



Delft University of Technology

Activating Learning at Scale

A Review of Innovations in Online Learning Strategies

Davis, Dan; Chen, Guanliang; Hauff, Claudia; Houben, Geert-Jan

DOI

[10.1016/j.compedu.2018.05.019](https://doi.org/10.1016/j.compedu.2018.05.019)

Publication date

2018

Document Version

Final published version

Published in

Computers & Education

Citation (APA)

Davis, D., Chen, G., Hauff, C., & Houben, G.-J. (2018). Activating Learning at Scale: A Review of Innovations in Online Learning Strategies. *Computers & Education*, 125, 327-344. <https://doi.org/10.1016/j.compedu.2018.05.019>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' – Taverne project

<https://www.openaccess.nl/en/you-share-we-take-care>

Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.



Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Activating learning at scale: A review of innovations in online learning strategies

Dan Davis*, Guanliang Chen, Claudia Hauff, Geert-Jan Houben

Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands



ARTICLE INFO

Keywords:

Teaching/learning strategies
 Adult learning
 Evaluation of CAL systems
 Interactive learning environments
 Multimedia/hypermedia systems

ABSTRACT

Taking advantage of the vast history of theoretical and empirical findings in the learning literature we have inherited, this research offers a synthesis of prior findings in the domain of *empirically evaluated active learning strategies in digital learning environments*. The primary concern of the present study is to evaluate these findings with an eye towards scalable learning. Massive Open Online Courses (MOOCs) have emerged as the new way to reach the masses with educational materials, but so far they have failed to maintain learners' attention over the long term. Even though we now understand how effective active learning principles are for learners, the current landscape of MOOC pedagogy too often allows for passivity — leading to the unsatisfactory performance experienced by many MOOC learners today. As a starting point to this research we took John Hattie's seminal work from 2008 on learning strategies used to facilitate active learning. We considered research published between 2009 and 2017 that presents *empirical evaluations* of these learning strategies. Through our systematic search we found 126 papers meeting our criteria and categorized them according to Hattie's learning strategies. We found large-scale experiments to be the most challenging environment for experimentation due to their size, heterogeneity of participants, and platform restrictions, and we identified the three most promising strategies for effectively leveraging learning at scale as Cooperative Learning, Simulations & Gaming, and Interactive Multimedia.

1. Introduction

In the dense landscape of scalable learning technologies, consideration for sound pedagogy can often fall by the wayside as university courses are retrofitted from a classroom to the Web. Up against the uncertainty of how to best rethink and conceive of pedagogy at scale, we here synthesize the previous findings as well as highlight the possibilities going forward with the greatest potential for boosting learner achievement in large-scale digital learning environments.

Now that the initial hype of Massive Open Online Courses (MOOCs) has passed and the Web is populated with more than 4000 of these free or low-cost educational resources, we take this opportunity to evaluate and assess the state-of-the art in pedagogy at scale while identifying the best practices that have been found to significantly increase learner achievement.

This study conducts a review of the literature by specifically seeking innovations in *scalable* (not requiring any physical presence or manual grading or feedback) learning strategies that aim to create a more active learning experience, defined in Freeman et al. (2014) as one that “engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work.” By limiting the selection criteria to

* Corresponding author.

E-mail addresses: d.j.davis@tudelft.nl (D. Davis), g.chen-2@tudelft.nl (G. Chen), c.hauff@tudelft.nl (C. Hauff), g.j.p.m.houben@tudelft.nl (G.-J. Houben).

empirical research that can be applied at scale, we aim for this survey to serve as a basis upon which future MOOC design innovations can be conceived, designed, and tested. We see this as an important perspective to take, as many learning design studies provide design ideas, but do not contain a robust empirical evaluation. We certainly do not intend to discount the value of observational or qualitative studies in this domain; rather, for the following analyses we are primarily concerned with results backed by tests of statistical significance because this offers a more objective, quantitative measure of effectiveness.

2. Method

The driving question underpinning this literature survey is: **Which active learning strategies for digital learning environments have been empirically evaluated, and how effective are they?**

To begin the literature search we utilized John Hattie's *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement* Hattie (2008) as a basis. It provides a comprehensive overview of findings in the domain of empirically tested learning strategies in traditional classroom environments. As Hattie's work was published in 2008, we used that as a natural starting point for our review, working forward to July 2017. It creates a narrow enough scope (nine years: 2009–2017) and temporally relevant (MOOCs went mainstream in 2012) time constraints for the review. We manually scanned all publications released from our selected venues in this time period and determined for each whether or not they met our criteria: (1) the learning strategy being analyzed must have been **scalable** — it must not require manual coding, feedback, physical presence, etc., (2) the evidence must come from empirical analyses of **randomized controlled experiments** with a combined **sample size of at least ten** across all conditions, and (3) the subjects of the studies must be **adult learners**, i.e. at least 18 years old. We included the age criterion based on the profile of the typical MOOC learner — aged 25–35 according to Tomkin and Charlevoix (2014), which aligns with our own institution's data as well.

From Hattie's synthesis of meta-analyses we identified the 10 core learning strategies that best apply to open online education — only selecting from those which Hattie found to be effective. With these learning strategies fixed, we systematically reviewed all publications in five journals and eight conferences (listed in Table 1) that have displayed a regular interest in publishing work on testing these categories of innovative online learning strategies. These venues were identified and selected based on an exploratory search through the literature—we began with a sample of studies we were previously familiar with that fit the scope of the present review and perused the references of each to identify more potential venues worth exploring. This process was repeated for each identified study thereafter. The lead author also reached out to experts in the field to assure that this method did not overlook any potential venues. The thirteen venues used for the final review are those which showed the most consistent interest in publishing studies that meet our criteria. We employed this method over a search/query term method because our criteria (namely that of being a randomized controlled trial among adult populations) are not reliably gleanable from standard search engine indexing.

We acknowledge there are other journals and conference proceedings that may have been applicable for this survey, but given our search criteria, we found these thirteen venues to be the most appropriate based on our initial exploratory search.

Of the 7706 papers included in our search, we found 126 (1.6%) to meet our criteria. The criterion requiring randomized controlled trials proved to be a strong filter with many studies not reporting randomization or a baseline condition to compare against. Overall, these 126 papers report on experiments with a total of 132,428 study participants. We then classified each work into one of the ten learning strategy categories (listed in Table 2).

Fig. 1 illustrates the number of studies that met our selection criteria organized by the year published. It shows the increasing frequency of such experiments in recent years, with the most notable increase from 2014 to 2015.

We could propose any number of explanations for the decrease in studies from 2015 to 2016, but it would be purely speculation. However, when examining the studies themselves, we do notice a prominent trend with some explanatory power. With the dawn of MOOC research emerging around 2013 and 2014, the experiments carried out in this window can be viewed now, in hindsight, as foundational. Such interventions in this era included sending out emails to learners (Kizilcec, Schneider, Cohen, & McFarland, 2014a) or dividing the course discussion forum and controlling instructor activity (Tomkin & Charlevoix, 2014). However, in 2016 and 2017 we begin to see an elevated level of complexity in interventions such as the *adaptive and personalized* quiz question delivery system

Table 1

Overview of included venues. The most recent included issue from each publication is indicated in parentheses. Unless otherwise indicated with a †, the full proceedings from 2017 have been included.

| |
|--|
| Computers & Education (Vol. 114) |
| Journal of Learning Analytics (Vol. 4, No. 2) |
| Journal of Educational Data Mining (Vol. 8, No. 2) |
| The Open Education Journal eLearning Papers (Issue 43) |
| IEEE Transactions on Learning Technologies (Vol. 10, Issue 1) |
| ACM Learning @ Scale (L@S) |
| Learning Analytics & Knowledge (LAK) |
| European Conference on Technology-Enhanced Learning (EC-TEL) † |
| International Conference on Educational Data Mining (EDM) |
| ACM Conference on Computer-Supported Cooperative Work (CSCW) |
| European MOOCs Stakeholders Summit (EMOOCs) |
| European Conference on Computer-Supported Collaborative Work (ECSCW) |
| Human Factors in Computing Systems (CHI) |

Table 2

Overview of considered learning categories. The selected papers per category are shown in parentheses. The sum of the numbers is 131 and not 126, as five papers apply to two categories.

| |
|--|
| <i>Mastery Learning</i> (1) |
| <i>Meta-Cognitive Strategies</i> (24) |
| <i>Questioning</i> (9) |
| <i>Spaced vs. Massed Practice</i> (1) |
| <i>Matching Learning Styles</i> (3) |
| <i>Feedback</i> (21) |
| <i>Cooperative Learning</i> (17) |
| <i>Simulations & Gaming</i> (18) |
| <i>Programmed Instruction</i> (6) |
| <i>Interactive Multimedia Methods</i> (31) |

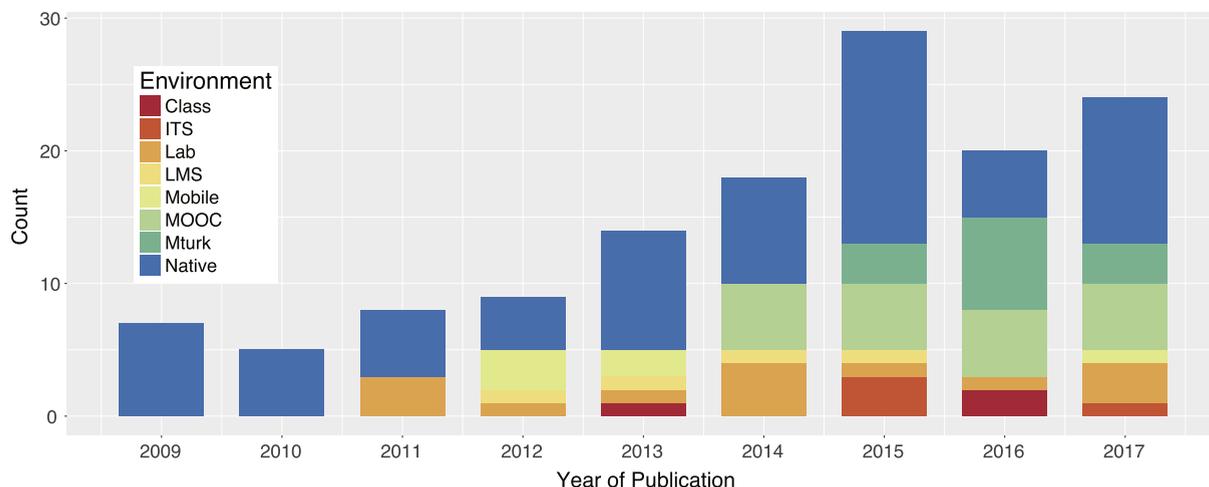


Fig. 1. The number of papers by year and learning environment meeting our selection criteria. Each environment is defined in detail in Section 3.1. Best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Rosen et al., 2017) implemented and evaluated at scale in a MOOC. It is also worth noting that a number of journal issues and conference proceedings from 2017 had not yet been released at the time of this writing (indicated in Table 1).

Fig. 2a shows the proportion of results (positive, null, or negative) with respect to the experimental environment employed by the selected articles/studies. Noting the difference between MOOCs and native environments (those designed and implemented specifically for the study), we see native environments yielding positive results at a much stronger rate than MOOCs (59% vs. 42% respectively). We see two main factors contributing to this difference: (i) native environments can be modeled specifically for the experiment/tested concepts, whereas experiments done in MOOCs must adapt to the existing platforms and (ii) no MOOC studies provide participants any incentive to participate, whereas this is common to experiments in native environments.

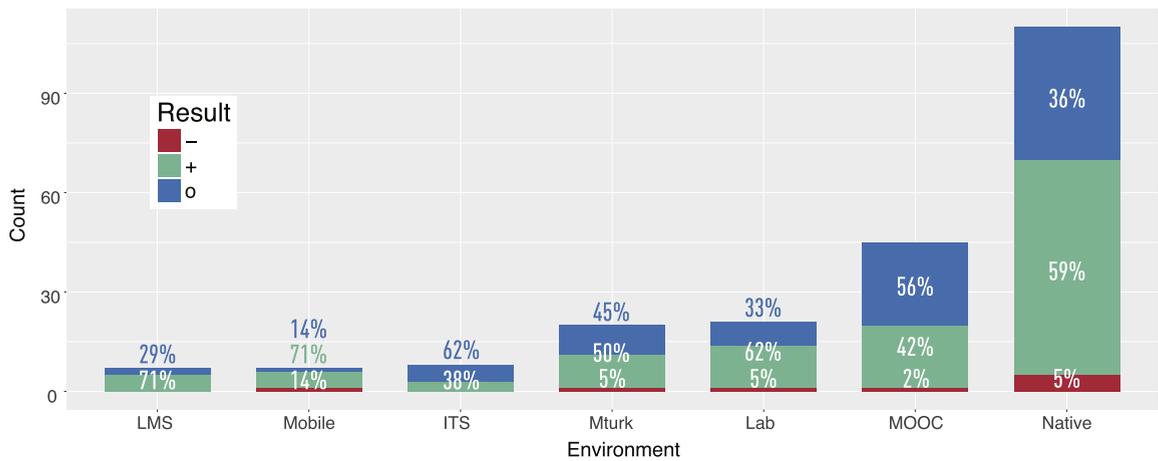
Fig. 2c further visualizes this discrepancy in illustrating the proportion of positive, negative, and null results across three subject pool sizes: small-scale studies with between 10 and 100 participants, medium-sized studies with 101–500 participants and large-scale studies with more than 500 study participants. We here find a statistically significant difference in the proportion of reported positive findings in large (42% in studies with 500 + participants) and small (60% in studies with 10–100 participants) studies using a χ^2 test ($p < 0.05$). As the focus of this study is on large-scale learning, we specifically ran this analysis to evaluate the impact that scale and, in turn, sample size and heterogeneity can have on an experiment.

We registered this project with the Center for Open Science,¹ and the registration which includes all data gathered as well as scripts used for analysis & visualization are available at <https://osf.io/jy9n6/>.

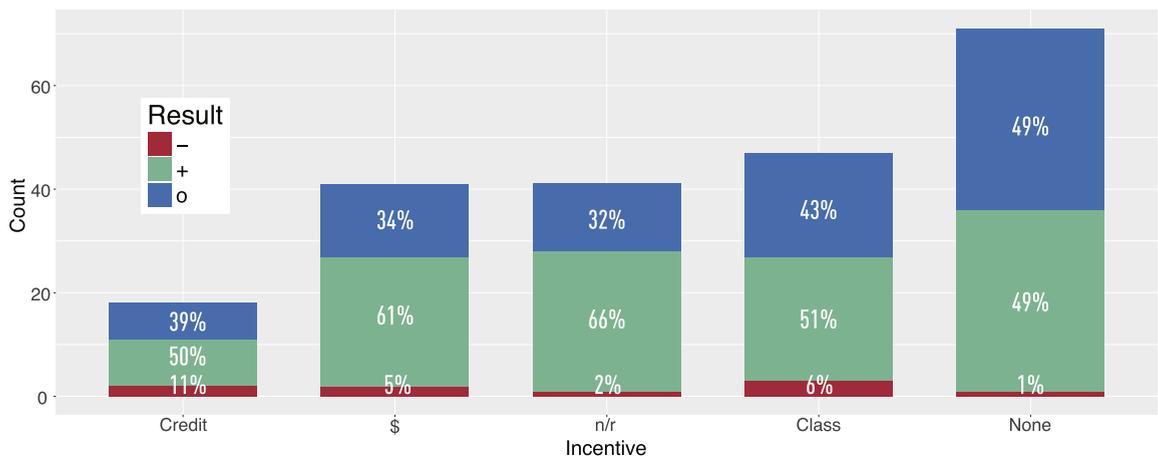
3. Terminology

We now define the terminology used in the reporting of our results. Not only is this explicit terminology elucidation important for the clarity of this review, it can also serve as a reference for future experiments in this area to ensure consistency in how results are reported and replicated. In discussing each study, we refer to “learners”, “students”, or “participants” as the authors do in the referenced work.

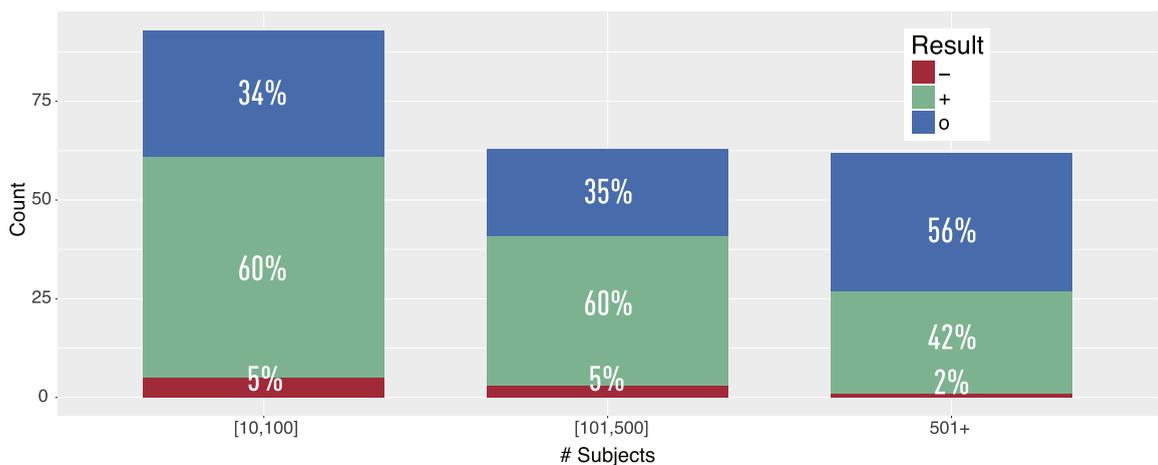
¹ <https://cos.io>.



(a) Papers partitioned by environment



(b) Papers partitioned by incentive.



(c) Papers partitioned by number of study participants.

Fig. 2. Reported results (y-axis) from papers meeting our selection criteria partitioned by environment, incentive, and size. The red - indicates studies reporting a significant negative effect of the intervention; the green + indicates a significant positive effect of the intervention; and the blue o indicates findings without a statistically significant effect. Best viewed in color. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.1. Environment

The first dimension by which we categorize the studies is the environment wherein the experiment/intervention took place. We distinguish between the following:

- **Intelligent Tutoring System (ITS):** Digital learning systems that monitor and adapt to a learner's behavior and knowledge state.
- **Laboratory Setting (Lab):** Controlled, physical setting in which participants complete the experimental tasks.
- **Learning Management System (LMS):** Software application used to host & organize course materials for students to access online at any time.
- **Mobile Phone Application (Mobile):** Participants must download and use an application on their mobile phone to participate in the experiment.
- **Massive Open Online Course (MOOC):** Online course which offers educational materials free of cost and with open access to all.
- **Amazon Mechanical Turk (Mturk):** Online marketplace used to host low-cost micro-payment tasks for crowdsourcing and human computation. Participants are recruited and paid through MTurk and often redirected to an external application.
- **Native:** A piece of software designed and implemented specifically for the study.

Fig. 1 shows the breakdown of our studies with respect to the environment. Note that despite the widespread availability of MOOC and LMS environments in 2015, native environments still dominated that year. We speculate that this may be because researchers find it more efficient to build their own environment from scratch rather than adapt their study to the limitations of a pre-existing platform—which is the case with all MOOC experiments included in this study; each intervention had to be designed within the confines of either the edX² or Coursera³ platforms. We also note a sudden spike in popularity for studies using Mturk from 2015 to 2016. While it is more expensive to carry out research with Mturk compared to MOOCs (which provide no incentive or compensation), Mturk ensures a certain level of compliance and engagement from the subjects in that they are rewarded for their time with money.

3.2. Incentive

The second dimension we distinguish is the incentive participants in each study received for their participation:

- **Monetary Reward (\$):** Participants receive either a cash reward or a gift certificate.
- **Required as part of an existing class (Class):** An instructor conducts an experiment in her own course where all enrolled students are participants.
- **Class Credit (Credit):** By participating in the study, participants receive course credit which can be applied to their university degree.
- **None:** Participants were not provided any incentive or compensation.
- **n/r:** Not reported.

3.3. Outcome variables

As experiments on learning strategies can evaluate a multitude of outcomes, here we provide an overview of all learning outcomes reported in the included studies.

- **Final Grade:** the cumulative score over the span of the entire course which includes all graded assignments.
- **Completion Rate:** the proportion of participants who earn the required final passing grade in the course.
- **Learning Gain:** the observed difference in knowledge between pre-treatment and post-treatment exams
- **Exam Score:** different from the final grade metric in that this only considers learner performance on one particular assessment (typically the final exam).
- **Long-Term Retention:** measured by assessing a learner's knowledge of course materials longitudinally, not just during/immediately after the experiment.
- **Learning Transfer:** measuring a learner's ability to apply new knowledge in novel contexts beyond the classroom/study.
- **Ontrackness:** the extent to which a learner adheres to the designed learning path as intended by the instructor.
- **Engagement:** a number of studies measure forms of learner activity/behavior and fall under this category. Specific forms of engagement include:
 - **Forum Participation:** measured by the frequency with which learners post to the course discussion forum (including posts and responses to others' posts).
 - **Video Engagement:** the amount of actions (pause, play, seek, speed change, toggle subtitles) a learner takes on a video component.

² www.edx.org.

³ www.coursera.org.

- **Revision:** the act of changing a previously-submitted response.
- **Persistence/Coverage:** the amount of the total course content accessed. For example, a learner accessing 75 out of the 100 components of a course has 75% persistence.
- **Self-Efficacy:** a learner's self-perceived ability to accomplish a given task.
- **Efficiency:** the rate at which a learner progresses through the course. This is most commonly operationalized by the amount of material learned relative to the total time spent.

4. Review

In the following review we synthesize the findings and highlight particularly interesting aspects of certain experiments. Unless otherwise indicated, all results presented below come from intention-to-treat (ITT) analyses, meaning all participants enrolled in each experimental condition are considered without exception. Each category has a corresponding a table detailing the total sample size (“N”), experimental environment (“Env.”), incentive for participation (“**Incentive**”), and reported results (“**Result**”). In the **Result** column, statistically significant positive outcome variables as a result of the experimental treatment are indicated with a +; null findings where no significant differences were observed are indicated with a 0; and negative findings where the treatment resulted in an adverse effect on the outcome variable are indicated with a -.

4.1. Mastery learning

Teaching for mastery learning places an emphasis on learners gaining a full understanding of one topic before advancing to the next (Bloom, 1968). Given that students' understanding of new topics often relies upon a solid understanding of prerequisite topics, teaching for mastery learning only exposes students to new material once they have mastered all the preceding material, very much in line with constructivist thinking as outlined by Cunningham and Duffy (1996, pp. 170–198). In the traditional classroom, teaching for mastery learning presents a major challenge for teachers in that they must constantly monitor each individual student's progress towards mastery over a given topic—a nearly impossible task in a typical classroom with 30 students, never mind 30,000. However, with the growing capabilities of education technologies, individualized mastery learning pedagogy can now be offered to students at scale.

While mastery learning is so frequently found to be an effective teaching strategy in terms of student achievement, it often comes at the cost of time. This issue of time could be a reason behind there being only one paper in this category. Mostafavi, Eagle, and Barnes (2015) implemented a data-driven knowledge tracing system to measure student knowledge and release content according to their calculated knowledge state. Students using this system were far more engaged than those using the default problem set or that with on-demand hints. A strict implementation of mastery learning — as in Mostafavi et al. (2015), where learners in an ITS are required to demonstrate concept mastery before advancing in the system — would be useful to understand its effect on the heterogeneous MOOC learner population (cf Table 3).

Table 3
Mastery learning.

| Mastery Learning: + : 1 | | | | |
|-------------------------|-----|------|-----------|--------------|
| Ref. | N | Env. | Incentive | Result |
| Mostafavi et al. (2015) | 302 | ITS | Class | + Engagement |

+ = positive effects.

4.2. Metacognitive strategies

Metacognitive behavior is defined by Hattie (2008) as “higher-order thinking which involves active control over the cognitive processes engaged in learning.” Metacognition is an invaluable skill in MOOCs, where learners cannot depend on the watchful eye of a teacher to monitor their progress at all times. Instead, they must be highly self-directed and regulate their own time management and learning strategies to succeed. The papers in this category explore novel course designs and interventions that are intended to make learners more self-aware, reflective, and deliberate in the planning of (and adherence to) their learning goals.

Davis, Chen, van der Zee, Hauff, and Houben (2016) conducted two experiments: in study “A” they provided learners with retrieval cue prompts after each week's lecture, and in study “B” they provided study planning support to prompt learners to plan and reflect on their learning habits. Overall, neither intervention had any effect on the learners in the experimental conditions, likely because the learners could ignore the prompts without penalty. However, when narrowing down to the very small sample of learners who engaged with the study planning module, the authors found desirable significant increases in learner behavior. Maass and Pavlik (2016) also ran an experiment testing support for retrieval practice. They found that (i) retrieval prompts increase learning gain and (ii) the complexity of the retrieval prompt had a significant impact on the prompts effect, with deeper prompts leading to better learning gains. In contrast, the retrieval prompts used by Davis et al. (2016) assessed shallow, surface-level knowledge, which could be a reason for the lack of a significant effect.

Even though the education psychology literature suggests that boosting learners' metacognitive strategies is highly effective for

increasing learning outcomes (Hattie, 2008), 23 of the 38 results (61%) in this category report null or negative findings. Furthermore, with the reporting of a negative impact of an intervention, Kizilcec et al. (2014a) found a certain form of participation encouragement (collectivist-framed prompting) to actually decrease learners' participation in the course discussion forum.

Nam and Zellner (2011) conducted a study evaluating the effect of framing a group learning activity in different ways. Compared to a “group processing” frame of mind (where group members are asked to assess the contribution of each group member), the “positive interdependence” frame of mind (where group members are reminded that boosting one's individual performance can have a great impact on the overall group achievement) group had higher post assessment scores.

In lieu of an actual learning platform, crowdworker platforms are also beginning to be used for learning research. One example is the study by Gadiraju and Dietze (2017), who evaluated the effect of achievement priming in information retrieval microtasks. While completing a crowdworker task aimed at teaching effective information retrieval techniques, the participants were also assessed on their learning through a test at the end of the task. By providing achievement primers (in the form of inspirational quotes) to these crowdworkers, the authors observed no significant difference in persistence or assessed learning. Given the ease with which these experiments can be deployed, more work should go into exploring the reproducibility of findings from a crowdworker context to an actual learning environment.

In summary, the current body of work in supporting learners' metacognitive awareness indicates how difficult it is to affect such a complex cognitive process, as more than half of the reported results from this category led to non-significant results (cf Table 4). While some studies do indeed report positive results, the overall trend in this category is an indication that we have not yet mastered the design and implementation of successful metacognitive support interventions that can effectively operate at scale. Setting this apart from other categories is the difficulty to measure metacognition; compared to other approaches such as questioning (where both the prompt and response are easily measurable), both eliciting and measuring responses to metacognitive prompts is far more challenging.

Table 4

Metacognitive strategies.

| Ref. | N | Env. | Incentive | Result |
|--|--------|--------|-----------|---|
| Metacognitive Strategies: +: 15 / °: 20 / -: 3 | | | | |
| Kizilcec, Prez-Sanagustn, and Maldonado (2016) | 653 | MOOC | None | °Persistence °Final Grade |
| Lang, Heffernan, Ostrow, and Wang (2015) | 950 | ITS | None | °Learning Gain °Engagement |
| Lamb, Smilack, Ho, and Reich (2015) | 4777 | MOOC | None | + ForumParticipation |
| Sonnenberg and Bannert (2015) | 70 | Native | \$ | + Learning Gain |
| Dodge, Whitmer, and Frazee (2015) | 882 | LMS | Class | °Final Grade |
| Tabuenca, Kalz, Ternier, and Specht (2015) | 60 | Native | \$ | °Exam Score |
| Kizilcec et al. (2014a) | 11,429 | MOOC | None | °ForumParticipation-ForumParticipation |
| Margulieux and Catrambone (2014) | 120 | Native | Credit | + Exam Score |
| Xiong, Wang, and Beck (2015) | 2052 | Native | None | + Learning Gain + Completion Rate |
| Noroozi, Weinberger, Biemans, Teasley, and Mulder (2012) | 56 | Native | n/r | + Learning Gain |
| Davis et al. (2016)A | 9836 | MOOC | None | °Final Grade °Engagement °Persistence |
| Davis et al. (2016)B | 1963 | MOOC | None | °Final Grade °Engagement °Persistence |
| Maass and Pavlik (2016) | 178 | Mturk | \$ | + Learning Gain |
| Kizilcec, Davis, and Cohen (2017a) | 1973 | MOOC | None | + Final Grade + Persistence + Completion Rate |
| Yeomans and Reich (2017)A | 293 | MOOC | None | -Completion Rate |
| Yeomans and Reich (2017)B | 3520 | MOOC | None | -Completion Rate °Engagement |
| Gadiraju and Dietze (2017) | 340 | Mturk* | \$ | °Final Grade °Persistence |
| Kim, Reinecke, and Hullman (2017) | 378 | Mturk | \$ | + Final Grade |
| Hwang and Mamykina (2017) | 225 | Native | n/r | + Learning Gain |
| De Grez, Valcke, and Roozen (2009) | 73 | Native | Class | °Learning Gain |
| Nam and Zellner (2011) | 144 | Native | Class | °Engagement + Final Grade |
| Huang, Wu, and Chen (2012) | 60 | Mobile | None | + Final Grade |
| Poos, van den Bosch, and Janssen (2017) | 80 | Lab | None | °Final Grade °Learning Transfer |
| Gamage et al. (2017) | 87 | MOOC | n/r | + Learning Gain |

+ = positive effects, ° = null results, - = negative effects.

4.3. Questioning

Hattie (2008) found questioning to be one of the most effective teaching strategies in his meta-analysis. Questioning (cf Table 5) is characterized by the posing of thoughtful questions that elicit critical thought, introspection, and new ways of thinking. The studies in this category explore new methods of prompting learners to retrieve and activate their prior knowledge in formative assessment contexts. Yang, Shih, and Chen (2016) evaluated the effectiveness of a two-tier questioning technique, described as “... a set of two-level multiple choice questions [in which the] first tier assesses students' descriptive or factual knowledge ... while the second tier investigates the reasons for their choices made in the first tier.” They found this questioning technique to be highly effective in their experiment, with learners in the two-tier condition achieving 0.5 standard deviations better learning gains than learners receiving standard one-tier questions.

Table 5
Questioning.

| Questioning: +: 7 / °: 7 | | | | |
|--|------|--------|-----------|---|
| Ref. | N | Env. | Incentive | Result |
| Yang et al. (2016) | 43 | Native | n/r | + Learning Gain |
| Thompson, Braude, Canfield, Halfond, and Sengupta (2015) | 43 | Native | Class | + Learning Gain |
| Williams et al. (2016a) | 659 | Mturk | \$ | + Revision |
| Şendağ and Ferhan Odabaşı (2009) | 40 | Native | Class | °Learning Gain °Final Grade |
| Chen (2010) | 84 | Native | Class | + Learning Gain |
| de Koning et al. (2010) | 76 | Native | Credit | + Final Grade °Persistence °Learning Transfer |
| Yang, Chen, and Hwang (2015) | 79 | Native | Class | + Final Grade °Engagement |
| Attali (2015) | 804 | Mturk | \$ | + Learning Gain |
| Attali and van der Kleij (2017) | 2445 | Native | None | °Persistence °Final Grade |

+ = positive effects, ° = null results.

Instructional questioning was explored in the Mturk setting by Williams, Lombrozo, Hsu, Huber, and Kim (2016a) who compared the effectiveness of different questioning prompt wordings. They found prompts that directly ask the learner to provide an explanation of why an answer is correct leads learners to revise their answers (to the correct one) more than a prompt asking for a general explanation of the answer.

de Koning, Tabbers, Rikers, and Paas (2010) conducted a study where half of the learners were cued to generate their own inferences through self-explaining and half were provided pre-written instructional explanations. Taking place in the context of a course about the human cardiovascular system, results show that learners prompted to self-explain performed better on the final test, but did not show any difference in persistence or learning transfer from the given explanation group.

Given its effectiveness and relative simplicity to implement, two-tier questioning should be further investigated in the MOOC setting to stimulate learners critical thought beyond surface-level factual knowledge.

Related to the tactic of questioning is the learning strategy known as retrieval practice, or the testing effect, which is characterized by the process of reinforcing prior knowledge by actively and repeatedly recalling relevant information Adesope, Trevisan, and Sundararajan (2017). Recent work has found retrieval practice to be highly effective in promoting long-term knowledge retention Adesope et al. (2017); Clark and Mayer (2016); Roediger and Butler (2011); Henry, Roediger, and Karpicke (2016); Lindsey, Shroyer, Pashler, and Mozer (2014); Karpicke and Roediger (2008); Karpicke and Blunt (2011). Accordingly, we recommend that future research interested in questioning tactics is designed to stimulate learners to engage in retrieval practice.

4.4. Spaced vs. massed practice

Hattie (2008) describes the difference between spaced learning (sometimes referred to as distributed practice) and massed practice as “the frequency of different opportunities rather than merely spending more time on task.” In other words, distributing one's study sessions over a long period of time (e.g., 20 min per day for 2 weeks) is characteristic of high spacing, whereas studying in intense, concentrated sessions (one 4-h session) is characteristic of massed practice (Wulf & Shea, 2002). Historically, studies have found that the desired effect of spaced learning (long-term knowledge retention) is found most commonly in tasks of low difficulty, and the effect decreases as the difficulty increases (Rohrer & Taylor, 2006).

Dearman and Truong (2012) developed a mobile phone “Vocabulary Wallpaper” which aimed to implicitly teach (through the learners mobile phone background) learners new vocabulary in a second language in highly spaced microlearning sessions. Their findings show that, compared to learners receiving the lessons at less distributed rates, learners with highly-spaced exposure showed a significant increase of second language vocabulary learned.

As evidenced by the lone study in the category (cf Table 6), it is difficult to design and implement experiments that effectively get learners to commit to high spacing (ideally enacted as a learned self-regulation skill). Even still, given its proven effectiveness elsewhere in the learning literature (Hattie, 2008), practitioners and researchers should tackle this design challenge in creating and evaluating environments that encourage spaced practice.

Table 6
Spaced vs. Massed Practice.

| Spaced vs. Massed Practice: +: 1 | | | | |
|----------------------------------|----|--------|-----------|-----------------|
| Ref. | N | Env. | Incentive | Result |
| Dearman and Truong (2012) | 15 | Mobile | \$ | + Learning Gain |

+ = positive effects.

4.5. Matching learning styles

Brown, Brailsford, Fisher, and Moore (2009) conducted an experiment testing the efficacy of “learning style-adapted e-learning environments.” In the study, where students self-proclaimed learning styles were either matched or unmatched, yielded no significant differences in terms of learner achievement between conditions. Consistent with the current popular literature on the topic (Kirschner, 2017; Kirschner & van Merriënboer, 2013), the authors found that adapting the courses to students’ learning styles did not result in any significant benefit.

Soflano, Connolly, and Hailey (2015) employed a game-based learning environment to evaluate the impact of adapting instruction to learning styles in a computer programming learning context. The authors report that compared to the groups using a non-adaptive version of the SQL language tutor software, the adaptive system yielded no difference in final grades (Soflano et al., 2015).

However, there does still exist some evidence in favor of this learning strategy (cf Table 7). Chen, Kinshuk, Wei, and Liu (2011) created an online learning environment where the teaching strategy was adapted to each of the learners’ individual thinking styles. With three teaching strategies (constructive, guiding, or inductive) either matched or unmatched to three thinking styles (legislative, executive, or judicial, respectively), the authors found that the group who had their thinking style matched accordingly outperformed those who did not.

Table 7

Matching learning styles.

| Matching Learning Styles: +: 2 / ◦ : 2 | | | | |
|--|-----|--------|-----------|---------------------------|
| Ref. | N | Env. | Incentive | Result |
| Brown et al. (2009) | 221 | Native | Class | ◦Exam Score |
| Chen et al. (2011) | 223 | Native | n/r | + Final Grade |
| Soflano et al. (2015) | 120 | Native | None | ◦Final Grade + Efficiency |

+ = positive effects, ◦ = null results.

Instead of adapting to a single modality that a learner prefers (such as being a “visual learner”), the literature on learning styles emphasizes that while one modality may be preferred by the learner (and can lead to positive experimental results in certain contexts), providing them instruction in a variety of modalities will provide the greatest benefit overall (Kirschner & van Merriënboer, 2013).

4.6. Feedback

Hattie (2008) defines feedback as “information provided by an agent (e.g., teacher, peer, book, parent, or one’s own experience) about aspects of one’s performance or understanding.” Strategically providing students with feedback offers them the chance to reflect and reassess their approach to a given situation. Feedback can best be thought of as a mirror for learners; it serves to encourage them to stop and mindfully evaluate their own behavior or learning processes — which are otherwise unconscious or unconsidered — and make them readily visible. However, this act of mindfully evaluating and altering one’s behavior should not be taken for granted. Self-regulating one’s own learning processes (especially in response to feedback) is a skill which is highly correlated with and caused by prior education (Winters, Greene, & Costich, 2008). Especially in the MOOC context, where the learners come from many diverse backgrounds, it is imperative that feedback offered to the learner is adaptive and aligned to their ability to process, understand, and act upon it.

While Hattie (2008) finds feedback to be the most effective teaching strategy in his entire meta-analysis, we find very mixed results in our selected studies in terms of its effectiveness. Of the 38 results reported within the 21 papers of this category, only 14 (37%) are positive findings.

Zooming in on two of the MOOC studies in this category, Coetzee, Fox, Hearst, and Hartmann (2014) and Tomkin and Charlevoix (2014) evaluated the effectiveness of feedback in the context of the discussion forum. Coetzee et al. (2014) tested the effectiveness of implementing a reputation system in a MOOC discussion forum — the more you post to the forum, the more points you accumulate (this paper also applies to the Simulations & Gaming category for this reason). The authors found that providing this positive feedback did indeed lead learners to post more frequently in the forum, but this did not have any impact on their final course grade. Tomkin and Charlevoix (2014) ran an experiment in which learners were divided into one of two course discussion forums — in one forum the instructor was active in providing individualized feedback to learners and engaging in discussion, and in the other no instructor feedback was provided. The authors report no differences in either completion rate or course engagement between the two conditions.

To address the challenge of providing in-depth feedback on students’ learning in a coding context, Wiese, Yen, Chen, Santos, and Fox (2017) tested the effectiveness of a code style tutor which offered adaptive, real-time feedback and hints to students learning to code. Compared to a control group receiving a simplified feedback system consisting of a single unified score assessing the code, students who used the adaptive feedback system did not show any difference in the extent to which they improved their coding style (Wiese et al., 2017, pp. 41–50).

Bian et al. (2016) developed and evaluated an animated pedagogical agent which was able to provide different types of emotional

feedback to participants in a simulated environment. They found that positive emotional feedback (expressing happiness and encouragement in response to desirable behavior) led to significantly higher test scores than negative feedback (where the agent expressed anger and impatience to undesirable behavior). Also taking place in a simulated environment, the experiment carried out by Sedrakyan, Snoeck, and Poelmans (2014) evaluated the effectiveness of a feedback-enabled simulation learning environment. Compared to students in the intervention group who interacted with the feedback-enabled simulation environment, those in the control condition, who did not have access to the simulation, performed more poorly on a final assessment.

While navigational feedback (support for learners in optimizing their learning path through a course) like that introduced by Borek, McLaren, Karabinos, and Yaron (2009) is common in ITS to help learners through problems, the challenge now arises to provide personalized feedback at scale on other factors such as learner behavior patterns (cf Table 8). This way, feedback can be used as a mechanism to make learners more aware of their learning habits/tendencies and, in turn, better at self-regulating. However, with only 37% of results reported in this category being positive, this highlights the fact that simply providing feedback is insufficient in promoting positive learning outcomes—these results are an indication that, even though we now have developed the technology to enable the delivery of feedback at scale, attention must now be shifted towards understanding the nuance of what type of feedback (and with what sort of frequency) will help the learner in a given context or state.

Table 8
Feedback.

Feedback: +: 14 / ° : 22 / -: 2

| Ref. | N | Env. | Incentive | Result |
|--|--------|--------|-----------|--|
| Kulkarni, Bernstein, and Klemmer (2015a) | 104 | Native | None | + Exam Score°Revision |
| Williams et al. (2016b) | 524 | MTurk | \$ | + Exam Score°Self-Efficacy |
| Eagle and Barnes (2013) | 203 | Native | Class | + Persistence |
| Kardan and Conati (2015) | 38 | Native | n/r | + Learning Gain°Exam Scores°Efficiency |
| Fossati et al. (2009) | 120 | Native | Class | + Learning Gain |
| Borek et al. (2009) | 87 | Native | Class | °Exam Score°Learning Transfer + Learning Gain |
| Coetzee et al. (2014) | 1101 | MOOC | None | °Final Grade°Persistence + ForumParticipation |
| Tomkin and Charlevoix (2014) | 20,474 | MOOC | None | °Completion Rate°Engagement |
| Beheshitha, Hatala, Gašević, and Joksimović (2016) | 169 | Class | None | °ForumParticipation |
| Rafferty, Jansen, and Griffiths (2016) | 200 | Mturk | \$ | °Learning Gain |
| Bian et al. (2016) | 56 | Lab | None | + Self-Efficacy + Exam Score |
| Nguyen, Garnarcz, Ng, Dabbish, and Dow (2017) | 205 | Mturk | \$ | + Final Grade - Revision |
| Wiese et al. (2017) | 103 | Native | Class | °Learning Gain °Revision |
| Davis et al. (2017) | 33,726 | MOOC | None | + Completion Rate °Engagement |
| Mitrovic, Ohlsson, and Barrow (2013) | 41 | Native | Class | °Learning Gain °Engagement °Final Grade + Efficiency |
| Corbalan, Paas, and Cuyppers (2010) | 34 | Native | n/r | °Engagement °Final Grade |
| González, Jover, Cobo, and Muñoz (2010) | 121 | Native | Class | + Final Grade |
| van der Kleij, Eggen, Timmers, and Veldkamp (2012) | 152 | Native | Class | °Final Grade |
| Erhel and Jamet (2013) | 41 | Lab | n/r | °Final Grade |
| Christy and Fox (2014) | 80 | Lab | Credit | -Final Grade |
| Sedrakyan et al. (2014) | 66 | Native | n/r | + Final Grade |

4.7. Cooperative learning

Interventions targeting cooperative learning explore methods to enable learners in helping and supporting each other in the understanding of the learning material. Cooperative learning (cf Table 9) is one of the major opportunity spaces in MOOCs for their unprecedented scale and learner diversity, as evidenced by the prevalence of reported positive findings (71%). The studies in this category develop and test solutions which try to find new ways to bring learners together no matter where they are in the world to complete a common goal.

Table 9
Cooperative learning.

Cooperative Learning: +: 17 / ° : 6 / -: 1

| Ref. | N | Env. | Incentive | Result |
|--|------|--------|-----------|------------------------|
| Bhatnagar et al. (2015) | 144 | Native | Class | + Revision |
| Lan, Graf, Lai, and Kinshuk (2011) | 54 | Native | Class | + Learning Gain |
| Tritz, Michelotti, Shultz, McKay, and Mohapatra (2014) | 396 | Native | Class | + Final grade |
| Ngoon, Gamero-Garrido, and Klemmer (2016) | 75 | Native | \$ | + Learning Gain |
| Kulkarni et al. (2015a) | 104 | Native | None | + Exam Score °Revision |
| Kulkarni et al. (2015b) | 2422 | MOOC | None | + Exam Score |
| Cambre, Kulkarni, Bernstein, and Klemmer (2014) | 2474 | MOOC | None | + Exam Score |
| Culbertson, Andersen, White, Zhang, and Jung (2016) | 42 | Native | \$/Credit | + Learning Gain |
| Coetzee, Lim, Fox, Hartmann, and Hearst (2015) | 1334 | MTurk | \$ | + Exam Scores |

(continued on next page)

Table 9 (continued)

| Cooperative Learning: + : 17 / ° : 6 / - : 1 | | | | |
|--|------|--------|-----------|--|
| Ref. | N | Env. | Incentive | Result |
| Zheng et al. (2014) | 1730 | MOOC | None | °Engagement °Persistence °Final Grade |
| Khandaker and Soh (2010) | 145 | Native | Class | °Exam Score |
| Konert, Bellhäuser, Röpke, Gallwas, and Zucik (2016) | 396 | Class | None | + Engagement + Persistence |
| Labarthe et al. (2016) | 8673 | MOOC | None | + Persistence + Completion Rate + Engagement |
| AbuSeileek (2012) | 216 | Lab | n/r | + Final Grade |
| Papadopoulos et al. (2012) | 54 | Native | Class | + Final Grade |
| Chang, Lin, and Tsai (2013) | 27 | LMS | Class | + Final Grade |
| Cho and Lee (2013) | 120 | Class | None | -Engagement °Learning Gain |

+ = positive effects, ° = null results, - = negative effects.

One successful example of this is the study by Kulkarni, Cambre, Kotturi, Bernstein, and Klemmer (2015b) where MOOC learners were divided into small groups (between 2 and 9 learners per group) and allowed to have discussions using real-time video calls over the internet. Each group was given prompts encouraging the learners to both discuss course materials and share general reflections of the course experience. The authors found that learners in groups with a larger diversity of nationalities performed significantly better on the course final exam than learners in groups with low diversity. This result shows promise that the scale and diversity of MOOC learners can actually bring something novel to the table in learners' apparent interest in cultural diversity.

On a similar note, Zheng, Vogelsang, and Pinkwart (2014) developed an algorithm which aimed to divide MOOC learners into small groups in a more effective fashion compared to randomization. This model took into consideration the following factors: collaboration preferences (local, email, Facebook, Google + or Skype), gender, time zone, personality type, learning goal, and language. The authors found this algorithmic sorting of students into groups to not have any effect on overall engagement, persistence, or final grade. Whereas this algorithm grouped largely for *similarity* (for example, grouping learners in the same time zone together), the study presented by Kulkarni et al. (2015b) suggests that *diversity* may be a better approach to automated group formation.

There are also possibilities for cooperative learning in which the learners do not meet face-to-face. In this light, Bhatnagar et al. (2015) evaluated a cooperative learning system which crowd-sourced learner explanations. After answering an assessment question, learners were prompted to give an explanation/justification. These explanations were then accumulated and shared with their peers; the authors found that providing learners the explanations of their peers increased the likelihood of a learner revising their answer to the correct one. In this scenario, the prompting for explanations not only serves as a reflective activity for the individual learner, it also leverages the social aspect by allowing him or her to contribute to the larger course community and potentially help a peer in need.

Cho and Lee (2013) investigated the effectiveness of co-explanation (where learners are instructed to collaboratively explain worked examples) as compared to self-explanation (where learners work alone) in a Design Principles course context. The authors found that learners in the co-explanation condition were not as engaged with the assessment questions (identifying a design's strengths and weaknesses) as their counterparts in the self-explanation condition.

In the domain of peer review, Papadopoulos, Lagkas, and Demetriadis (2012) compared "free-selection" peer review (where students could freely choose which of their peers' work to review) against an "assigned-pair" design (where the peer review pairings are assigned by the instructor) in the context of a computer networking course. The authors found that students in the free-selection group achieved greater learning outcomes and provided better reviews than those in the assigned-pair group (Papadopoulos et al., 2012).

Labarthe, Bouchet, Bachelet, and Yacef (2016) developed a recommender system within a MOOC to provide each learner with a list of peers in the same course who they would likely work well with based on profile similarity modeling. Compared to the control group with no recommendations, the experimental group (receiving the list of peer recommendations) displayed significantly improved persistence, completion rate and engagement.

Given the consistency of positive results in this category (71%—the highest of any category), the above studies should be used as building blocks or inspiration for future work in finding new ways to bring learners together and increase their sense of community and belonging in the digital learning environment. Advances made in this vein would work towards harnessing the true power of large-scale open learning environments where learners not only learn from the instructor but from each other as well through meaningful interactions throughout the learning experience Siemens (2005).

4.8. Simulations & Gaming

Hattie (2008) categorizes simulations and games together and defines them as a simplified model of social or physical reality in which learners compete against either each other or themselves to attain certain outcomes. He also notes the subtle difference between simulations and gaming in that simulations are not always competitive. The studies in this category (cf Table 10) are carried out predominantly in native environments. While understandable given the games could have been developed for purposes other than experimentation, this raises potential issues with an eye towards reproducibility. However, considering 19 of the 28 reported results (68%) in the category pertain to desirable benefits in learner achievement or behavior, this also indicates a very strong trend towards the generalizable effectiveness of using simulations and gamification to help learners.

Table 10
Simulations & gaming.

| Simulations & Gaming: +: 19 / ◦ : 9 | | | | |
|---|------|--------|--------------|---|
| Ref. | N | Env. | Incentive | Result |
| Bumbacher, Salehi, Wierzchula, and Blikstein (2015) | 36 | Native | n/r | + Learning Gain |
| Cox, Cairns, Shah, and Carroll (2012) | 41 | Native | \$ | + Engagement |
| Culbertson et al. (2016) | 42 | Native | \$/Credit | + Engagement + Exam Score |
| Ibanez, Di-Serio, and Delgado-Kloos (2014) | 22 | Native | Class | + Exam Score |
| Krause, Mogalle, Pohl, and Williams (2015) | 206 | Native | n/r | + Persistence + Exam Score |
| Li, Grossman, and Fitzmaurice (2014) | 24 | Native | n/r | ◦Efficiency + Learning Transfer |
| Schneider, Jermann, Zufferey, and Dillenbourg (2011) | 82 | Native | Class | + Learning Gain + Engagement |
| Cutumisu and Schwartz (2016) | 264 | Mturk | \$ | ◦Engagement◦Learning Gain |
| Coetzee et al. (2014) | 1101 | MOOC | None | ◦Final Grade◦Persistence + ForumParticipation |
| Cheng, Yang, and Andersen (2017) | 68 | Native | n/r | ◦Learning Gain |
| Pozo-Barajas, Pablo-Romero, and Caballero (2013) | 194 | Native | n/r | + Final Grade |
| Smith et al. (2013) | 57 | Native | Class | + Learning Gain |
| Brom, Bromová, Děchtěrenko, Buchtová, and Pergel (2014) | 75 | Lab | \$ or Credit | ◦Final Grade + Engagement |
| Attali and Arieli-Attali (2015) | 1218 | Mturk | \$ | ◦Final Grade ◦Efficiency |
| Cagiltay et al. (2015) | 142 | Native | n/r | + Final Grade |
| Hooshyar et al. (2016) | 52 | Native | Class | + Learning Gain |
| Nebel, Schneider, Beege, and Rey (2017) | 103 | Native | \$ or Credit | + Final Grade + Efficiency |
| Barr (2017) | 72 | Lab | n/r | + Final Grade |

+ = positive effects, ◦ = null results.

While each game or simulation is unique in its own right, the underpinning theme in all of these studies is as follows: the learner earns and accumulates rewards by exhibiting desirable behavior as defined by the instructor/designer. While creating native educational games or gamifying existing learning environments (especially MOOCs as in Coetzee et al. (2014)) is a complex, time-consuming process, based on the predominantly positive findings in the literature, we conclude that it is an area with high potential for boosting learning performance.

Cutumisu and Schwartz (2016) ran a study evaluating the effect of choosing versus receiving feedback in a game-based assignment. Compared to the group which passively received feedback, the group which was forced to actively retrieve feedback was more engaged with the environment, but showed no difference in learning or revision behavior. Note that this study is not in the Feedback category, as both cohorts received the same feedback; the element being tested in the experiment was the manner in which it was delivered within the simulated environment.

Cagiltay, Ozcelik, and Ozcelik (2015) employed a serious game design and put students either in a competitive (showing a scoreboard and ranking of peer performance) or non-competitive environment. The authors found that the competitive environment led to significantly higher test scores and more time spent answering questions. On a similar note, Attali and Arieli-Attali (2015) evaluated the effect of implementing a points system within a computer-based mathematics learning environment. Although participants in the conditions with the points system answered questions faster, there was no effect on the accuracy of their responses.

Hooshyar et al. (2016) created a formative assessment game for a computer programming learning task. The authors found that participants in a traditional, non-computer-based environment performed worse on problem solving tasks than those who received the computer-based formative assessment system. Barr (2017) also compared a game-based learning environment to a non-computer-based experience. In their experiment, the participants in the game-based learning condition displayed better scores on a post-test.

With 68% of the reported results being positive findings—the second highest among all categories—we see great potential for the effectiveness of learning experiences where learners are afforded the ability to interact with and explore simulated environments. Due to the substantial cost of developing such environments, future research is needed to evaluate whether this trend of positive findings continues so that institutions can be assured in justifying their investment in these instructional strategies.

4.9. Programmed instruction

According to Hattie (2008), programmed instruction is a method of presenting new subject matter to students in a graded sequence of controlled steps. Its main purposes are to (i) manage learning under controlled conditions and (ii) promote learning at the pace of the individual learner.

Programmed instruction (cf Table 11) is inherently adaptive—it presents material to the learner according to that learner's unique set of previous actions. As they stand now, MOOCs are simply online course content resources that remain static irrespective of a learner's behavior. Unlike the native and lab environments used in Brinton et al. (2015) and Karakostas and Demetriadis (2011), the current MOOC technology has not yet accounted for a learner's past behavior in delivering personalized content accordingly. By developing and implementing these types of systems in a MOOC, MOOCs could then become more adaptable and able to cater instruction based on the individual learner. To enable this would require a real-time tracking system for learners where their behavior could be collected, modeled/analyzed, and then acted upon (e.g., with the delivery of a personalized recommendation for a next activity or resource) in real time on a large scale.

Table 11
Programmed instruction.

| Programmed Instruction: +: 4 / ◦ : 7 | | | | |
|--|-----|--------|-----------|---|
| Ref. | N | Env. | Incentive | Result |
| Brinton et al. (2015) | 43 | Native | None | + Engagement |
| Karakostas and Demetriadis (2011) | 76 | Lab | n/r | + Learning Gain |
| Rosen et al. (2017) | 562 | MOOC | None | + Learning Gain + Persistence ◦Final Grade |
| Zhou, Wang, Lynch, and Chi (2017) | 153 | ITS | Class | ◦Final Grade ◦Engagement |
| Arawjo, Wang, Myers, Andersen, and Guimbretière (2017) | 24 | Native | n/r | ◦Engagement ◦Completion Rate ◦Learning Transfer |
| van Gog (2011) | 32 | Lab | n/r | ◦Final Grade |

+ = positive outcomes ◦ = null results.

4.10. Interactive multimedia methods

As lecture videos are currently the backbone of MOOC instructional content, it is imperative that they effectively impart knowledge to learners in an engaging, understandable fashion. Also among the most effective strategies with 64% of reported results being positive, interactive multimedia methods, though not limited to video, test various methods of content delivery through multimedia application interfaces (cf Table 12).

Table 12
Interactive multimedia methods.

| Interactive Multimedia Methods: +: 28 / ◦ : 12 / -: 4 | | | | |
|---|-----|--------|--------------|---|
| Ref. | N | Env. | Incentive | Result |
| Lee et al. (2015) | 102 | Native | n/r | + Learning Gain |
| Monserrat, Zhao, McGee, and Pandey (2013) | 15 | Native | None | + Efficiency |
| Monserrat et al. (2014) | 18 | Native | None | + Learning Gain |
| Nicholson, Huber, Jackson, and Olivier (2014) | 40 | Native | None | + Efficiency |
| Trusty and Truong (2011) | 21 | Native | \$ | ◦Learning Gain◦Engagement |
| Dearman and Truong (2012) | 15 | Mobile | \$ | + Learning Gain |
| Kizilcec et al. (2014b) | 22 | Lab | Credit | ◦Exam Score |
| Kwon and Lee (2016) | 120 | Mturk | \$ | + Exam Score + Efficiency |
| Kizilcec, Pérez-Sanagustín, and Maldonado (2017b) | 104 | Native | Credit | ◦Final Grade |
| Zhu et al. (2017) | 22 | Native | n/r | + Learning Gain + Engagement |
| Pandey et al. (2017) | 44 | Native | n/r | ◦Engagement - Final Grade |
| Culbertson, Shen, Jung, and Andersen (2017) | 27 | Native | \$ | ◦Learning Gain + Engagement |
| Austin (2009) | 75 | Native | Credit | + Learning Transfer |
| Yamada (2009) | 40 | Native | n/r | + Engagement ◦Revision |
| Wang, Vaughn, and Liu (2011) | 123 | Lab | n/r | + Final Grade |
| Pastore (2012) | 154 | LMS | n/r | + Efficiency |
| Chen, Wang, Chen, and Chen (2014) | 81 | LMS | n/r | + Final Grade |
| Chuang and Tsao (2013) | 111 | Mobile | n/r | + Learning Gain ◦Long Term Retention |
| AbuSeileek and Qatawneh (2013) | 30 | Native | Class | -Engagement |
| Imhof, Scheiter, Edelmann, and Gerjets (2013) | 71 | Native | \$ or Credit | -Final Grade |
| Urquiza-Fuentes and Velázquez-Iturbide (2013) | 132 | Native | Credit | ◦Final Grade ◦Long Term Retention + Completion Rate |
| Aldera and Mohsen (2013) | 50 | Native | Class | + Final Grade + Long Term Retention |
| Martin and Ertzberger (2013) | 109 | Mobile | Class | -Final Grade |
| Chen and Wu (2015) | 37 | Native | n/r | + Learning Gain |
| Song et al. (2014) | 144 | Native | None | + Learning Transfer + Engagement |
| van Gog, Verveer, and Verveer (2014) | 43 | Lab | \$ or Credit | + Final Grade |
| Limperos et al. (2015) | 259 | Lab | None | + Final Grade |
| Türkay (2016) | 621 | Mturk | \$ | ◦Persistence + Final Grade |
| Jang, Vitale, Jyung, and Black (2017) | 76 | Native | n/r | + Final Grade |
| Jeno, Grytnes, and Vandvik (2017) | 71 | Mobile | \$ | + Final Grade |
| van Wermeskerken and van Gog (2017) | 69 | Lab | \$ or Credit | ◦Final Grade |
| Sharma, Jermann, and Dillenbourg (2015) | n/r | MOOC | None | + VideoEngagement |

+ = positive effects, ◦ = null results, - = negative effects.

Kizilcec, Papadopoulos, and Sritanyaratana (2014b), for example, compared lecture videos which included a small overlay of the instructor's face talking versus the same lecture videos without the overlay. Results show that while learners preferred videos showing face and perceived it as more educational, there were no significant differences in the groups' exam scores.

A more interactive approach to lecture videos was explored by Monserrat, Li, Zhao, and Cao (2014) who integrated several interactive components into lecture videos by integrating quiz, annotation, and discussion activities within a video player. Compared

to a baseline interface, which separates the videos and assessments, the integrated interface was favored by learners and enabled them to learn more content in a shorter period of time.

Looking beyond video delivery methods, [Kwon and Lee \(2016\)](#) compare four delivery methods of a tutorial on the topic of data visualization. The four conditions are: (i) a baseline which only included text, (ii) baseline plus static images, (iii) video tutorial, and (iv) interactive tutorial where learners worked with a Web interface to manipulate and create their own data visualizations. The authors found that learners with the interactive tutorial performed better on the exam and did so while spending less overall time in the platform — an indication of increased efficiency.

[Aldera and Mohsen \(2013\)](#) evaluated the effectiveness of captioned animation with keyword annotation (a note explaining the meaning of a given word) in multimedia listening activities for language learning. Compared to participants who received either just animations or animations with captions, the captioned animations with keyword annotation condition performed significantly better on recognition and vocabulary tests. However, the participants just receiving the animations significantly outperformed the other conditions on listening comprehension and recall over time.

[Martin and Ertzberger \(2013\)](#) deployed a “here and now” learning strategy (where learners have 24/7 access to learning activities on their mobile phones) to compare its effectiveness against computer-based instruction. While the “here and now” conditions expressed more positive attitudes towards the learning experience after the experiment, the computer-based learning cohort earned higher scores on a post-test.

[Limperos, Buckner, Kaufmann, and Frisby \(2015\)](#) tested the impact of modality (text vs. audio + text) on learning outcomes. They found that the multimodal format (audio + text) led to better learning outcomes than receiving text alone. [Pastore \(2012\)](#) ran a study to see the effect of compressing the time of instruction (decreasing time to train/learn the materials) on learning. They found that decreasing (accelerating) the time by 25% leads to similar learning outcomes, whereas decreasing by 50% causes a decrease in learning.

Pursuing new research in this category is important going forward in trying to truly leverage the Web for all of its learning affordances. The possibilities for digital interfaces, sensors, and devices are expanding rapidly, and more immersive, interactive, and intelligent environments promise to make a significant impact on online learning environments in the future. Even before these exciting technologies have become widely explored, we still observe an encouraging trend in this category in terms of positive results reported; we therefore recommend future research continue to explore the new possibilities in highly dynamic, interactive learning environments.

5. Conclusion

Based on both the quantitative and qualitative analyses from this review, we identify Cooperative Learning, Simulations & Gaming, and Interactive Multimedia as the three most promising strategies for most effectively activating learning at scale. We draw this conclusion from the proportion of positive results from each category: 71% for Cooperative Learning, 68% for Simulations & Gaming, and 64% for Interactive Multimedia—compared to all other groups with more than one study which have an average of 43% positive results.

According to Hattie's meta-analysis including over 50,000 studies, the ten learning strategies shown in [Table 2](#) are among the most effective. And yet, in so many instances do we here find null results for the studies employing them. Based on Hattie's work and sheer volume of studies included, we cannot yet dismiss the strategies themselves as ineffective; rather, translating them to the digital age of scalable learning has emerged as the primary challenge for the future. We identify a key factor in meeting this challenge to be that of incentive. Compared to an even balance of positive and null findings in experiments without any incentive for the participant (49% + and 49% ◦), positive results are twice as likely as null results in experiments which provide a monetary incentive. Given that the main application area for the present review is concerned with self-directed, informal online learning environments such as MOOCs, we see this discrepancy as one that calls for thorough investigation in future research. This trend suggests that the applicability and effectiveness of instructional and/or learning strategies could potentially be context-dependent—that the same intervention might be highly effective in a context where participants have a monetary incentive and ineffective in a context where participants are rewarded with course credit or are intrinsically motivated, for example. This bears significant implications for the generalizability of online learning research in that (i) researchers must take great care in contextualizing findings and (ii) readers must be attentive in interpreting results and mindful of the study's full context.

We are in the beginning stages now of constructing this new narrative of pedagogy at scale, and would be naive to think this could have been perfected in just three years. Guided by the proven efficacy of Cooperative Learning, Simulations & Gaming, and Interactive Multimedia learning strategies, the community should now work through iterative cycles of designing, testing, and evaluating new solutions in formalizing this emergent body of theory and literature.

Acknowledgements

Dan Davis' research is supported by the Leiden-Delft-Erasmus Centre for Education and Learning. Guanliang Chen's research is supported by the Extension School of the Delft University of Technology. This work is co-funded by the Erasmus + Programme of the European Union. Project: STELA 62167-EPP-1-2015-BE-EPPKA3-PI-FORWARD.

References

- van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72, 323–327.
- van Wermeskerken, M., & van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education*, 113, 98–107.
- AbuSeileek, A. F. (2012). The effect of computer-assisted cooperative learning methods and group size on the EFL learners' achievement in communication skills. *Computers & Education*, 58(1), 231–239.
- AbuSeileek, A. F., & Qatawneh, K. (2013). Effects of synchronous and asynchronous computer-mediated communication (CMC) oral conversations on English language learners' discourse functions. *Computers & Education*, 62, 181–190.
- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). *Rethinking the use of tests: A meta-analysis of practice testing*. Review of Educational Research.
- Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education*, 68, 60–75.
- Arawjo, I., Wang, C.-Y., Myers, A. C., Andersen, E., & Guimbretière, F. (2017). *Teaching programming with gamified semantics*. New York, New York, USA: ACM.
- Attali, Y. (2015). Effects of multiple-try feedback and question type during mathematics problem solving on performance in similar problems. *Computers & Education*, 86, 260–267.
- Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education*, 83, 57–63.
- Attali, Y., & van der Kleij, F. (2017). Effects of feedback elaboration and feedback timing during computer-based practice in mathematics problem solving. *Computers & Education*, 110, 154–169.
- Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers & Education*, 53(4), 1339–1354.
- Barr, M. (2017). Video games can develop graduate skills in higher education students: A randomised trial. *Computers & Education*, 113, 86–97.
- Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learning analytics visualizations. *LAK '16* (pp. 54–63).
- Bhatnagar, S., Desmarais, M., Whittaker, C., Lasry, N., Dugdale, M., & Charles, E. S. (2015). An analysis of peer-submitted and peer-reviewed answer rationales, in an asynchronous Peer Instruction based learning environment. *EDM '15* (pp. 456–459). http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_48.pdf.
- Bian, Y., Yang, C., Guan, D., Xiao, S., Gao, F., Shen, C., et al. (2016). Effects of pedagogical Agent's personality and emotional feedback strategy on Chinese students' learning experiences and performance. *CHI '16* (pp. 433–444). ACM.
- Bloom, B. S. (1968). Learning for mastery. *Instruction and Curriculum*, 1(2), 1–12.
- Borek, A., McLaren, B. M., Karabinos, M., & Yaron, D. (2009). How much assistance is helpful to students in discovery learning? *EC-TEL '09* (pp. 391–404). Springer. http://link.springer.com/chapter/10.1007/978-3-642-04636-0_38.
- Brinton, C. G., Rill, R., Ha, S., Chiang, M., Smith, R., & Ju, W. (2015). Individualization for education at scale: Miic design and preliminary evaluation. *IEEE TLT: Vol. 8*, (pp. 136–148). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6955856.
- Brom, C., Bromová, E., Děchtěrenko, F., Buchtová, M., & Pergel, M. (2014). Personalized messages in a brewery educational simulation: Is the personalization principle less robust than previously thought? *Computers & Education*, 72, 339–366.
- Brown, E. J., Brailsford, T. J., Fisher, T., & Moore, A. (2009). Evaluating learning style personalization in adaptive systems: Quantitative methods and approaches. *IEEE TLT: Vol. 2*, (pp. 10–22). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4785451.
- Bumbacher, E., Salehi, S., Wierzchula, M., & Blikstein, P. (2015). Learning environments and inquiry behaviors in science inquiry Learning: How their interplay affects the development of conceptual understanding in physics. *EDM '15* (pp. 61–68).
- Cagiltay, N. E., Ozcelik, E., & Ozcelik, N. S. (2015). The effect of competition on learning in games. *Computers & Education*, 87, 35–41.
- Cambre, J., Kulkarni, C., Bernstein, M. S., & Klemmer, S. R. (2014). Talkabout: Small-group discussions in massive global classes. *L@S '14* (pp. 161–162). <http://dl.acm.org/citation.cfm?id=2567859>.
- Chang, M.-M., Lin, M.-C., & Tsai, M.-J. (2013). A study of enhanced structured web-based discussion in a foreign language learning class. *Computers & Education*, 61, 232–241.
- Chen, C.-H. (2010). Promoting college students knowledge acquisition and ill-structured problem solving: Web-based integration and procedure prompts. *Computers & Education*, 55(1), 292–303.
- Cheng, A., Yang, L., & Andersen, E. (2017). *Teaching language and culture with a virtual reality game*. New York, New York, USA: ACM.
- Chen, N.-S., Kinsbuk, Wei, C.-W., & Liu, C.-C. (2011). Effects of matching teaching strategy to thinking style on learner's quality of reflection in an online learning environment. *Computers & Education*, 56(1), 53–64.
- Chen, M.-P., Wang, L.-C., Chen, H.-J., & Chen, Y.-C. (2014). Effects of type of multimedia strategy on learning of Chinese characters for non-native novices. *Computers & Education*, 70, 41–52.
- Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, 80, 108–121.
- Cho, Y. H., & Lee, S. E. (2013). The role of co-explanation and self-explanation in learning from design examples of PowerPoint presentation slides. *Computers & Education*, 69, 400–407.
- Christy, K. R., & Fox, J. (2014). Leaderboards in a virtual classroom: A test of stereotype threat and social comparison explanations for women's math performance. *Computers & Education*, 78, 66–77.
- Chuang, Y.-H., & Tsao, C.-W. (2013). Enhancing nursing students' medication knowledge: The effect of learning materials delivered by short message service. *Computers & Education*, 61, 168–175.
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons.
- Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? *CSCW '14* (pp. 1176–1187). <http://dl.acm.org/citation.cfm?id=2531657>.
- Coetzee, D., Lim, S., Fox, A., Hartmann, B., & Hearst, M. A. (2015). Structuring interactions for large-scale synchronous peer learning. *CSCW '15*(pp. 1139–1152). <http://dl.acm.org/citation.cfm?id=2675251>.
- Corbalan, G., Paas, F., & Cuypers, H. (2010). Computer-based feedback in linear algebra: Effects on transfer performance and motivation. *Computers & Education*, 55(2), 692–703.
- Cox, A., Cairns, P., Shah, P., & Carroll, M. (2012). Not doing but thinking: The role of challenge in the gaming experience. *CHI '12* (pp. 79–88). <http://dl.acm.org/citation.cfm?id=2207689>.
- Culbertson, G., Andersen, E., White, W., Zhang, D., & Jung, M. (2016). Crystallize: An immersive, collaborative game for second language learning. *CSCW '16*(pp. 636–647). ACM.
- Culbertson, G., Shen, S., Jung, M., & Andersen, E. (2017). *Facilitating development of pragmatic competence through a voice-driven video learning interface*. New York, New York, USA: ACM.
- Cunningham, D., & Duffy, T. (1996). *Constructivism: Implications for the design and delivery of instruction*. Handbook of research for educational communications and technology.
- Cutumisu, M., & Schwartz, D. L. (2016). Choosing versus receiving Feedback: The impact of feedback valence on learning in an assessment game. *EDM '16* (pp. 341–346).
- Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G.-J. (2016). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-regulated learning strategies at scale. *EC-TEL '16* (pp. 57–71). Springer.
- Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G.-J. (2017). *Follow the successful crowd - raising MOOC completion rates through social comparison at scale*. LAK.

- De Grez, L., Valcke, M., & Roozen, I. (2009). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education*, 53(1), 112–120.
- Dearman, D., & Truong, K. (2012). Evaluating the implicit acquisition of second language vocabulary using a live wallpaper. *CHI '12* (pp. 1391–1400). . <http://dl.acm.org/citation.cfm?id=2208598>.
- Dodge, B., Whitmer, J., & Frazee, J. P. (2015). Improving undergraduate student achievement in large blended courses through data-driven interventions. *LAK '15* (pp. 412–413). . <http://dl.acm.org/citation.cfm?id=2723657>.
- Eagle, M., & Barnes, T. (2013). Evaluation of automatically generated hint feedback. *EDM '13* (pp. 372–374). . http://www.educationaldatamining.org/EDM2013/papers/rn_paper_87.pdf.
- Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67, 156–167.
- Fossati, D., Eugenio, B. D., Brown, C. W., Ohlsson, S., Cosejo, D. G., & Chen, L. (2009). Supporting computer science curriculum: Exploring and learning linked lists with iList. *IEEE TLT: Vol. 2*, (pp. 107–120). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4906985.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gadiraju, U., & Dietze, S. (2017). *Improving learning through achievement priming in crowdsourced information finding microtasks*. LAK.
- Gamage, D., Whiting, M., Rajapakse, T., Thilakarathne, H., Perera, I., & Fernando, S. (2017). Improving assessment on MOOCs through peer identification and aligned incentives. *The fourth (2017) ACM conference* (pp. 315–318). New York, New York, USA: ACM Press.
- van Gog, T. (2011). Effects of identical example–problem and problem–example pairs on learning. *Computers & Education*, 57(2), 1775–1779.
- González, J. A., Jover, L., Cobo, E., & Muñoz, P. (2010). A web-based learning tool improves student performance in statistics: A randomized masked trial. *Computers & Education*, 55(2), 704–713.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Henry, I., Roediger, L., & Karpicke, J. D. (2016). The power of testing Memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Hooshyar, D., Ahmad, R. B., Yousefi, M., Fathi, M., Horng, S.-J., & Lim, H. (2016). Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills. *Computers & Education*, 94, 18–36.
- Huang, H.-W., Wu, C.-W., & Chen, N.-S. (2012). The effectiveness of using procedural scaffolds in a paper-plus-smartphone collaborative learning context. *Computers & Education*, 59(2), 250–259.
- Hwang, M. L., & Mamykina, L. (2017). *Monster Appetite: Effects of subversive framing on nutritional choices in a digital game environment, effects of subversive framing on nutritional choices in a digital game environment*. New York, New York, USA: ACM.
- Ibanez, M.-B., Di-Serio, A., & Delgado-Kloos, C. (2014). Gamification for engaging computer science students in learning activities: A case study. *IEEE TLT: Vol. 7*, (pp. 291–301). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6827214.
- Imhof, B., Scheiter, K., Edelmann, J., & Gerjets, P. (2013). Learning about locomotion patterns: Effective use of multiple pictures and motion-indicating arrows. *Computers & Education*, 65, 45–55.
- Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education*, 106, 150–165.
- Jeno, L. M., Grytnes, J.-A., & Vandvik, V. (2017). The effect of a mobile-application tool on biology students' motivation and achievement in species identification: A self-determination theory perspective. *Computers & Education*, 107, 1–12.
- Karakostas, A., & Demetriadis, S. (2011). Adaptive domain-specific support to enhance collaborative learning: Results from two studies. *EC-TEL '11* (pp. 210–219). Springer. http://link.springer.com/chapter/10.1007/978-3-642-23985-4_17.
- Kardan, S., & Conati, C. (2015). Providing adaptive support in an interactive simulation for learning: An experimental evaluation. *CHI '15* (pp. 3671–3680). . <http://dl.acm.org/citation.cfm?id=2702424>.
- 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.
- Khandaker, N., & Soh, L.-K. (2010). ClassroomWiki: A collaborative wiki for instructional use with multiagent group formation. *IEEE TLT: Vol. 3*, (pp. 190–202). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5339122.
- Kim, Y.-S., Reinecke, K., & Hullman, J. (2017). *Explaining the Gap: Visualizing One's predictions improves recall and comprehension of data, visualizing One's predictions improves recall and comprehension of data*. New York, New York, USA: ACM.
- Kirschner, P. A. (2017). Stop propagating the learning styles myth. *Computers & Education*, 106, 166–171.
- Kirschner, P. A., & van Merriënboer, J. J. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist*, 48(3), 169–183.
- Kizilcec, R. F., Davis, G. M., & Cohen, G. L. (2017a). Towards equal opportunities in MOOCs. *The fourth (2017) ACM conference* (pp. 121–130). New York, New York, USA: ACM Press.
- Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014b). Showing face in video instruction: Effects on information retention, visual attention, and affect. *CHI '14* (pp. 2095–2102). . <http://dl.acm.org/citation.cfm?id=2557207>.
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017b). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education*, 104, 18–33.
- Kizilcec, R. F., Prez-Sanagustn, M., & Maldonado, J. J. (2016). Recommending self-regulated learning strategies does not improve performance in a MOOC. *L@S '16* (pp. 101–104). . <http://dl.acm.org/citation.cfm?id=2893378>.
- Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014a). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. *eLearning Papers* (pp. 17–26). <http://www.moocsandco.com/sites/default/files/elearning>.
- van der Kleij, F. M., Eggen, T. J. H. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263–272.
- Konert, J., Bellhäuser, H., Röpke, R., Gallwas, E., & Zucik, A. (2016). MoodlePeers: Factors relevant in learning group formation for improved learning outcomes, satisfaction and commitment in e-learning scenarios using GroupAL. *EC-TEL '16* (pp. 390–396). Springer.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations: Two approaches to enhance attention cueing in animations. *Computers & Education*, 55(2), 681–691.
- Krause, M., Mogalle, M., Pohl, H., & Williams, J. J. (2015). A playful game changer: Fostering student retention in online education with social gamification. *L@S '15* (pp. 95–102). . <http://dl.acm.org/citation.cfm?id=2724665>.
- Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015a). PeerStudio: Rapid peer feedback emphasizes revision and improves performance. *L@S '15* (pp. 75–84). . <http://dl.acm.org/citation.cfm?id=2724670>.
- Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M. S., & Klemmer, S. R. (2015b). Talkabout: Making distance matter with small groups in massive classes. *CSCW '15* (pp. 1116–1128). . <http://dl.acm.org/citation.cfm?id=2675166>.
- Kwon, B. C., & Lee, B. (2016). A comparative evaluation on online learning approaches using parallel coordinate visualization. *CHI '16* (pp. 993–997). ACM.
- Labarthe, H., Bouchet, F., Bachelet, R., & Yacef, K. (2016). Does a peer recommender foster students' engagement in MOOCs? *EDM '16* (pp. 418–423). .
- Lamb, A., Smlilack, J., Ho, A., & Reich, J. (2015). Addressing common analytic challenges to randomized experiments in MOOCs: Attrition and zero-inflation. *L@S '15* (pp. 21–30). . <http://dl.acm.org/citation.cfm?id=2724669>.
- Lang, C., Heffernan, N., Ostrow, K., & Wang, Y. (2015). The impact of incorporating student confidence items into an intelligent tutor: A randomized controlled trial. *EDM '15* (pp. 144–149). . http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_152.pdf.
- Lan, C. H., Graf, S., Lai, K. R., & Kinschuk, K. (2011). Enrichment of peer assessment with agent negotiation. *IEEE TLT: Vol. 4*, (pp. 35–46). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5560631.

- Lee, Y.-C., Lin, W.-C., Cherng, F.-Y., Wang, H.-C., Sung, C.-Y., & King, J.-T. (2015). Using time-anchored peer comments to enhance social interaction in online educational videos. *CHI '15* (pp. 689–698). . <http://dl.acm.org/citation.cfm?id=2702349>.
- Li, W., Grossman, T., & Fitzmaurice, G. (2014). CADament: A gamified multiplayer software tutorial system. *CHI '14* (pp. 3369–3378). . <http://dl.acm.org/citation.cfm?id=2556954>.
- Limperos, A. M., Buckner, M. M., Kaufmann, R., & Frisby, B. N. (2015). Online teaching and technological affordances: An experimental investigation into the impact of modality and clarity on perceived and actual learning. *Computers & Education*, *83*, 1–9.
- Lindsey, R. V., Shroyer, J. D., Pashler, H., & Mozer, M. C. (2014). Improving students long-term knowledge retention through personalized review. *Psychological Science*, *25*(3), 639–647.
- Maass, J. K., & Pavlik, P. I., Jr. (2016). Modeling the influence of format and depth during effortful retrieval practice. *EDM '16* (pp. 143–150). .
- Margulieux, L. E., & Catrambone, R. (2014). Improving problem solving performance in computer-based learning environments through subgoal labels. *L@S '14* (pp. 149–150). . <http://dl.acm.org/citation.cfm?id=2567853>.
- Martin, F., & Ertzberger, J. (2013). Here and now mobile learning: An experimental study on the use of mobile technology. *Computers & Education*, *68*, 76–85.
- Mitrovic, A., Ohlsson, S., & Barrow, D. K. (2013). The effect of positive feedback in a constraint-based intelligent tutoring system. *Computers & Education*, *60*(1), 264–272.
- Monserat, T.-J. K. P., Li, Y., Zhao, S., & Cao, X. (2014). L. IVE: An integrated interactive video-based learning environment. *CHI '14* (pp. 3399–3402). . <http://dl.acm.org/citation.cfm?id=2557368>.
- Monserat, T.-J. K. P., Zhao, S., McGee, K., & Pandey, A. V. (2013). NoteVideo: Facilitating navigation of blackboard-style lecture videos. *CHI '13* (pp. 1139–1148). . <http://dl.acm.org/citation.cfm?id=2466147>.
- Mostafavi, B., Eagle, M., & Barnes, T. (2015). Towards data-driven mastery learning. *LAK '15* (pp. 270–274). . <http://dl.acm.org/citation.cfm?id=2723622>.
- Nam, C. W., & Zellner, R. D. (2011). The relative effects of positive interdependence and group processing on student achievement and attitude in online cooperative learning. *Computers & Education*, *56*(3), 680–688.
- Nebel, S., Schneider, S., Beege, M., & Rey, G. D. (2017). Leaderboards within educational videogames: The impact of difficulty, effort and gameplay. *Computers & Education*, *113*, 28–41.
- Ngoon, T., Gamero-Garrido, A., & Klemmer, S. (2016). Supporting peer instruction with evidence-based online instructional templates. *L@S '16* (pp. 301–304). . <http://dl.acm.org/citation.cfm?id=2893439>.
- Nguyen, T. T. D. T., Garnarz, T., Ng, F., Dabbish, L. A., & Dow, S. P. (2017). *Fruitful feedback, in: The 2017 ACM conference*. New York, New York, USA: ACM Press1024–1034.
- Nicholson, J., Huber, M., Jackson, D., & Olivier, P. (2014). Panopticon as an elearning support search tool. *CHI '14* (pp. 1221–1224). . <http://dl.acm.org/citation.cfm?id=2557082>.
- Noroozi, O., Weinberger, A., Biemans, H. J., Teasley, S. D., & Mulder, M. (2012). Fostering multidisciplinary learning through computer-supported collaboration script: The role of a transactive memory script. *EC-TEL '12* (pp. 413–418). . http://link.springer.com/chapter/10.1007/978-3-642-33263-0_36.
- Pandey, V., Amir, A., Debelius, J., Hyde, E. R., Kosciulek, T., Knight, R., et al. (2017). *Gut Instinct: Creating scientific theories with online learners, creating scientific theories with online learners*. New York, New York, USA: ACM.
- Papadopoulos, P. M., Lagkas, T. D., & Demetriadis, S. N. (2012). How to improve the peer review method: Free-selection vs assigned-pair protocol evaluated in a computer networking course. *Computers & Education*, *59*(2), 182–195.
- Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education*, *58*(1), 641–651.
- Poos, J. M., van den Bosch, K., & Janssen, C. P. (2017). Battling bias: Effects of training and training context. *Computers & Education*, *111*, 101–113.
- Pozo-Barajas, R., Pablo-Romero, M. d. P., & Caballero, R. (2013). Evaluating a computer-based simulator program to teach the principles of macroeconomic equilibria. *Computers & Education*, *69*, 71–84.
- Rafferty, A. N., Jansen, R. A., & Griffiths, T. L. (2016). Using inverse planning for personalized feedback. *EDM '16* (pp. 472–477). .
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20–27.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, *20*(9), 1209–1224.
- Rosen, Y., Rushkin, I., Ang, A., Federicks, C., Tingley, D., & Blink, M. J. (2017). Designing adaptive assessments in MOOCs. *The fourth (2017) ACM conference* (pp. 233–236). New York, New York, USA: ACM Press.
- Schneider, B., Jermann, P., Zufferey, G., & Dillenbourg, P. (2011). Benefits of a tangible interface for collaborative learning and interaction. *IEEE TLT: Vol. 4*, (pp. 222–232). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5654494.
- Sedrakyan, G., Snoeck, M., & Poelmans, S. (2014). Assessing the effectiveness of feedback enabled simulation in teaching conceptual modeling. *Computers & Education*, *78*, 367–382.
- Şendağ, S., & Ferhan Odabaşı, H. (2009). Effects of an online problem based learning course on content knowledge acquisition and critical thinking skills. *Computers & Education*, *53*(1), 132–141.
- Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Displaying teachers gaze in a MOOC: Effects on students video navigation patterns. *EC-TEL '15* (pp. 325–338). Springer. http://link.springer.com/chapter/10.1007/978-3-319-24258-3_24.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International journal of instructional technology and distance learning*, *2*(1), 3–10.
- Smith, G. G., Li, M., Drobisz, J., Park, H.-R., Kim, D., & Smith, S. D. (2013). Play games or study? Computer games in eBooks to learn English vocabulary. *Computers & Education*, *69*, 274–286.
- Soflano, M., Connolly, T. M., & Hainey, T. (2015). An application of adaptive games-based learning based on learning style to teach SQL. *Computers & Education*, *86*, 192–211.
- Song, H. S., Pusic, M., Nick, M. W., Sarpel, U., Plass, J. L., & Kalet, A. L. (2014). The cognitive impact of interactive design features for learning complex materials in medical education. *Computers & Education*, *71*, 198–205.
- Sonnenberg, C., & Bannert, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics*, *2*(1), 72–100. <http://eprint.lib.uts.edu.au/journals/index.php/JLA/article/view/4090>.
- Tabuenca, B., Kalz, M., Ternier, S., & Specht, M. (2015). Stop and think: Exploring mobile notifications to foster reflective practice on meta-learning. *IEEE TLT: Vol. 8*, (pp. 124–135). . http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6990417.
- Thompson, M. M., Braude, E., Canfield, C. D., Halfond, J., & Sengupta, A. (2015). Assessment of KNOWLA: Knowledge assembly for learning and assessment. *L@S '15* (pp. 267–271). . <http://dl.acm.org/citation.cfm?id=2728673>.
- Tomkin, J. H., & Charlevoix, D. (2014). Do professors matter?: Using an a/b test to evaluate the impact of instructor involvement on MOOC student outcomes. *L@S '14* (pp. 71–78). . <http://dl.acm.org/citation.cfm?id=2566245>.
- Tritz, J., Michelotti, N., Shultz, G., McKay, T., & Mohapatra, B. (2014). Peer evaluation of student generated content. *LAK '14* (pp. 277–278). . <http://dl.acm.org/citation.cfm?id=2567598>.
- Trusty, A., & Truong, K. N. (2011). Augmenting the web for second language vocabulary learning. *CHI '11* (pp. 3179–3188). . <http://dl.acm.org/citation.cfm?id=1979414>.
- Türkay, S. (2016). The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics. *Computers & Education*, *98*, 102–114.
- Urquiza-Fuentes, J., & Velázquez-Iturbide, J.Á. (2013). Toward the effective use of educational program animations: The roles of student's engagement and topic complexity. *Computers & Education*, *67*, 178–192.
- Wang, P.-Y., Vaughn, B. K., & Liu, M. (2011). The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education*, *56*(1), 300–311.

- Wiese, E. S., Yen, M., Chen, A., Santos, L. A., & Fox, A. (2017). *Teaching students to recognize and implement good coding style*. L@S.
- Williams, J. J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K. Z., Lasecki, W. S., et al. (2016b). AXIS: Generating explanations at scale with learnersourcing and machine learning. *LS '16* (pp. 379–388). . <http://dl.acm.org/citation.cfm?id=2876042>.
- Williams, J. J., Lombrozo, T., Hsu, A., Huber, B., & Kim, J. (2016a). Revising learner misconceptions without Feedback: Prompting for reflection on anomalies. *CHI '16* (pp. 470–474). ACM.
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review*, 20(4), 429–444.
- Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, 9(2), 185–211.
- Xiong, X., Wang, Y., & Beck, J. B. (2015). Improving students' long-term retention performance: A study on personalized retention schedules. *LAK '15* (pp. 325–329). . <http://dl.acm.org/citation.cfm?id=2723636>.
- Yamada, M. (2009). The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: Experimental study. *Computers & Education*, 52(4), 820–833.
- Yang, T.-C., Chen, S. Y., & Hwang, G.-J. (2015). The influences of a two-tier test strategy on student learning: A lag sequential analysis approach. *Computers & Education*, 82, 366–377.
- Yang, T. C., Shih, D. L., & Chen, M. C. (2016). An investigation of the effects of online test strategy on students' learning behaviors. *L@S '16* (pp. 281–284). . <http://dl.acm.org/citation.cfm?id=2893434>.
- Yeomans, M., & Reich, J. (2017). *Planning prompts increase and forecast course completion in massive open online courses*. LAK.
- Zheng, Z., Vogelsang, T., & Pinkwart, N. (2014). The impact of small learning group composition on student engagement and success in a MOOC. *EDM '14* (pp. 500–503). . <http://www.educationaldatamining.org/EDM2015/proceedings/short500-503.pdf>.
- Zhou, G., Wang, J., Lynch, C. F., & Chi, M. (2017). Towards closing the Loop: Bridging machine-induced pedagogical policies to learning theories. *The 10th international conference on educational data mining* (pp. 112–119). .
- Zhu, Y., Wang, Y., Yu, C., Shi, S., Zhang, Y., He, S., et al. (2017). *ViVo: Video-Augmented dictionary for vocabulary learning, video-augmented dictionary for vocabulary learning*. New York, New York, USA: ACM.