

PeerWise: Evaluating the Effectiveness of a Web-Based Learning Aid in a Second-Year Psychology Subject

Psychology Learning & Teaching

2018, Vol. 17(2) 166–176

© The Author(s) 2018

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1475725718764181

journals.sagepub.com/home/plat**Piers D. L. Howe**  and **Meredith McKague**

Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia

Jason M. Lodge

School of Education, Faculty of Humanities and Social Sciences, The University of Queensland, St Lucia, Queensland, Australia.

Anthea G. Blunden and Geoffrey Saw

Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Australia

Abstract

Testing can do more than just determine what a student knows; it can aid the learning process, a phenomenon known as the testing effect. There is a growing trend for students to create and share self-assessment questions in their subject, as advocated by the contributing-student pedagogy (CSP). For subjects with large enrolments, this process can be facilitated by educational technology. PeerWise is an example of such technology. It is free, web-based software that allows students to author, share, answer, and provide feedback on multiple-choice quizzes in a collaborative and constructivist fashion. While it is popular, it is unclear to what degree it facilitates student learning. To evaluate its effectiveness, we introduced PeerWise into a second-year psychology subject. We measured the extent to which it increased scores in the final exam. We found that PeerWise did significantly increase exam scores, so was a useful learning aid.

Keywords

Computer-based learning, contributing-student pedagogy, collaborative learning, online learning, interactive learning environments, learning communities

Introduction

In higher education, assessments can be either summative or formative (Boud, 2000; Nicol & Macfalane-Dick, 2007). Summative assessments are designed to evaluate what students

Corresponding author:

Piers D. L. Howe, Melbourne University, 12th Floor Redmond Barry Building, Parkville, Victoria 3010, Australia.

Email: pdhowe@unimelb.edu.au

know. They are typically administered at the conclusion of the subject or course and will usually provide quantitative feedback, allowing the students to be ranked relative to their peers (Boud, 2000; Nicol & Macfalane-Dick, 2007). Conversely, formative assessments are primarily designed to aid the learning process (Black & Wiliam, 2009; Nicol & Macfalane-Dick, 2007). They are typically used to indicate shortcomings in the knowledge of the student in order to guide subsequent learning (Black & Wiliam, 2009; Nicol & Macfalane-Dick, 2007). In particular, they can help the students identify which of their preconceptions need to be refined (Smith, diSessa, & Roschelle, 1993). Students do not start a subject as a blank slate. Rather, they bring to the subject a range of knowledge and preconceptions. Crucially, these preconceptions are not necessarily misconceptions (Clement, Brown, & Zietsman, 1989). Rather, they often represent an unrefined or oversimplified understanding of a topic that can serve as a useful starting point for their subsequent learning (Clement et al., 1989). Formative assessments help to identify in which ways these preconceptions require refinement and can further be used throughout a subject to aid students in progressively building their understanding. Formative assessments can also give feedback on how effective the student's learning strategies have been, and can motivate the student to learn more in the future (Pastotter, Schicker, Niedernhuber, & Bauml, 2011).

While formative assessment has been identified as a critical practice for enabling student progress, it inevitably exists within a broader pedagogical approach. Recent innovations emphasize a more interactive role for students. Whereas in the traditional teaching model the instructor prepares and administers learning resources, in the contributing-student pedagogy (CSP), the students share responsibility for creating learning resources (Hamer et al., 2008; Hamer, Sheard, Purchase, & Luxton-Reilly, 2012). This pedagogy is based on the philosophy that students learn best when they are actively engaged in helping to create their own learning resources, and draws on constructivist (Hamer et al., 2008) and socio-cultural constructivist (Ben-Ari, 2001) theories of learning. CSP activities will often involve students creating assessments for their peers, sharing solutions and feedback with each other, and sometimes reviewing the work of their classmates (Vygotsky, 1978). While this may be achieved using only pen and paper, in subjects with larger enrolments, CSP will usually rely heavily on computer-based technologies to administer and provide feedback on formative assessment tasks.

PeerWise (<http://peerwise.cs.auckland.ac.nz>) is an example of web-based software that allows students to create and share formative self-assessments in a convenient manner (Denny, Hamer, Luxton-Reilly, & Purchase, 2008; Denny, Luxton-Reilly, & Hamer, 2008a). As such it facilitates CSP. It is free and widely used. In brief, it works as follows (Luxton-Reilly & Denny, 2010): The lecturer creates a PeerWise website and gives the students access to it. The students can then create, share, and answer multiple-choice questions (MCQs). When a question is created, the author of the question designates which of the potential answers is correct and provides a written explanation to act as feedback. Once the question has been attempted, the answerer receives this feedback and then has the opportunity to rate the question based on its difficulty and quality, as well as provide general comments. These comments can form the basis of an online discussion where the veracity of the question can be debated with other students, including the author of the question. This allows useful questions to be identified and permits those questions whose answers are potentially incorrect or ambiguous to be flagged. The correct answer can then be provided by other students. A central aim of PeerWise is to guide subsequent learning by the student. By revealing in which areas the student's knowledge is incomplete or otherwise needs

refinement, PeerWise helps the student focus their learning more effectively. It is in this sense that we regarded PeerWise as formative (Black & Wiliam, 2009) and well aligned with CSP.

Testing is a powerful means of improving learning (Roediger & Karpicke, 2006a, 2006b). Indeed, there is growing evidence that it is one of the most effective ways of aiding student learning (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). For example, taking a test has been shown to improve retention more than spending an equivalent amount of time restudying the material, even when test performance is poor and no feedback is given (Roediger & Karpicke, 2006a). This *testing effect*, whereby engaging in effortful retrieval improves memory for the material studied, has been widely demonstrated in laboratory settings (Roediger & Butler, 2011). However, testing may not always be beneficial (Nguyen & McDaniel, 2015; Roediger, Putnam, & Smith, 2011). For example, practice testing with MCQs can result in students memorizing the false lures (Marsh, Roediger, Bjork, & Bjork, 2007), although this effect can be reduced or eliminated by supplying feedback (Butler & Roediger, 2008). It has also been reported that testing some material may not increase performance on untested material, if the two sets of materials are not closely related (Mayer et al., 2009; Wooldridge, Bugg, McDaniel, & Liu, 2014). Indeed, testing only part of the material can increase the forgetting of the untested portion (Anderson, Bjork, & Bjork, 1994; Little, Storm, & Bjork, 2011). Finally, it has been suggested that the benefits of testing decrease or even disappear when the complexity of the material to be learned increases (Van Gog & Sweller, 2015). For these reasons, it was not a foregone conclusion that PeerWise would assist students with their learning. Furthermore, even if it were to increase student learning, instructors would still need to verify that this increase was large enough to justify the time students spent using it.

While some studies have indeed shown that using PeerWise is correlated with higher marks in exams taken at the end of the subject (Bates, Galloway, & McBride, 2012; Bottomley & Denny, 2001), these studies did not attempt to control for student aptitude. Here we are using the term “aptitude” to refer to how easily and quickly a student can learn the subject matter. In particular, if two students are equally motivated, use the same revision strategy and spend the same amount of time learning, the one with the greater aptitude will learn more. When determining how effective PeerWise is as a learning aid, it is necessary to control for student aptitude. In particular, it could be that the students with a greater aptitude for the subject tend to use PeerWise more. Since these students would also tend to perform better in the exam taken at the end of the subject, this could give rise to a spurious correlation between PeerWise usage and exam performance, giving the impression that PeerWise contributes more to learning than it really does (Luxton-Reilly, 2012).

The most rigorous way to address this concern would be to randomly divide students into cohorts and allow only some cohorts access to PeerWise. Student aptitude could then be decoupled from PeerWise usage. However, this is generally considered unethical as it would mean denying some students in a subject a potentially beneficial learning aid while allowing other students in the same subject access to it. As a compromise option, Humpage (2014) adopted a quasi-experimental design. She conducted a 6-year study of a sociology subject where PeerWise was made available for only two of those years. Comparing student outcomes across all 6 years, she concluded that there was little evidence that the introduction of PeerWise was associated with any improvement in student performance.

McQueen, Shields, Finnegan, Higham, & Simmen (2014) argued that PeerWise is helpful for some students. They studied the usage of PeerWise in a second-year genetics subject over a period of 3 years. For each year, they divided their students into four quartiles based on

their performance in a prior genetics subject. Within each quartile, they performed a median split to identify those students with high PeerWise activity (HPA) and low PeerWise activity (LPA). They reported mixed findings. In the majority of their comparisons (7/12), they found that the HPA group did not significantly outperform the LPA group. In the remaining comparisons, they found a 4–6% improvement in the overall marks for the subject. However, it is unclear to what extent even this increase can be attributed to PeerWise. Because they operationalized student aptitude as performance on a *prior* genetics subject, they were unable to measure the aptitude of students for the *current* subject. In particular, they were unable to determine whether PeerWise did improve performance or whether students who were going to perform better anyway happen to be more likely to engage with PeerWise, thereby giving the false impression that using PeerWise resulted in better performance in the final exam.

There are several reasons why students with greater aptitude for the subject material might be more likely to engage with PeerWise. For example, such students might derive more satisfaction from learning, so might readily engage with all learning aids. Additionally, these students would be more likely to get the answers correct, which might encourage them to remain engaged with PeerWise. Since aptitude is likely to be correlated with exam performance, if the students with greater aptitude tend to engage more with PeerWise this would cause PeerWise usage to be correlated with exam performance, even if PeerWise itself was not an effective learning aid. Therefore, to test whether PeerWise is effective, it is necessary to control for student aptitude. While correlation can never prove causation (Aldrich, 1995), we can better test if participation with PeerWise does indeed improve final exam scores if we use partial correlations to control for student aptitude.

There were a number of different ways students could participate in PeerWise: they could write questions, comment on questions, and answer questions. Our initial plan was to investigate to what degree each way of utilizing PeerWise increased final exam scores. However, as we discuss later, relatively few students either wrote or commented on questions, so we decided to quantify PeerWise usage solely as the number of questions a student answered.

In summary, different studies have come to conflicting conclusions as to whether PeerWise facilitates learning. While one multi-year study concluded that PeerWise does not improve student performance (Humpage, 2014), another study found that it might do so for some students (McQueen et al., 2014). However, for the latter study, it is unclear to what extent this improvement is due to the students with greater aptitude for the subject matter tending to use PeerWise more.

The purpose of our study was to answer the following question: To what extent does utilizing PeerWise by answering questions increase final test performance? To address this question, we needed to measure the degree to which PeerWise facilitates learning while controlling for student aptitude. We measured student aptitude by measuring performance on two assignments. Crucially, these assignments were part of the same subject as the one to which PeerWise was applied (Luxton-Reilly, 2012). This allowed us to better discount student aptitude when measuring the degree to which PeerWise facilitates student learning.

Method

Participants

To assess the effect of answering questions on PeerWise on final exam performance, we introduced it into a second-year psychology subject, *Biological Psychology*, which was taught

at an Australian University. The subject had two essay-based assessments that were due during the 12-week semester and a multiple-choice final exam that occurred shortly after the end of the semester. We excluded from our study any student that did not complete all the assessments. In total, 647 students attended the subject and completed all assessments. Out of these, 387 (60%) participated in PeerWise. While PeerWise does not collect demographics, based on subject enrolment, more than 90% of the students would have been 18–21 years old, and approximately 75% would have been female. Most would be Caucasian, but a large minority would have been south-east Asian.

Materials

Each student completed two essays, answered one or more questions on the PeerWise website, and sat the final exam. Thus, for each student, we had the marks for the two essays, the number of questions he/she authored, answered and commented on in PeerWise, and his/her score in the final exam.

Procedure

We constructed a PeerWise account, which was made available to students at the start of the second week of the semester. For ethical reasons, it was made available to all students enrolled in the subject. Because we were unsure whether PeerWise would be an effective learning aid, participation in it was not mandatory. However, students were encouraged to use it on a number of occasions.

The two essays were marked using a points-based marking scheme. Essentially, there was a list of criteria that each essay needed to satisfy and a number of points was assigned to each criterion. For example, some of the criteria covered the key concepts that each essay needed to explain. Other criteria evaluated how well students followed APA formatting guidelines. Double marking of a subset (approximately 10%) of the essays ensured that all essays were being marked in a consistent manner. The essays were on a topic that was covered only in the tutorials and was not the focus of any of the lectures. As the PeerWise questions were focused solely on the lecture content and not on the tutorial content, the PeerWise questions did not relate to the essays. As such, the essay marks represent a useful, independent measure of student aptitude for the subject in question.

In the final exam, there were 120 MCQs. These questions covered the lecture content, so covered the same material as the PeerWise questions. However, because the PeerWise questions were constructed by students, they were written independently of the final exam.

Design and Analysis

As discussed above, it is possible that the students with greater aptitude for the subject matter might be more likely to use PeerWise. Since these students are likely to do better in the final exam, this may cause a positive correlation between the exam scores and participation in PeerWise, even if PeerWise were not an effective learning aid (Luxton-Reilly, 2012). Following the lead of Luxton-Reilly (2012), we used performance on the two assignments during the subject to give a measure of each student's subject-specific aptitude.

For each subject, we therefore measured the partial correlation between the level of participation in PeerWise and the student's exam mark, controlling for the student's subject-specific aptitude (as measured by their separate marks on the two assignments). We also

performed a linear regression where we attempted to predict students' exam scores based on their two essay scores and their level of participation in PeerWise. This regression allowed us to determine the degree to which PeerWise participation affected the final exam score, separate from student aptitude.

Results

In total, 176 questions were written and 134 comments were made for this subject. Out of the 387 students who completed all assessments and participated with PeerWise, only 28 wrote questions. Most of these students did not write many questions, whereas a small proportion wrote a large number. Consequently, for these students, the modal number of contributed questions was two whereas the mean was 6.3. Forty-three of the 387 students commented on at least one question, and all 387 students answered at least one question with each, on average, answering 92.2 questions. Given the relatively low number of students authoring questions or writing comments, we did not include these two measures in our analysis. This, unfortunately, meant that we could not quantitatively investigate the extent to which creating or commenting on MCQs aids learning and affects exam performance.

Figure 1 shows the marks for the two assignments and the exam as a function of whether or not the student participated in PeerWise. For the two assignments, students who participated in PeerWise scored higher than those that did not (Assignment 1: $t(645)=4.47$, $p < .001$, $r^2 = .030$; Assignment 2: $t(645)=5.98$, $p < .001$, $r^2 = .052$). Consequently, we need to take into account student aptitude when attempting to determine to what degree participating in PeerWise increases final exam scores.

We calculated the Pearson's partial correlation between the degree of participation in PeerWise and the participant's final exam score, controlling for their subject-specific aptitude, as operationalized as performance on the two assignments. The result was statistically significant ($r(645) = .250$, $p < .001$), demonstrating that participation in PeerWise is positively associated with exam performance, even when student aptitude is controlled for, with participation in PeerWise uniquely accounting for approximately 6.3% of the total variance in the final exam score. We then performed a linear regression to determine to

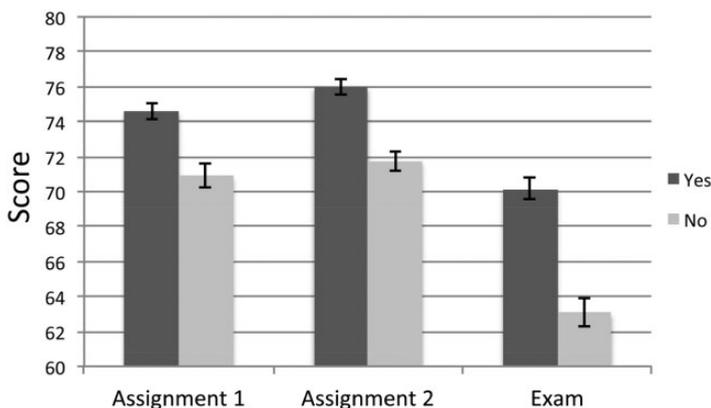


Figure 1. The mean assignment marks as a function of whether the student participated ("yes") or did not participate ("no") in PeerWise. Error bars represent the standard error of the mean.

what extent answering questions in PeerWise increased the final exam score. Specifically, we used linear regression to predict the final exam score based on the student's performance in the two assignments and their level of participation in PeerWise. The results of the regression are shown in Table 1. This regression model accounted for a significant amount of the variance in the final exam score ($r^2 = .323$, $F(3,643) = 102.2$, $p < .001$). According to this model, each question answered in PeerWise increased the final exam score by 0.04%.

Discussion

Self-assessments can aid learning, and it is becoming increasingly common for students to create and share these assessments, as advocated by the CSP (Luxton-Reilly, 2012). PeerWise is an online resource that can facilitate this process (Denny, Hamer, et al., 2008; Denny, Luxton-Reilly, et al., 2008a). Although it has proven popular with students (Denny, Luxton-Reilly, & Hamer, 2008b), it was unclear whether it increases student learning and, if so, whether the extent to which it increases student learning justifies the time students spend on it. The purpose of our study was to address this concern. We found that the number of questions that a student answered in PeerWise was positively correlated with their final exam score even when subject-specific student aptitude was controlled for. A linear regression revealed that each question answered increased the final exam score by 0.04%. This means that if a student were to answer all 176 questions that were posted on PeerWise, then the predicted increase in their final exam score would be 7.0%. Given that we would expect students to take approximately three hours to answer these questions (based on the fact that in the final exam they are required to answer 120 questions in two hours), this would appear to be an efficient use of their time.

Our results add to the growing literature that PeerWise is an effective learning aid. A concern with much of this literature was that it did not control for student aptitude. A notable exception to this trend was a study by McQueen et al. (2014) who quantified student aptitude as performance on a prior subject. Like us, they concluded that PeerWise was an effective tool for formative assessment, at least for some students. Combined with our findings, this suggests that regardless of whether student aptitude is defined as performance on a prior subject or performance on non-MCQ assessment tasks in the current subject, one can still find evidence that usage of PeerWise increases final exam scores, controlling for student aptitude.

The McQueen et al. (2014) study investigated the effectiveness of PeerWise in the context of a second-year genetics subject. Conversely, we investigated its effectiveness in the context of a second-year psychology subject. The fact that both studies found PeerWise to be an effective tool for formative assessment suggests that finding is likely to generalize to other

Table 1. The coefficients of the regression fit. "Number of Questions Answered" denotes how many PeerWise MCQs were answered.

	B	SE B	β	t	p
Constant Term	12.7	3.60		3.53	<.001
Assignment 1	0.304	0.049	.252	6.25	<.001
Assignment 2	0.406	0.056	.290	7.19	<.001
Number of Questions Answered	0.040	0.006	.217	6.54	<.001

second-year science subjects. However, it is possible that PeerWise may not be as effective in other contexts. For example, Humpage (2014) concluded that there was no compelling evidence that PeerWise was an effective learning aid for a first-year sociology subject.

As noted earlier, PeerWise relies on the testing effect and there is evidence that the testing effect may not apply to more complex subject matter (Van Gog & Sweller, 2015). Van Gog and Sweller found that testing was most effective as a learning aid for material containing distinct elements, where the individual elements could be learned independently and without reference to the other elements. They described such material as having low element interactivity. The subject that was the focus of the present study, *Biological Psychology*, was an introductory survey subject that was designed to cover a large number of different topics, at a relatively superficial level of detail. The final exam was focused on testing individual facts, which mostly could be learned without reference to each other. As such, it would seem to be the sort of subject for which one would expect testing to be an effective learning aid (Van Gog & Sweller, 2015).

Other subjects may contain material with much higher element interactivity. In particular, the subject for which Humpage (2014) investigated the effectiveness of PeerWise, *Social Policy, Social Justice*, involved material that was complex, interrelated, and nuanced. Using the terminology of Van Gog and Sweller, we could say that her material had high element interactivity. Such material places less emphasis on fact-based learning, for which testing is an effective revision strategy, and more emphasis on integrative, critical thinking, which Humpage argued was less facilitated by PeerWise's multiple-choice format. In particular, she argued that the multiple-choice format may have encouraged students to overly simplify the material. The findings of Humpage (2014) suggest that PeerWise may not be effective for all subjects.

So far, we have been agnostic as to the mechanism by which PeerWise aids learning. It could be that improved learning occurred only for the items that were tested, what Roediger and Karpicke (2006a) call "direct" testing effects. If true, then training with PeerWise would increase performance for those questions in the final exam that happened to be similar to or identical to questions in PeerWise (recall that the students who wrote the questions for PeerWise had no knowledge of the questions that would appear on the final exam). However, it has also been reported that memory for the untested items may also improve, what is often described as an "indirect" testing effect (Roediger & Karpicke, 2006a). It has been suggested that indirect testing effects are caused by the participation in the test activating other related information, thereby causing the participant to review that information as well (Roediger & Karpicke, 2006a). However, testing can also increase learning for unrelated information that is studied *after* the test is administered (Szpunar, McDermott, & Roediger, 2008). How this occurs is still unclear (Roediger & Karpicke, 2006a). It seems to be due to both improvements in encoding (Chan, 2009) and retrieval (Pastotter & Bauml, 2014), possibly by encouraging the participants to pay more attention during the subsequent learning process (Pastotter & Bauml, 2014). So, it could be that PeerWise improves learning by encouraging students to pay more attention when reviewing their notes subsequent to using it (Pastotter et al., 2011).

In conclusion, we have found that PeerWise is an effective learning aid in a second-year psychology subject. Not only did PeerWise participation result in a statistically significant improvement in exam scores, but the effect size was large enough to make its usage worthwhile for the students. This supports previous work that has also found PeerWise to be a useful learning aid (McQueen et al., 2014). However, it is possible that these results will only

hold for subject matter with low element interactivity (Pastotter et al., 2011; Van Gog & Sweller, 2015) and may not be appropriate for material that has higher element interactivity (Humpage, 2014).

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by a Learning Teaching Initiative grant from the University of Melbourne (no grant number).

ORCID iD

Piers D. L. Howe  <http://orcid.org/0000-0001-6171-1381>

References

- Aldrich, J. (1995). Correlations genuine and spurious in Pearson and Yule. *Statistical Science*, 10(4), 364–376.
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1063–1087.
- Bates, S. P., Galloway, R. K., & McBride, K. L. (2012). *Student-generated content: Using PeerWise to enhance engagement and outcome in introductory physics courses*. College Park, Maryland, USA: AIP.
- Ben-Ari, M. (2001). Constructivism in computer science education. *Journal of Computers in Mathematics and Science Teaching*, 20(1), 45–73.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–21.
- Bottomley, S., & Denny, P. (2001). A participatory learning approach to biochemistry using student authored and evaluated multiple-choice questions. *Biochemistry and Molecular Biology Education*, 5, 352–361.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 24(2), 151–167.
- Butler, A. C., & Roediger, H. L. I. (2008). Feedback enhances the positive effects and reduces the negative effects. *Memory & Cognition*, 36, 604–616.
- Chan, J. C. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language*, 61, 153–170.
- Clement, J., Brown, D. E., & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding ‘anchoring conceptions’ for grounding instructions on students’ intuitions. *International Journal of Science Education*, 11(5), 554–565.
- Denny, P., Hamer, J., Luxton-Reilly, A., & Purchase, H. (2008). PeerWise: Students sharing their multiple choice questions. *Proceedings of the Fourth International Workshop on Computing Education Research*, 51–58.
- Denny, P., Luxton-Reilly, A., & Hamer, J. (2008a). The PeerWise system of student contributed assessment questions. *Proceeding of the tenth conference on Australasian computing education*, 78, 69–74.

- Denny, P., Luxton-Reilly, A., & Hamer, J. (2008b). *Student use of the PeerWise system*. Paper presented at the 13th Annual Conference On Innovation and Technology in Computer Science Education, Madrid, Spain.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–48.
- Hamer, J., Cutts, Q., Jackova, J., Luxton-Reilly, A., McCartney, R., Purchase, H., . . . Sheard, J. (2008). Contributing student pedagogy. *ACM SIGCSE Bulletin*, *40*(4), 194–212.
- Hamer, J., Sheard, J., Purchase, H., & Luxton-Reilly, A. (2012). Contributing student pedagogy. *Computer Science Education*, *22*(4), 315–318.
- Humpage, L. (2014). PeerWise: A useful learning tool for sociology? *New Zealand Sociology*, *29*(1), 135.
- Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory*, *19*, 346–359.
- Luxton-Reilly, A. (2012). *The design and evaluation of StudySieve: A tool that supports student-generated free-response questions, answers and evaluations*. (PhD), The University of Auckland.
- Luxton-Reilly, A., & Denny, P. (2010). Constructive evaluation: A pedagogy of student-contributed assessment. *Computer Science Education*, *20*(2), 145–167.
- Marsh, E. J., Roediger, H. L. I., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, *14*(2), 194–199.
- Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., . . . Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*(1), 51–57.
- McQueen, H. A., Shields, C., Finnegan, D. J., Higham, J., & Simmen, M. W. (2014). PeerWise provides significant academic benefits to biological science students across diverse learning tasks, but with minimal instructor intervention. *Biochemistry and Molecular Biology Education*, *42*(5), 371–381.
- Nguyen, K., & McDaniel, M. A. (2015). Using quizzing to enhance student learning in the classroom: The good, the bad and the ugly. *Teaching of Psychology*, *42*, 87–92.
- Nicol, D., & Macfalane-Dick, D. (2007). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199–218.
- Pastotter, B., & Bauml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology*, *5*, 1–5.
- Pastotter, B., Schicker, S., Niedernhuber, J., & Bauml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *37*, 287–297.
- Roediger, H. L. I., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science*, *15*(1), 20–27.
- Roediger, H. L. I., & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210.
- Roediger, H. L. I., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.
- Roediger, H. L. I., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. *Psychology of Learning and Motivation*, *55*, 1–36.
- Smith, J. P., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of Learning Sciences*, *3*(2), 115–163.
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. I. (2008). Testing during study insulates against a buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*, 1392–1399.

- Van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decrease or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, 27(2), 247–264.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wooldrige, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory and Cognition*, 3, 214–221.

Author biographies

Piers D. L. Howe is currently an Associate Professor and the Director of Teaching and Learning in the School of Psychological Sciences at the University of Melbourne. His research focuses on Situation Awareness, Decision-Making, Behavioral Change and Teaching and Learning. He currently teaches Biological Psychology, Visual Perception and Mathematical Models of Decision-Making. As some of his class sizes are very large (~1000 students), he has an interest in the evaluation of computational teaching aids, such as PeerWise.

Meredith McKague is currently a Senior Lecturer (Teaching Specialist) and is the Convener of Academic Innovation in the School of Psychological Sciences at the University of Melbourne. Her research focuses on the cognitive processes underlying word recognition and reading skills, bilingualism, and teaching and learning. Meredith teaches throughout the undergraduate psychology curriculum, in the areas of Learning and Cognition, and she coordinates Cognitive Psychology, and the Capstone subjects.

Jason M. Lodge is a Principal Research Fellow in the ARC-SRI Science of Learning Research Centre and Associate Professor of Educational Psychology in the School of Education, The University of Queensland. Jason's research focuses on the cognitive, meta-cognitive and affective foundations of conceptual learning. He investigates how we might support students to achieve conceptual change and to self-regulate their development of conceptual understanding in digital environments. Jason is coordinator of the Australian Psychological Society Psychology Education Interest Group.

Anthea G. Blunden is a Teaching Specialist and PhD candidate at the University of Melbourne. Her work focuses on visual perception, categorization, and decision-making. She teaches courses in Biological Psychology, Cognitive Psychology, and Organizational Psychology and has an interest in collaborative, peer-driven learning for both teachers and learners in higher education.

Geoffrey Saw is a Teaching Specialist at the University of Melbourne. His research focuses on Teaching and Learning, Decision-Making, and the Wisdom of the Crowd effect. He lectures in Cognitive Psychology and Research Methods and has a strong interest in research into higher education pedagogical techniques.