficial because it engages students in an active learning process (Wieman, 2014), it should be associated with better learning outcomes even if students spend comparatively less time engaging in that activity. To investigate this question, we compare the time spent reading and doing, and its relative impact on learning outcomes. Is it the case that reading for longer periods results in better learning outcomes than doing for shorter periods?

Additionally, to probe the generalizability of learning by doing even for declarative knowledge, we investigate students' behavior in two courses. An introductory psychology course focusing mostly on declarative knowledge and a computation course focusing on both declarative and procedural (learning how to code) knowledge.

Finally, we investigate students' beliefs about the usefulness of using learning by doing in their study. At the start of each course, as part of an optional unit, students completed a question on what they thought was the best strategy to study for the course. Are students' *a priori* beliefs on how to study biased towards explicit learning (i.e., reading)?

The "doer effect" in a Psychology MOOC

Method

Sample. Our analyses include data from 783 students enrolled in an online "Introduction to Psychology as a Science" MOOC offered by the Georgia Institute of Technology through Coursera. We included in the analyses students registered in OLI for whom pretest, quizzes and the final exam data were available. For a description of the entire sample see Koedinger et al. (2015).

Description of the course. The course "Introduction to Psychology as a Science" was designed as a 12-week introductory survey course, and is often taught during the first year of college. For each week of class, the course targeted a major topic area (e.g. Memory, Sense and Perception, Abnormal Behavior). Elements of CMU's Open Learning Initiative (OLI) "Introduction to Psychology" course were incorporated into Georgia Tech's "Introduction to Psychology as a Science" MOOC. OLI materials including text and interactive activities were available to students, in addition to the lectures, quizzes and other Coursera-based activities of the larger course. Each sub-topic was supported by a pre-recorded video lecture (10-15 minutes, with downloadable slides) and included matched modules and learning outcomes in the OLI learning environment. A high-stakes quiz assessed students against these outcomes at the end of each week.

The OLI modules included a variety of expository content (text, examples, images, and video clips) and a large number of interactive activities. Broadly, these activities serve two purposes. "Learn By Doing" activities, intended to support student outcome achievement, provide feedback and robust hints to support students. Figure 1 shows an example of a "Learn by Doing" activity from the Personality module covered in week 9 of the course. Another type of activity, "Did I Get This" activities, provided a self-comprehension check for students. These activities were created in conjunction with the OLI text materials and complement it by providing testing ("Did I Get This") or active learning ("Learn by Doing") activities that cover the concepts described in the text.

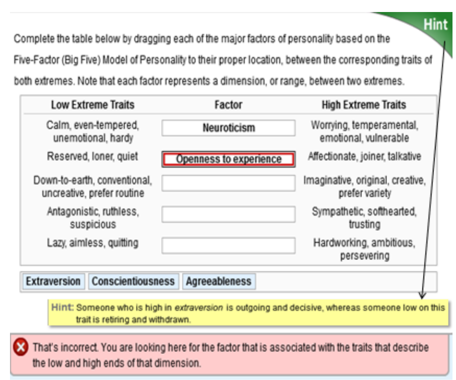

Figure 1: Screenshot of an OLI "Learn By Doing" activity from the module on Personality.

Research questions, measures, and analysis plan. We explore three main research questions: Q1: Does completing more practice ("doing") activities, compared to completing more reading activities, predict better learning outcomes?; Q2: Does spending more time on practice ("doing") activities, compared to time spent on reading activities, predict better learning outcomes?; Q3: What are students' beliefs regarding "best study strategies" for an online course?

To approach these questions, we created the following analyses plan. First we calculated our dependent measures:

number of reading and doing activities and total time spent in each. We started by identifying for each student the number of doing and reading activities. A doing activity was identified as responding to at least one practice activity in the OLI. A reading activity was identified as opening a text webpage in the OLI. Because opening a text webpage does not necessarily mean a student was reading, we adjusted the total number of activities by removing extremely short reading activities (below the 10^{th} percentile of reading time for all reading activities for that student). Our reasoning is that when students took extremely short amounts of time in a page, they might not have in fact read the associated text.

Table 1: Study strategies students could choose from. The metacognitive activity in the Psychology MOOC course included only the first 4 strategies. The Computing OLI course included all strategies in this table.

Strategy	Description
Game the System	Do activities without reading text. Select different answer choices until correct.
Do-Read	Do first activity and if cannot answer, read relevant text.
Read-Do	Read text and complete do activities as they come up
Read	Read text, skip doing activities.
Do	Complete some activities, then go back to text and look for a similar example to read.

Because doing activities were contained inside webpages, often surrounded by text, time spent doing and reading for each activity/page was inferred from recorded data as follows. Timestamps were recorded for when a student opened a OLI page, when a student made a choice in each step of an activity (e.g., selected an option in a multiple-choice question), checked their responses in activities, asked for hints in the activities, and closed the page. From these logs, we could infer doing time as the time difference between the initial step and the final step of an activity. All other time spent (from opening to closing) in the page was considered reading time. However, this process does not include the time spent reading the doing activity text before starting the activity. To correct for this, the time spent completing each doing activity also includes a proportion of the time right before completing the first step (the other portion being classified as time spent reading). For reading time, we calculated the difference between the time when a webpage was initially accessed and the time an activity was started, as well as the time between an activity was finished and the another one started or the webpage was closed plus the portion of the time immediately before the first step of each activity. However, initial analyses of the time spent in each page revealed a number of large outliers (several standard deviations above the student mean for the student). These times might be indicative that a student left the webpage opened while completing other activities (potentially not related to the course). To correct for this, in addition to removing very short reading times (for consistency with the number of reading activities analysis above), we also replaced very long reading times (above the 90^{th} percentile for that student) with the average reading time for that student. This way, we hope to reduce the influence of situations during which the student had the page opened but was not actively reading the text presented.

Our dependent measures included the summed quiz score across the 11 quizzes and the final exam score, all multiple-choice questions. Each quiz was worth 10 points. The final exam had 35 questions (each worth 1 point). To account for differences in student prior knowledge, we entered pretest score as a predictor in the models. The pretest, completed at the start of the course, was composed of multiple-choice questions from content covered in most of the units of the course and was graded from 0-20 points.

We converted the raw scores for the independent measures of student behavior as well as the dependent measures into standardize *z*-scores for ease of comparison across measures. We analyzed the effect of each independent measure on each dependent measure separately using a logistic regression model (in *R* code):

zQuiz[zExam] = lm(zPretest+zNumDoAct [zTotalDoTime] + zNumReadAct[zTotalReadTime] + zNumDoAct[zTotalDoTime]*zNumReadAct[zTotalReadTime], data = oli_do_read)

Finally, to identify students' beliefs regarding best studying strategies (Q3), we took the students' responses in an activity during the "Learning Strategies" module included in the beginning of the OLI course. In this optional module (not included in the other analyses), students were introduced to several key research findings in learning sciences, and "best strategies" to achieve best learning. In one of the activities included in this unit students were asked to choose which of four study strategies they thought would yield best results in the course (see Table 1). To describe students *a priori* study strategy judgments, we calculated the proportion of students who chose each of these alternatives before starting their study in the course. Each student could choose one or more of the options, out of the four offered: "Game the system", "Do-Read", "Read-Do", and "Read" (see Table 1).

Results and Discussion

Table 2: Descriptive statistics for the main measures of students' study behavior and independent measures in the Psych MOOC

	M (*SD*)	Median	25^{th} Prctl.	75^{th} Prctl.
Read Time (mins)	9408 (5377)	9091	6080	12265
Doing Time (mins)	833 (748)	478	244	1240
#Read Activities	287 (210)	245	156	384
#Doing Activities	435 (265)	541	152	683
Pretest	11 (3.5)	11	9	13
Quizzes	89 (18.3)	94	83	101
Final Exam	27 (5.8)	28	24	31

Descriptive measures of student behavior. As it can be seen in Table 2, students spent on average more time reading than doing (9000 min vs. 800 min, respectively); conversely students completed more doing than reading activities (435 vs. 287, respectively). This overall descriptive data is consistent with the nature of the OLI course, which included a large number of short doing activities and text passages.

Q1: More "doing" activities predicts better learning outcomes. Results of the logistic regression predicting Quiz and Exam performance using number of doing and reading activities are presented in Table 3.

The regression analysis showed that higher quiz and exam scores are predicted by completing a larger number of doing activities (β = 0.40, p < .0001 and (β = 0.24, p < .0001, respectively), and by completing more reading activities (β = 0.11, p = .001 and β = 0.11, p =.03, respectively). Importantly, the relative benefit of completing more doing activities was 2.4 to 3.6 times larger than completing more reading activities.

Overall, these results support those found by Koedinger et al. (2015, 2016), showing that completing more doing activities predicts better learning outcomes to a greater degree than completing more reading, even when we correct for the existence of very short (potentially off-task) reading events.

Finally, contrary to some intuitive predictions of the complementary nature of the two types of learning activities, their positive effect on learning outcomes are not additive. Completing more doing activities is more beneficial when students completed less reading activities (and vice-versa; β = -0.15, p < .0001 and β = -0.04, p =.40, respectively).

Table 3: Results of logistic regression for both courses. Coefficients are standard deviations from the mean (*z-scores*).

		Quiz					Exam				
Course	DV	Adj R^2	Doing Coef.	Reading Coef.	Interact Coef.	Effect Ratio	Adj R^2	Doing Coef.	Reading Coef.	Interact Coef.	Effect Ratio
Psych MOOC	Number Activities	.29	0.40 (0.04)	0.11 (0.04)	-0.15 (0.04)	3.6	.14	0.24 (0.05)	0.10 (0.05)	-0.04 (0.04)	2.4
	Total Time	.19	0.39 (0.04)	0.11 (0.04)	-0.19 (0.03)	3.5	.11	0.19 (0.05)	0.13 (0.04)	-0.09 (0.03)	1.5
Computing OLI	Number Activities	.42	0.56 (0.02)	0.23 (0.02)	-0.16 (0.06)	2.43	.08	0.28 (0.02)	0.05 (0.02)	-0.05 (0.02)	5.6
	Total Time	.10	0.19 (0.03)	0.27 (0.03)	-0.03 (0.003)	0.70	.02	0.15 (0.04)	0.08 (0.03)	-0.02 (0.003)	1.9

Q2: More time in doing activities predicts better learning outcomes. The results of the logistic regression predicting Quiz and Exam performance using total time doing and reading are also presented in Table 3. The regression analyses showed that higher quiz and exam scores were predicted by spending more time doing (β = 0.39, p < .0001 and β = 0.19, p <.0001, respectively), as well as reading (β = 0.11, p = .006 and β = 0.13, p =.001, respectively).

Importantly, because reading requires, on average, more time than doing (see mean and standard deviations in Table 3), for each 1 standard-deviation (18.3 points, 17% total score) improvement in the total quiz score, students had to complete only a total of 18.45 hours of doing work during the 12 weeks of the course (or 1.5 hours/week), but 166.22 hours of reading work during the same period (or 13.8 hours/week). Similar improvements in final exam score require 16.16 hours of doing work but 168.77 hours of reading work over the entire course. Finally, similarly to what we saw when analyzing number of activities completed, spending more time completing doing activities is more beneficial when students spend less time reading (and vice-versa; β = -0.19, p < .0001 and β = -0.09, p =.005, respectively). This result further indicates that the benefits of the two types of activity is not additive.

Q3: Students overestimate the benefits of reading. Only a subset of students (N = 389) from the original sample described above also completed the "Learning Strategies" module (the module was optional). Table 4 shows the percentage of students who chose each possible study strategy as well as the percentage of students who chose exclusively each option. As it can be seen from the table, the large majority of students (93%) chose "reading and completing the activities as they appear" ("Read-Do") as the best strategy. In fact, Read-Do was the most popular as the exclusive choice. Did students who chose a strategy focused on learning by doing spend more time doing than reading? To evaluate this question, we looked at the relative time spent doing vs. reading depending on the strategy the student chose.

Table 4: Percentage of students who selected each strategy as best for learning in the course.

Course	N	Game the system		Do-Read		Read-Do		Read		Do	
		Selected	Only selection	Selected	Only selection	Selected	Only selection	Selected	Only selection	Selected	Only Selection
Psychology MOOC	389	8%	8%	32%	5%	93%	61%	6%	0.05%	N/A	N/A
Computing OLI	950	3%	0%	36%	4%	94%	40%	4%	0%	36%	0.05%

For each student, we calculated the difference between total time doing and total time reading (doing-reading). More positive values in this measure indicate more time doing relative to time spent reading. We compared students who chose only the strategy "Do-Read", those who chose that strategy and another strategy, and those who chose any other strategy. Students who chose only the "Do-Read" strategy (M = -7202, SD = 6139), or that strategy in addition to another (M = -7694, SD = 4414), spent relatively more time completing doing activities than those who did not choose that option (M = -9746, SD = 4983; t (386) = 2.238, p = .026 and t (386) = 3.63, p < .0001, respectively).

The "doer effect" in an online Computing Course

One of the goals of this research was to investigate whether engaging in learning by doing is an effective learning strategy for different types of knowledge. To extend the nature of the types of knowledge covered, we ran the same analyses with data from students' study behavior in an online version of a computing course. The content of this course is substantially different from the more expositive nature of an introductory psychology course. The course design followed the same overall principles and was similar to the Psychology course in terms of number of activities available to the students (see Koedinger et al., 2016 for details).

Sample. Our analyses include data from 2261 students enrolled in the online computing course "Information Systems" at University of Maryland University College (UMUC) using the OLI platform. We included in the analyses below students registered in OLI for whom quiz scores and a final grade were available. No pretest was available in this course.

Research questions and analyses plan. The same research questions and analyses plan as for the Psychology MOOC were used. The dependent measures used were the percentage correct across all quizzes and the final grade in number (1-5). Regression models do not include a pretest score.

Results and Discussion

Descriptive measures of student behavior. Similar to what we found in the Psychology MOOC course, students spent on average more time reading than doing; conversely students completed more doing than reading activities (see Table 5).

Q1: More "doing" activities predicts better learning outcomes. Better quiz and exam scores are predicted by completing more doing activities (β = 0.56, p < .0001 and β = 0.27, p <.0001, respectively), as well as more reading activities (β = 0.22, p < .0001 and β = 0.05, p =.02, respectively; see Table 3). Moreover, we found similar ratios of benefit of doing over reading (2.43-5.6) as in the Psych MOOC, as well as a counter-intuitive interaction whereby the effect of doing activities is greater for lower amounts of reading, but only when predicting quiz scores (and vice-versa; β = -0.16, p < .0001 and β = -0.05, p =.02, respectively).

Table 5: Descriptive statistics for the main measures of students' study behavior and independent measures in the Computing course

	M (SD)	Median	25th Prctl.	75th Prctl.
Read Time (mins)	13714 (21948)	4679	613	19481
Doing Time (mins)	2830 (7447)	234	26	1749
#Read Activities	30 (31)	24	14	36
#Doing Activities	70 (55)	64	14	130
Percent Correct Quizzes	7.98 (2.8)	8.5	5.82	10.44
Final Grade	3.89 (1.25)	4	3	5

Q2: More time in doing activities predicts better learning outcomes. Better quiz and exam scores are predicted by spending more time completing doing activities (β = 0.19, p < .0001 and β = 0.14, p <.0001, respectively), as well as more time reading (β = 0.27, p < .0001 and β = 0.08, p =.01, respectively; see Table 3). There is also an interaction, whereby the positive effect of more time spent in doing activities is larger when students spend less time reading (and vice-versa; β = -0.03, p < .0001 and β = -0.02, p <.0001, respectively). Although for the quiz scores we see a larger impact of more reading time compared to more doing time (as evidence by a ratio smaller than 1), for both quiz and exam scores it is clear that spending more time completing doing activities is more beneficial and efficient because it takes on average less time to complete more doing activities and this has an impact on performance. For example, for a 1 standard deviation (2.8%) improvement in quiz scores, students would have to spend 70.9 hours over the duration of the course completing doing activities, but a whopping 326.13 hours reading – a gain of more than 20%.

Q3: Students overestimate the benefits of reading. Among the subset of students who completed the question on what they believed was the best learning strategy (N = 950), the large majority of students indicated that they should read all the text and complete all activities as they show up (94%, see Table 4). Only a small number of students indicated that they should focus mostly on the doing activities (36%). Moreover, the students' *a priori* strategy preference did not predict their relative time spent doing, F (2, 938) < 1, p = .558, demonstrating that even students who completed more doing activities are probably unware of its benefits.

General Discussion

The results of this research indicate that self-regulated learning by doing is associated with larger learning gains than learning by reading. More importantly, besides being a desirable learning strategy, it might also be more efficient. Across two different online courses focusing on different types of content, we found that students who completed more doing activities showed larger learning gains in shorter time (between 10 and 20% less time to achieve similar improvements). This result is important for two reasons: (1)

it emboldens efforts to include more active, doing activities in lessons, as an alternative to reading activities, and (2) it shows the generalizability of learning by doing to different kinds of materials, even materials often thought of as involving declarative, as opposed to procedural, knowledge.

Learning by doing as described here involved effortful (Roediger & Karpicke, 2006), active engagement and knowledge manipulation by the student (Wieman, 2014), with timely feedback (Roediger & Karpicke, 2006). All these properties have been associated with better learning outcomes compared to passive learning situations such as reading. Any of these factors might have contributed to the benefits of spending more time completing doing activities. Interestingly, the benefits of learning by doing were larger when students spent less time reading, suggesting that the two types of activity might be non-additive. An interesting hypothesis for future research is whether learning by doing could replace some or all of the learning that takes place from reading. Can effective learning of declarative knowledge be done exclusively by doing with feedback?

Importantly, we found that students do not realize the potential of learning by doing. Students seem to overestimate the value of explicit, verbal, learning and underestimate the value of active learning, as seen by their overwhelming support for strategies that emphasize reading and weak support for strategies that emphasize doing. Similar dichotomies between best learning outcomes and students *a priori* judgements of best study practices have been described before (see Roediger & Karpicke, 2006), and underscore the important role of familiarizing students with empirically tested best-practices.

Finally, the naturalistic character of the data and the approach used here have great potential. Natural datasets (such as the two used in this investigation) are increasingly available and allow for a wider investigation of the generalizability, effectiveness and adequacy of learning methods, theories, and approaches developed in the laboratory. This approach can play a key role for the future of learning science because of the novel insights that can only be gained from studying how learning takes place in natural contexts by their natural agents (Jones, 2016). However, admittedly, the research presented here does not allow us to establish causal links or discriminate between alternative theories of why learning by doing is a more efficient learning strategy. It is possible that the differences in learning by completing reading and doing activities presented here are due to a third variable; though previous research suggests that might not be the case (Koedinger et al., 2016). Nonetheless, the research presented here can stimulate future controlled studies that establish causal links, and investigate which characteristics of learning by doing in classroom contexts contribute to its benefits.

Acknowledgments

This work was supported by a National Science Foundation grant (ACI-1443068) toward the creation of LearnSphere.org and by funding from Google. All data presented here is available from LearnShepere.org's DataShop (https://pslcdatashop.web.cmu.edu).

References

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. *Advances in Instructional Psychology*, *5*, 1–34.

Carvalho, P.F., Braithwaite, D.W., de Leeuw, J.R., Motz, B.A., & Goldstone, R.L. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PLoS ONE*, *11*(3).

Ciccone, D.S., & Brelsford, J.W. (1976). Spacing repetitions in paired-associate learning: Experimenter versus subject control. Journal of Experimental Psychology: Human Learning and Memory, 2(4), 446.

Jones, M. N. (2016). Developing cognitive theory by mining large-scale naturalistic data. In M. N. Jones (Ed.), *Big Data in Cognitive Science.* New York: Taylor & Francis.

Karpicke, J.D., & Blunt, J.R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772-775.

Karpicke, J.D., & Roediger, H.L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966-968.

Karpicke, J.D., Butler, A. C., & Roediger III, H.L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own?. *Memory*, *17*(4), 471-479.

Koedinger, K.R., Corbett, A.T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757-798.

Koedinger, K.R., Kim, J., Jia, J.Z., McLaughlin, E.A., & Bier, N.L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale* (pp. 111-120). ACM.

Koedinger, K.R., McLaughlin, E.A., Jia, J.Z., & Bier, N.L. (2016). Is the doer effect a causal relationship? How can we tell and why it's important. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 388-397). ACM.

McDaniel, M.A., Anderson, J.L., Derbish, M.H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4), 494-513.

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210.

Sweller, J., & Cooper, G.A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction*, *2*(1), 59-89.

Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message. *Proceedings of the National Academy of Sciences, 111*(23), 8319-8320.

Zhu, X., Lee, Y., Simon, H. A., & Zhu, D. (1996). Cue recognition and cue elaboration in learning from examples. *Proceedings of the National Academy of Sciences*, *93*(3), 1346-1351.