The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms

Muhsin Menekse

To cite this article: Muhsin Menekse (2020) The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms, The Journal of Experimental Education, 88:2, 183-199, DOI: 10.1080/00220973.2019.1620159

To link to this article: https://doi.org/10.1080/00220973.2019.1620159

Published online: 12 Jun 2019.

THE JOURNAL OF EXPERIMENTAL EDUCATION
2020, VOL. 88, NO. 2, 183–199
https://doi.org/10.1080/00220973.2019.1620159

Routledge
Taylor & Francis Group

Check for updates

LEARNING, INSTRUCTION, AND COGNITION

The Reflection-Informed Learning and Instruction to Improve Students' Academic Success in Undergraduate Classrooms

Muhsin Menekse

Purdue University, West Lafayette, IN, USA

ABSTRACT

This study addressed the role of the reflection-informed learning and instruction (RILI) model on students' academic success by using CourseMIRROR mobile system. We hypothesized that prompting students to reflect on confusing concepts stimulates their self-monitoring activities according to which students are expected to review their understanding, search for related knowledge, and try to identify the confusing concepts. With this student-reflection information, instructors can thus address students' difficulties effectively, which can lead to enhanced academic success. We tested our hypothesis by conducting a semester-long quasi-experimental study in undergraduate industrial engineering classes ($N = 153$). The analyses revealed that students in the RILI condition performed significantly better than students in the control condition (Cohen's $d = .82$). In addition, reflection analysis showed that both quality and quantity of reflections were significantly associated with exam performance. Surveys indicated users highly valued the RIFI model; they rated CourseMIRROR favorably and said they would continue to use it in future classes.

KEYWORDS

College students; higher education; reflective thinking; achievement; classroom research; engineering education; mobile technologies

STUDIES HAVE SHOWN that classroom environments that actively involve students in their learning process and provide opportunities to reflect on their learning experiences have the potential to enhance students' academic success (e.g., National Research Council, 2012; Tuckman & Kennedy, 2011). However, initiating and managing the active involvement of undergraduate students in introductory science, technology, engineering, and mathematics (STEM) courses is challenging, since these courses are generally taught in large lecture halls due to the number of students enrolled (Mervis, 2013). The sheer size of these lecture classes is often associated with reduced frequency of instructor interaction and feedback to students. Large class sizes also typically result in traditional lecture-based instruction that hinders learning due to the placing of students in mostly passive roles (e.g., Cuseo, 2007; Menekse, Stump, Krause, & Chi, 2013; Walker, Cotner, Baepler, & Decker, 2008). Kokkelenberg and colleagues (Kokkelenberg, Dillon, & Christy, 2008), using a very large data set (764,000 observations) from 1992 to 2004, empirically showed that large class sizes have a negative effect on undergraduate students' academic outcomes. Furthermore, there is an increasing demand for higher education, and financial troubles within universities (i.e., budget cuts by states) make it safe to predict that the class size problem will continue to grow. So, how can we change the passive nature of large-lecture classes to enhance students' academic success?

To address this problem, we explored the effectiveness of a pedagogical model, called reflection-informed learning and instruction (RILI), which utilizes core learning sciences concepts. Also, we developed a mobile learning system by utilizing natural language processing (NLP) to

CONTACT Muhsin Menekse menekse@purdue.edu Neil Armstrong Hall of Engineering, Room 1233 West Lafayette, IN 47907.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/vjxe.

implement the RILI more effectively in college classrooms. Drawing from a self-regulated learning (SRL) perspective (Zimmerman, 2013), we hypothesized that the RILI can potentially enhance student academic success through multiple pathways. In technology-mediated learning contexts, agents and artifacts reciprocally shape both learning environment and outcomes (Overdijk, van Diggelen, Kirschner, & Baker, 2012). For example, students' reflections on their learning experiences help both students and their instructors to engage in diagnosis and to identify gaps or difficulties in comprehension and/or the application of knowledge and skills. Following diagnosis, students might independently engage in a variety of strategic responses to improve their learning such as (a) seeking new information (e.g., identifying supplementary materials/resources); (b) deploying targeted study strategies, such as explanation, elaboration and integration, or extra problem solving practice, individually or in study groups; or (c) seeking additional tutoring support from the instructor, teaching assistants, or peers. Simultaneously, instructors can use diagnostic information from student reflections to provide feedback that scaffolds strategic learning responses (e.g., providing additional information/learning resources, study guides, or targeted tutoring sessions).

In the current study, we hypothesized that prompting students to reflect on confusing concepts, wherein students are expected to review their understanding, search for related knowledge, and try to identify the confusing or missing concepts, stimulates students' self-monitoring activities. With this student reflection information, instructors can thus address students' difficulties effectively, which can lead to enhanced academic success. We tested our hypothesis by conducting a quasi-experimental study in undergraduate industrial engineering classes. Our specific research questions in this study were

1. How does the reflection-informed learning and instruction affect students' academic success?
2. To what degree do the quality and quantity of reflections relate to students' academic success?
3. How do students and the instructor evaluate the reflection-informed learning and instruction model and the associated mobile learning system?

Literature review

Reflection and learning

A review of the term *reflection* in the context of educational sciences reveals that there have been many studies undertaken to explain different ideas about reflection in different domains. Based on multiple definitions provided by Dewey (1933), Boud, Keogh, and Walker (1985); Von Wright (1992); Rogers (2001); and Davis (2003), reflection has been defined as a cognitive process that requires learners to step back from an experience in order to process their existing and new knowledge and/or experiences. In the current study, we specifically operationalized the reflection term similar to Boud and colleagues' (1985) definition in which they described reflection as a fundamental learning activity in which people "recapture their experience, think about it, mull it over and evaluate it" (p. 19).

Reflection is listed as a key component of the self-regulated learning (SRL) processes as well (e.g., Winne, 2001). Zimmerman (1986, 2002) described self-regulated learners as people who are responsible for their own decision-making processes that regulate the selection and use of various forms of knowledge. In addition, self-regulated learners set goals, organize and use strategies to attain these goals, manage and optimize available resources, monitor progress, and continuously reflect (Zimmerman & Kitsantas, 2005). There are multiple SRL models, such as operant, constructivist, or social cognitive, based on different theoretical perspectives. All of these SRL models include student reflection as a fundamental component.

Studies have shown that some students intuitively engage in reflection activities; however, the majority of students need external support to partake in reflection (e.g., Menekse, Stump, Krause, & Chi, 2011; Quinton, Smallbone, & Zimmerman, 2002). Without some guidance or scaffolding, reflection can become ineffective, so expected outcomes may not emerge (Boud & Walker, 1998). Some studies have illustrated the value of learners reflecting on what they have done or engaged in (e.g., Baird, Fensham, Gunstone, & White, 1991; Davis, 2003; Daumiller & Dresel, 2018; Katz, O'Donnell, & Kay, 2000; Lee & Hutchison, 1998; Moreno & Mayer, 2005; Murphy, 2004; Peters & Kitsantas, 2010). In these studies, reflection activities supported students in monitoring their comprehension, enhancing their performance, and facilitating self-regulated learning. For example, Gama (2001) tested the effectiveness of self-reflection on students' knowledge monitoring, evaluation of the learning process, and learning outcomes. She embedded a reflection assistant (RA) into an interactive learning environment (ILE) for algebra word problems. The main role of the RA was to promote students' knowledge monitoring during the problem-solving process. The hypotheses behind the design of the RA were that reflection enhances student's awareness of their own abilities to solve problems and that the selection and use of metacognitive strategies improves as students reflect on their decisions. The experimental group interacted with ILE to translate problems into equations, and they performed reflective activities before and during the problem-solving session. The control group also interacted with ILE but did not perform reflective activities. Results showed that students in the experimental condition solved more problems correctly than the control group did. Also, the experimental group spent more time on tasks and were more persistent when solving problems.

While reflection has been studied in educational psychology literature, a majority of these studies on reflection have been conducted in laboratory settings (e.g., Lee & Hutchison, 1998; McGrath, 2014). Also, some studies are conducted with K-12 students (e.g., Peters & Kitsantas, 2010; Veenman, Kok, & Blöte, 2005), psychology students (e.g., Berthold, Nuckles, & Renkl, 2007), or business students (e.g., Quinton & Smallbone, 2010). On the other hand, there are relatively few investigations in undergraduate STEM classrooms that have explored the role of reflection on the learning of fundamental concepts (e.g., Hirsch & McKenna, 2008; Turns et al., 2015). Recently, the Consortium to Promote Reflection in Engineering Education (CPREE) has drawn attention to how reflective practices are used in engineering classrooms. The CPREE primarily studies the basic mechanism and purpose of critical reflection and educators' roles in developing students' reflection skills (e.g., Csavina, Nethken, & Carberry, 2016; Thomas, Orand, Shroyer, Turns, & Atman, 2016).

To date, the limited reflection studies in STEM education have mostly focused on the nature of students' reflective thinking (e.g., Sabag, Trotskovsky, & Waks, 2014), the role of reflective practice on engineering design (e.g., Adams, Turns, & Atman, 2003; Shekar, 2007), and the reflection activities, especially the "muddiest points" activity (e.g., Brooks, Gilbuena, Krause, & Koretsky, 2014; Carberry, Krause, Ankeny, & Waters, 2013; Menekse et al., 2011; Mosteller, 1989). For example, Adams and her colleagues (2003) explored the question, To what extent do engineering students behave as reflective practitioners? by using Schön's (1983) theory of the professional as a reflective practitioner. They identified characteristics of reflective practices by studying freshman and senior students while they solved a design problem. Brooks and colleagues (2014) studied how to use word clouds to summarize students' written explanations and reflections. However, none of these reflection-focused studies employed experimental or quasi-experimental research design in classrooms to explore the effectiveness of reflection on students' academic success.

Reflection-informed learning and instruction

Prompting students to reflect on their learning experiences after each class throughout the semester could enhance students monitoring skills by supporting and encouraging students to systematically monitor their confusions, performance, and progress and to revise their learning strategies to accomplish their goals (e.g., Carberry, Krause, Ankeny, & Waters, 2013; Zimmerman, 2002). Student reflections can provide diagnostic information to learners and to the instructor about the students' current knowledge and skill and about gaps or problems in the understanding and application of knowledge. This diagnostic information can directly impact students' strategic response choices—for example, to seek more background information on the topic, to engage in extra study, either individually or with peers in study groups, or to seek tutoring help from instructors. Reflections also provide diagnostic information to the instructor that can be used to improve teaching and instructor feedback to students. Instructors can use the reflections to provide more-specific diagnostic feedback (identifying specific sources of learning problems) and to provide helpful learning strategies (directing students to seek additional relevant information, seek out tutoring support, or engage in extra study and practice). Instructors can also use the diagnostic information to directly adapt their teaching, for example, to (a) provide students with extra background information through instructor notes or other library/online resources, (b) adapt pacing and spend more time explaining a topic in greater depth or with additional examples, (c) provide students with extra practice and application problems and give them feedback on problem solutions to help them achieve mastery, and (d) provide extra tutoring sessions to individual students or small groups of students with common problems—for example, holding tutoring sessions to review and discuss annotated, worked solutions to difficult problems. After students have received a cycle of instructor feedback and have been able to engage in strategic responses, the recursive process can and should continue. Students might reflect on the impact of instructor feedback and their engagement in strategic responses, and this information can help instructors further refine their teaching-related decisions and strategies in future cycles.

To evaluate the effectiveness of the reflection-informed learning and instruction on students' academic success in real classroom settings, we conducted a quasi-experimental study with industrial engineering students and gathered semester-long data by utilizing a mobile learning system that was developed to support the reflection-informed learning and instruction in college classrooms.

A mobile learning system to support reflection-informed learning and instruction

In a typical large lecture, prompting students to individually reflect on confusing concepts after each class and summarize these reflections efficiently in a meaningful way is predictably unrealistic. To address this need, we developed a mobile learning system, CourseMIRROR (i.e., mobile in-situ reflections and review with optimized rubrics), and the server-side infrastructure to effectively implement the RILI in college classrooms (Fan, Luo, Menekse, Litman, & Wang, 2015; Luo, Fan, Menekse, Wang, & Litman, 2015). The CourseMIRROR system prompts students to write and submit their reflections on confusing concepts and problems at the end of each class (and/or after the class) for an entire semester by using their own devices (smart phones, tablets, PCs). Then, the natural language processing (NLP) algorithm creates relevant summaries of reflections for each lecture by clustering them based on the common themes. Available to both instructors and students, these summaries allow users to understand the difficulties and misunderstandings that their students or peers encountered from the lecture. Figure 1 shows the interfaces of CourseMIRROR.

Once students log into CourseMIRROR, they can first access their enrolled courses (Figure 1a). By clicking a course item, students have access to the corresponding lectures (Figure 1b). In the lectures list, students can find the lectures for which they can submit reflections. The status icon on the right side of each lecture item shows the status of the lecture (e.g., open for reflection, reflection submitted). For lectures that are open for reflection, clicking the lecture item will lead to the reflection submission page; students can start writing reflections (Figure 1c) using the CourseMIRROR reflection writing page. At the end of each lecture, students receive notifications

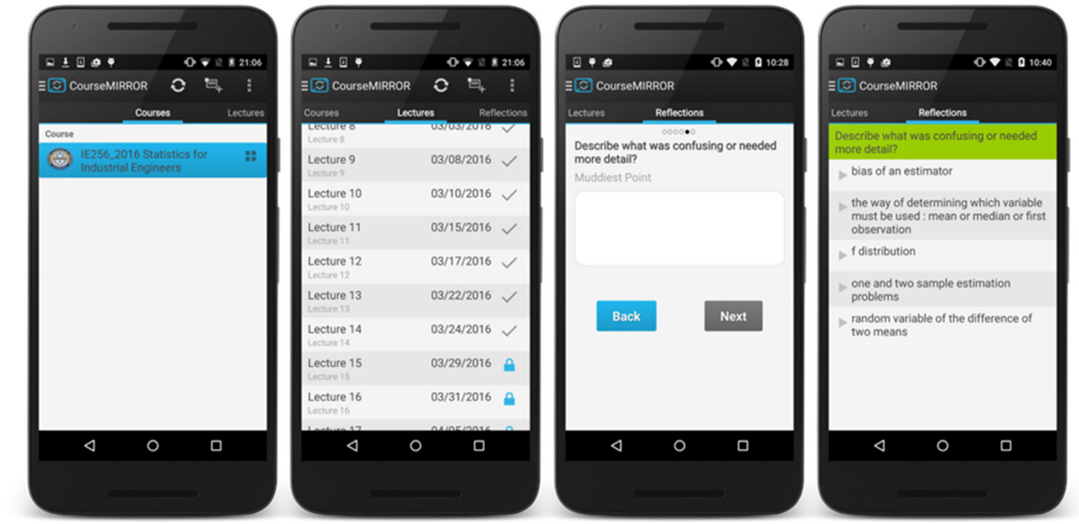

Figure 1. Primary interfaces of CourseMIRROR, left to right: (a) course list, (b) lecture list, (c) reflection writing page, and (d) reflection summary and sharing page.

on their mobile devices to remind them to write reflections for the corresponding lecture. After submission, we leverage the NLP techniques to generate summaries of these reflections. Once the summaries are ready, they are shared with both instructors and students (Figure 1d).

An example of human-generated versus CourseMIRROR summary

By using student reflections, CourseMIRROR generates its summaries using our NLP summarization algorithm. Table 1 shows a set of exemplary student reflections for 30 students and Table 2 shows a human-generated summary created by two different human coders and the CourseMIRROR summary based on the same reflections in Table 1. Human raters were asked to create summaries by clustering the most common issues based on students' reflections on confusing concepts, problems, or activities. As Table 2 indicates, the NLP-generated CourseMIRROR summary includes all the phrases that were included in the human-generated summaries. Prior studies explored the level of agreement between human coders and automated algorithms to create summaries (e.g., Luo & Litman, 2015; Luo, Liu, & Litman, 2016). Luo and colleagues (2016) studied the NLP algorithm of the CourseMIRROR system and found a good level of agreement in summaries created by human coders and the CourseMIRROR NLP algorithms.

Methods

Participants

We implemented the RILI in an engineering class for industrial engineering students. This is a required class for all industrial engineering students, which teaches the fundamentals of statistical methods for collecting, analyzing, and interpreting data. The main topics covered in this class are sampling distributions, estimation, hypothesis testing, regression analysis, nonparametric statistics, application in statistical quality control, and demand forecasting. Also, this is the first engineering statistics class for regularly admitted engineering students; therefore, for almost all students, this was the first time they would learn about these statistics concepts at the college level.

Seventy-four sophomore engineering students participated in the study and participation was voluntary. There were no financial or grade-related incentives to participate. Twenty-three of

Table 1. Sample student reflections (with no editing).

Prompt: Describe what was confusing or needed more detail in this class.
Student Reflections
S1: Nothing
S2: sampling distribution
S3: Distribution of the sample mean is same as the distribution of population, why?
S4: Last 2 questions of PS
S5: problem solutions
S6: ps question 2 and 5
S7: I don't know why, but it was kind of boring and difficult to follow
S8: Problem session
S9: Clt
S10: Not much.
S11:
Sampling distribution but i think it was because i missed the last lecture
S12: The second question of ps
S13: sampling
S14:
Sampling
S15: clt
S16: central limit theorem
S17:
Standard deviation
S18: 30 being the lower limit to convert to clt
S19: Conversion of normal approximation to binomial distribution
S20: all staff. i should work hard.
S21: Ps question 5 was not clearly asked and question 2
S22: sample distributions
S23: we are too slow
S24: Ps questions
S25: the difference between x bar and xi. and the application of central limit theorem using these terms.
S26:
Second question of ps
S27: PS questions
S28: the concept of being greater than 30 for validty of central limit theorem
S29: the difference between x bar and xi.
S30: The difference between x, x(i) and x-bar.

Table 2. Human-generated versus CourseMIRROR summary from the same reflections.

Type of Summarization	Corresponding Example
Human-Generated 1	1. Sampling distribution 2. Problem session questions (#2, #5) 3. Central limit theorem 4. Difference between x-bar and xi
Human-Generated 2	1. PS questions, especially the 2nd one 2. Sampling distribution 3. Concepts related to CLT 4. Difference between x-bar and x-i
CourseMIRROR (Phrase-clustering method)	1. Sampling distribution 2. Last 2 questions of ps 3. Central limit theorem 4. The lower limit to convert to clt 5. The difference between x x i and x-bar

these 74 students were women and 51 were men. In addition, we used baseline data from another group of students ($N = 79$) that had already taken the same class with the same instructor in the previous spring semester. Twenty-four of these 79 sophomore students were women and 55 were men. Students in both semesters were regularly admitted industrial engineering students, and there were no differences between students across semesters in terms of the number of classes/credits that they completed or their prior knowledge. There was no significant difference on students' college level GPA prior to taking this class, $F(1, 145) = .189$, $p = .66$.

This class is offered only in spring academic semesters and the entire cohort of sophomore industrial engineering students must take it in order to enroll in some of the upper level courses. Overall, we had data for 74 students in the intervention group (i.e., the RILI condition) and 79 students for the control condition.

Procedure for the implementation

The instructor included the CourseMIRROR mobile application information in the syllabus and told students at the beginning of the semester that they will be using this mobile application during the semester. Students were asked to install the application into their mobile devices. During the semester, students were prompted to reflect after each class and there were two classes per week for this course. Overall, students were prompted to submit reflections for 21 lectures over 12 weeks of an academic semester. There were no reflection submissions on exam days since there was no instruction on those days. Individual reflections were not shared with the instructor and reflections' contributors' names were unknown to the instructor. Each summary was generated and made available to all students and the instructor within 36 hours after the associated lecture.

We asked the instructor to remind students to submit their reflections at the end of each class. Based on our prior studies we found that instructor involvement is critical for students' participation throughout the semester (Fan, Luo, Menekse, Litman, & Wang, 2017; Luo et al., 2015). Also, we found that student participation was high when students realized that their instructor paid attention to reflection summaries (Fan et al., 2017; Luo et al., 2015). Furthermore, the application has a notification feature, which reminds students to submit their reflections after each class.

We suggested that the instructor read the summaries before each class period and discuss with the whole class the most common issues at the beginning of the following class period. We also suggested that the instructor provide additional learning resources, study guides, or practice problems based on student reflections. However, all of these decisions were at the discretion of the instructor, and the research team had no involvement in making instructional decisions or changes to instructional resources. There were no classroom observations during the implementation.

Measures

The learning measures included first, second, and final exams for all students. The maximum score for each exam was 100 and the minimum was zero. All exams were developed, administered, and graded by the course instructor and her teaching assistants. The instructor and her teaching assistants were not part of the research team. Exam 2 and the final exam were identical for both the RILI and the control groups across two semesters. However, three out of six assessment items on Exam 1 were not identical across two conditions, but they assessed the same statistical concepts. Based on our reliability analysis, the coefficient alpha is (1) .74 for Exam 1 for RILI condition and .72 for control group; (2) .76 for exam 2 (Exam 2 was identical across groups); and (3) .69 for the final exam (final exam was identical across groups). These values for coefficient alphas indicate satisfactory reliability. Reynolds, Livingston, Willson, and Willson (2010) recommended the reliability estimate of .70 or above for instructor-made classroom exams.

We also calculated the overall learning score, which resembled the instructor's method of course grade calculation. It was computed by adding 30% of Exam 1 score, 30% of Exam 2 score, and 40% of the final exam score for each individual student.

Each exam included a combination of short answer, multiple-choice, and true-false type questions. We also conducted a user experience survey at the end of the semester to understand to what degree students valued engaging with the reflection-informed learning and instruction by using the CourseMIRROR system. Another survey was given to the instructor to understand how she used the CourseMIRROR system, how she modified her teaching practices, and whether she planned to use the CourseMIRROR system in future classes.

Data analysis

Learning outcomes across conditions

Since our first research question addressed how the RILI affects students' academic successes, we compared students' exam scores across the control and the RILI conditions by conducting three separate one-way analyses of variances (ANOVAs) for Exam 1, Exam 2, and the final exam. In addition, we conducted another one-way ANOVA by using the students' scores on identical assessment items for Exam 1. Please note that Exam 2 and the final exam were identical for both conditions. Furthermore, we conducted an additional ANOVA by using the overall learning score as the dependent variable. Finally, we conducted a mixed-design ANOVA (repeated measures with a between-subjects factor) to evaluate how students' performance changed over time. In these analysis, we used an alpha of 0.05 as the cutoff for statistically detectable findings.

We also conducted a priori power analysis to determine the required sample size for F tests by using G*Power software (version 3.1). Using effect size of .30, alpha level at .05, and power at .90, the required sample size for one-way ANOVAs was 120, which is lower than our sample size of 153. Therefore, the power analysis indicated that we have a high probability of detecting a true effect by correctly rejecting the null hypothesis.

Table 3. Exemplary reflections for each scoring category for the reflection quality scores.

Reflection Type	Reflection Score	Exemplary Reflections
Specific reflection	4	The decision of sample size for normal distribution confused me
General reflection	3	Questions about hypothesis testing
Vague reflection	2	Chi-square formulation
No-confusion statement	1	There wasn't a confusing point in this lecture
No-reflection submission	0	N/A

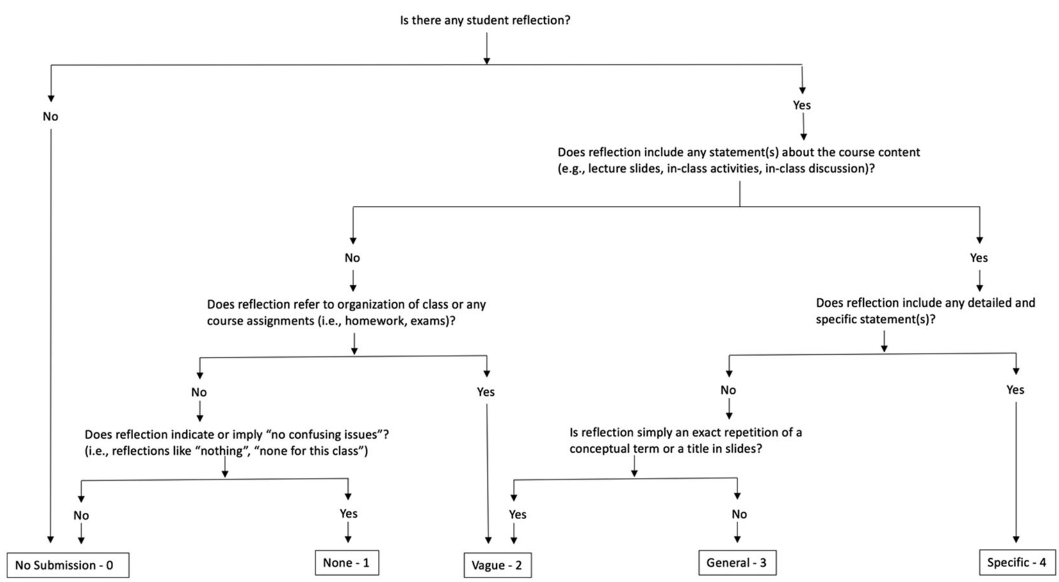

Figure 2. The flowchart for coding students' reflections for the reflection quality scores.

Reflection quality and quantity

Students were prompted to submit reflections for 21 lectures over 12 weeks of an academic semester, but not all students submitted all 21 possible reflections, as their participation was voluntary so we explored how the number of reflections was related to students' learning outcomes. On average, students submitted 12.29 reflections with a standard deviation of 5.89, and the median was 13. In addition, students' reflections were substantially varied in terms of quality. The quality here indicated the completeness and details in one's reflection. In Menekse et al. (2011) study, we developed a coding schema to classify reflections, which follows a scale from 1 to 4 to indicate the degree of deepness or quality of reflection (See Table 3 for the examples of reflections and Figure 2 for the flowchart of the coding schema). We used the same coding schema to classify reflections in this study.

Two raters individually coded students' reflections by using the coding schema. The interrater reliability was calculated and the Cronbach's alpha value was .95; Cohen's kappa was .66. Based on the guidelines from Altman (1990), a kappa of .66 indicates a substantial strength of agreement. Furthermore, since $p < .001$, our kappa coefficient is statistically significantly different from zero. The mean quality score was 2.18 with a standard deviation of .50. Table 3 shows

Table 4. Mean, standard deviation, and standard error values across conditions for all exams.

Exams	Conditions	Mean	Standard Deviation	Effect Size (Cohen's *d*)
Exam 1*	RILI condition	79.88	15.02	.04 (ns)
	Control condition	79.27	13.46	
Exam 1**	RILI condition	39.37	8.85	.09 (ns)
	Control condition	38.67	6.52	
Exam 2	RILI condition	75.36	15.94	.77
	Control condition	62.13	18.11	
Final exam	RILI condition	74.53	12.96	.77
	Control condition	63.61	15.10	

*Indicates data based on all items for Exam 1.
**Indicates data based on the identical items for Exam 1.

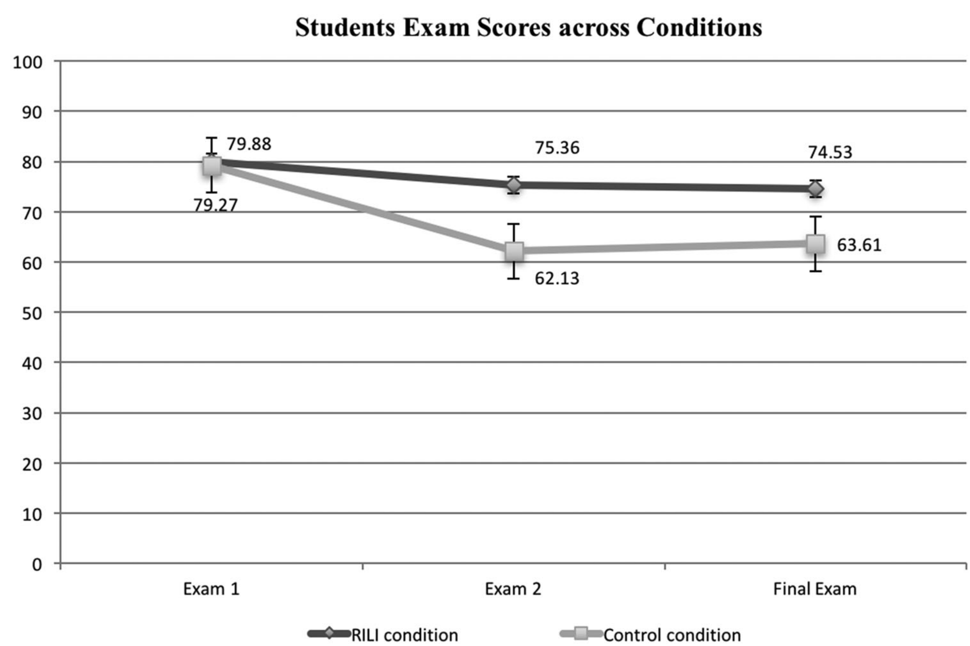

Figure 3. Mean scores of each exam for the RILI and control conditions.

exemplary reflections for different scores. By using the quality of reflections data and number of reflections that students submitted throughout the semester, we conducted a multiple linear regression analysis to explore how quality and quantity of reflection predict students' overall learning score.

Results

Learning outcomes

We conducted three separate one-way ANOVAs to evaluate whether there was a statistically detectable difference between the RILI and control conditions on each exam. The results indicated that there was no statistically detectable difference on Exam 1 (based on all items data), $F(1, 149) = .068$, $p = .79$, or based on the identical items data across conditions, $F(1, 149) = .317$, $p = .57$. On the other hand, students in the intervention condition performed better than the students in the control condition on Exam 2 and the final exam; $F(1, 146) = 21.93$, $p < .001$ and $F(1, 150) = 22.73$, $p < .001$, respectively. The effect size for both Exam 2 and the final exam is .77 (Cohen's d). In addition, the ANOVA for the overall learning score was statistically detectable, $F(1, 142) = 23.88$, $p < .001$, with an effect size of .82 (Cohen's d). Table 4 shows all the values for means, standard deviations, and Cohen's d values across conditions. Figure 3 shows the means and error bars, and Figure 4 shows the histograms for both conditions across exams. We also conducted mixed-design ANOVA (repeated measures with a between subject factor) to evaluate how students' performance changed over time. Results showed that there was a statistically detectable exam main effect, Wilks's lambda = .54, $F(2, 141) = 59.87$, $p < .001$; a statistically detectable difference between conditions, $F(1, 142) = 22.31$, $p < .001$; and a statistically detectable interaction effect for exams and conditions, Wilks's lambda = .81, $F(2, 141) = 16.61$, $p < .001$.

Results for reflection quality and quantity

We conducted a multiple linear regression analysis to evaluate how well the reflection quality and the number of students' reflections predicted the students' overall learning scores. The linear combination of the reflection quality and quantity was related to students' overall learning outcomes, $F(2, 58) = 10.50$, $p < .001$. The sample multiple correlation coefficient was .52, indicating that approximately 27% of the variance of the overall learning outcomes can be accounted for by the linear combinations of the reflection quality and quantity. Table 5 presents the relative strength of the individual predictors. All the bivariate and partial correlations between the reflection quality and quantity, and the overall learning outcomes were statistically detectable.

User experience survey results

At the end of the semester we asked students' opinions on their learning experience with the CourseMIRROR mobile learning system. The survey included eight items with a 5-point Likert scale (e.g., 5 indicates *strongly agree* and 1 indicates *strongly disagree*). It also included two open-ended questions asking for the time it takes to write a reflection and whether students think the benefit of the reflection and feedback is worth the time it takes to use this system. Table 6 shows the means and standard deviations for each survey item. The majority of the students gave positive ratings in terms of the usability and effectiveness of the application. Among the students that chose a negative or a positive option rather than the neutral alternative, 67% of the students indicated that they would like to use the CourseMIRROR in other courses as well; 70% said they had no problems in writing and submitting reflections, and 63% indicated that they benefitted from writing reflections. In addition, most students indicated that they could write a reflection within

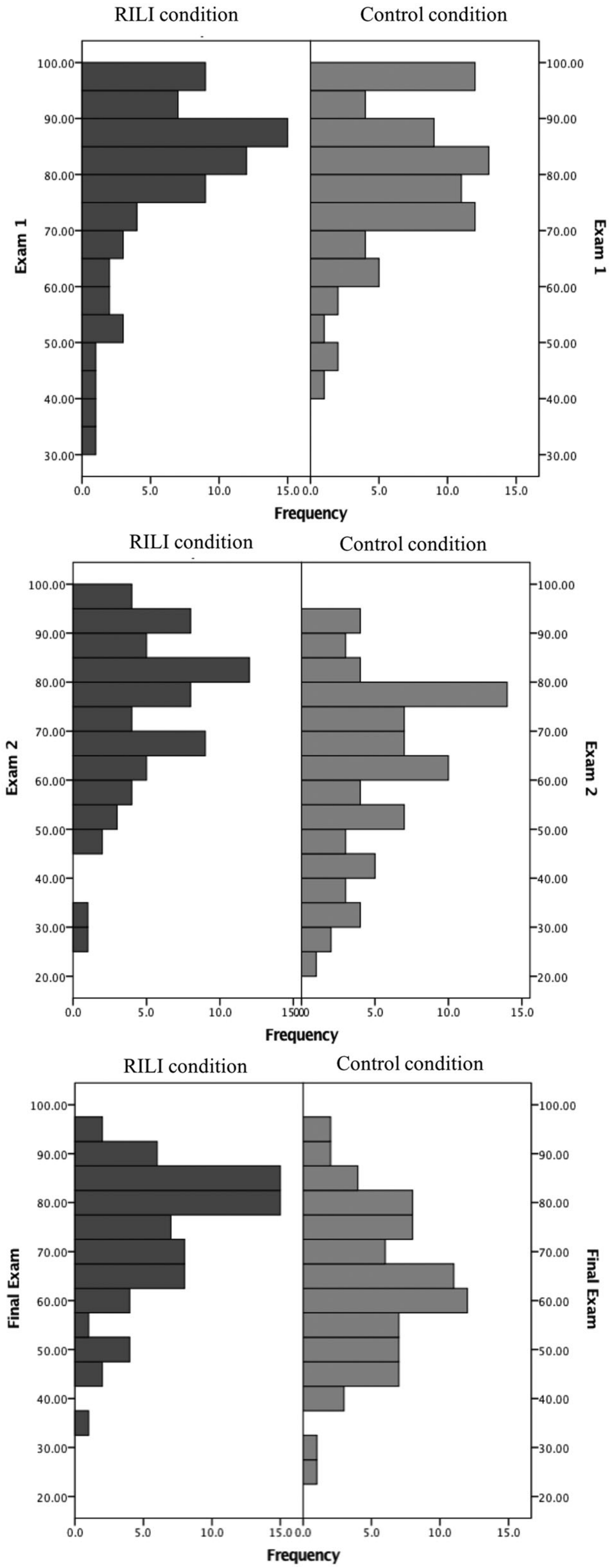

Figure 4. Histograms across the RILI and control conditions for all exams.

Table 5. The bivariate and partial correlations of the predictors with the overall learning score.

Predictors	Bivariate correlation between each predictor and the overall learning score	Partial correlation between each predictor and the overall learning score
Reflection Quality	.28*	.30*
Reflection Quantity	.44**	.45**

*$p < .05$.
**$p < .01$.

Table 6. Students' ratings on the 5-point Likert scale for usability and effectiveness of the CourseMIRROR system.

Survey Item	Mean	Standard Deviation
The CourseMIRROR app is easy to learn and use.	3.90	1.05
I had no trouble in writing and submitting reflections.	3.85	1.07
I wrote and submitted reflections on time.	3.82	1.23
It took me minimal effort to compose and submit reflections.	3.73	.99
I would like to use CourseMIRROR in the future courses.	3.58	1.22
I benefitted from writing reflections.	3.40	1.14
I benefitted from reading the reflection summaries.	3.04	1.17
I often read reflection summaries generated from the classmates' reflections.	2.95	1.23

"2-3 minutes," and a few of them mentioned they needed "at most 5 minutes." Furthermore, average ratings for the items "benefitted from reading the reflection summaries" and "often read the reflection summaries before coming to class" were 2.77 and 2.75, respectively. In addition, the average rating for the item "benefitted from writing reflections" was 3.28. This difference suggests that students perceived that generating reflections was more beneficial compared to reading reflection summaries. Answers to how reflections benefitted students were clustered by researchers according to two categories: (a) improved retention by revisiting what students learned in lecture and (b) made possible for students to receive timely and specific feedback on confusing points.

We asked the instructor to evaluate the CourseMIRROR system and her teaching experiences with it during the semester. Similarly, there were eight Likert-type items and one open-ended question asking how she used the reflection summaries to provide feedback. The Likert-type items included five options and, on average, the instructor rated 4.38 (out of 5) on this survey. The instructor stated that she read the reflection summaries before each class and that these student reflections helped her to improve her teaching strategies. The instructor said that she would continue to use the CourseMIRROR system.

Discussion

This study explored the effectiveness of the reflection-informed learning and instruction model on students' academic success with a quasi-experimental research design. Our findings shed light on how the integration of innovative technologies can improve teaching and learning. By examining students' learning through reflection and feedback loops in a semester-long study, this study attempted to uncover the factors that contribute to students' academic success in engineering classes. How students learn through iterative cycles of reflection and how instructors effectively utilize this information were not yet well understood or studied. Equally important is how instructors use the process of reflective practice to inform and transform instruction. Our primary hypothesis was that prompting students to reflect on confusing concepts stimulates their self-monitoring activities, wherein students are expected to review their understanding, search for related knowledge, and try to identify the confusing or missing elements. With this student-reflection information, instructors can thus address students' difficulties effectively, which can lead to enhanced academic success. The results supported our hypothesis by showing that students' exam scores were significantly improved in the RILI condition compared to the scores from the control condition. Although there was no difference based on the first exam, students in the RILI condition using the Course MIRROR mobile learning system performed better on Exam 2 and the final exam than the students in the baseline condition with a moderate to large effect size of .77. This is an important finding since concepts become more complex over time in this class. While Exam 1 assessed relatively basic concepts such as sampling distributions and data descriptions, Exam 2 and the final exam assessed students' understanding of more-complex topics such as regression analysis and nonparametric procedures. Also, data from the control condition indicated that the students' scores on the second and final exams were less than their scores on the first exam. On the other hand, this drop was minimal and insignificant for students in the RILI condition. Overall, by using the CourseMIRROR learning system as an instructional tool, the course instructor provided opportunities for her students to reflect on their understandings and classroom experiences as they proceeded throughout the semester. These reflective practices complemented by reflection-informed instructional strategies promoted the development of deeper understandings of complex engineering concepts for students.

Furthermore, reflection analysis showed that both the quality and quantity of reflections were positively associated with students' exam scores. This finding is also important in the way that it provides insights and guidelines for future improvements to the CourseMIRROR system. First, there is room to increase the number of reflections that students submit. In this study, the mean number of reflections was around 12 (out of 21). Instructors play a critical role in encouraging students to submit more (quantitatively) and more-timely reflections during the semester. Future studies will need to further explore the role of instructors for successful classroom implementation and for increasing the rate of students' reflection submissions. It could also be helpful to create guidelines regarding how instructors can encourage students to submit timely reflections and how they can utilize the reflection summaries more effectively. Currently, we are in the process of creating web tools for instructors. These web tools will include instructional strategies and specific examples, such as using case studies or problem-solving activities to address students' reflections concerning confusing concepts. The guidelines will also address the nature of students' reflective thinking, how to assess the quality of reflections, and contextual influences that can hinder or enable the development of reflection and reflective thinking. In addition, the user-friendliness aspects of the mobile app could be improved upon to encourage students to submit more reflections, which could then lead to further improvement in students' learning outcomes. User-friendliness aspects could include improvements for the user interface that could make the app more intuitive and easy to navigate. Likewise, the mobile app will be improved by including easy troubleshooting features for users in case something goes wrong with the mobile app.

Furthermore, there is room to improve the quality of reflections. In the current study the mean reflection quality score was 2.18 (and the maximum score was 4.00). One method could be to incorporate adaptive questions to encourage students to write more in-depth reflections. Also, we could provide hints or other scaffolding tools within the mobile system to help students generate more-detailed reflections. Finally, future studies will be needed to evaluate how these changes effect students' reflection behaviors and their learning outcomes.

Survey results indicated that most students and the course instructor considered the CourseMIRROR technology easy to learn and use. The majority of students indicated that they would like to use the CourseMIRROR in other courses, that they benefitted from writing reflections, and that they easily learned the system and had no trouble while writing or submitting reflections. In addition, students stated that generating reflections was beneficial in encouraging them to revisit the key concepts of the course and most students found that the instructor feedback was very helpful for addressing the most common confusing concepts. Likewise, the course instructor rated the CourseMIRROR system favorably and expressed interest in continuing to use it in future classes.

Limitations and future work

While the results provide generally good evidence of the effectiveness of the RILI on improving students' academic success, there are limitations to this study. First, since the research design was quasi-experimental, there was no random assignment of participants at the student level. However, participants in both control and experimental classes were highly similar based on their college level GPA. Also, all participants in both groups were regularly admitted industrial engineering students in that university. Furthermore, this specific class was a required course and not an elective for all industrial engineering majors so there was no possibility of bias for choosing to enroll in this class in terms of students' interest in the subject area and the concepts covered in the class. The second limitation was the relatively small sample size. While a priori power analysis showed that we had enough statistical power to detect a true effect, future replication studies in different classes across different institutions are needed to make stronger claims for our findings. Also, this quasi-experimental study was conducted in two classrooms with one instructor. Further large-scale studies across different courses and multiple instructors are needed to evaluate the transferability of the findings. The third limitation was the novelty effect. The novelty of the mobile application may influence students or instructor's behaviors. However, since our data collection had been continued across 21 lectures throughout the academic semester, there is a good chance that the novelty effect diminished as students and instructor become more familiar with the technology over time. In addition, mobile applications and mobile devices are no longer "novel" for most college students and instructors. A recent survey indicates that 95% of undergraduate students own at least one mobile device and the projections imply that ownership levels will increase steadily (Chen, Seilhamer, Bennett, & Bauer, 2015). Yet, the novelty of the technology for teaching a large-lecture class could still play a role in instructors' teaching behaviors and strategies and further studies are needed to explore the role of novelty on our findings. Additionally, there was no classroom observation in this study. Therefore, we don't have data to evaluate how summaries affected instructors' teaching strategies. In our next studies, we plan to conduct systematic classroom observation to improve understanding of the classroom dynamics and document how reflection summaries are utilized by instructors. Likewise, we will conduct interviews with students to understand how constructing reflections and reading reflection summaries are affecting their study strategies.

Conclusion and implications for practice

By conducting a quasi-experimental study in college classrooms, this study investigated how the reflection-informed learning and instruction affects students' academic success; the relationship between the quality and quantity of reflections and academic success; and how students and instructors value the reflection-informed learning and instruction as a pedagogical model. Our findings showed that the students in the RILI condition performed better than the students in the control condition based on exam scores. In addition, we found that the reflection quality and quantity were significantly related to students' academic success. Finally, the instructor and the majority of the students valued generating reflections and reading summaries, and they rated the mobile learning technology positively in terms of usability and effectiveness. These findings suggest that the RILI has the quality of being an instructional model that educators can use in their classes to enhance students' academic success in lecture-type classes. This instructional model enhances students' critical thinking by using reflections as a method for students to make meaning of what they are learning and what kind of difficulties they are experiencing as they are learning these new concepts. From the cognitive perspective, the reflection activity not only allows students to retrieve prior knowledge and recent learning experiences but also lets students reevaluate their initial understandings. They are better able to differentiate what they already know and what they don't know and to think about their next steps to improve their understanding (Boud, Keogh, & Walker, 1985). Furthermore, the frequent reflection activity could promote the development of a habit of using a critical lens throughout the learning process, which can also enhance students' ability to integrate new understandings with existing knowledge and can transform learners into reflective practitioners over time (Schön, 1983). We believe the reflection-informed learning and instruction plays a significant role in fostering reflective thinking, and our results showed that utilizing student reflections to improve teaching strategies throughout the semester is an effective method of supporting students in achieving a better understanding of fundamental concepts.

Funding

The research reported here was partially supported by the Purdue Research Foundation and the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180477 to Purdue University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

ORCID

Muhsin Menekse http://orcid.org/0000-0002-5547-5455

References

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design Studies*, *24*(3), 275–294. doi:10.1016/S0142-694X(02)00056-X

Altman, D. G. (1990). *Practical statistics for medical research*. Boca Raton, FL: CRC Press.

Baird, J., Fensham, P., Gunstone, R., & White, R. (1991). The importance of reflection in improving science teaching and learning. *Journal of Research in Science Teaching*, *28*(2), 163–182. doi:10.1002/tea.3660280207

Berthold, K., Nuckles, A., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction*, *17*(5), 564–577. doi:10.1016/j.learninstruc.2007.09.007

Boud, D., & Walker, D. (1998). Promoting reflection in professional courses: The challenge of context. *Studies in Higher Education*, *23*(2), 191–206. doi:10.1080/03075079812331380384

Boud, D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning*. Routledge.

Brooks, B. J., Gilbuena, D. M., Krause, S. J., & Koretsky, M. D. (2014). Using word clouds for fast, formative assessment of students' short written responses. *Chemical Engineering Education*, *48*(4), 190–198.

Carberry, A., Krause, S., Ankeny, C., & Waters, C. (2013). "Unmuddying" course content using muddiest point reflections. In 2013 IEEE Frontiers in Education Conference (FIE) (pp. 937–942). IEEE.

Chen, B., Seilhamer, R., Bennett, L., & Bauer, S. (2015). *Students' mobile learning practices in higher education; A multi-year study*. Retrieved from http://er.educause.edu/articles/2015/6/students-mobile-learning-practices-in-higher-education-a-multiyearstudy.

Csavina, K. R., Nethken, C. L. R., & Carberry, A. R. (2016). *Assessing student understanding of reflection in engineering education*. In 123rd American Society for Engineering Education Annual Conference and Exposition, New Orleans, LA.

Cuseo, J. (2007). The empirical case against large class size: Adverse effects on the teaching, learning, and retention of first-year students. *Journal of Faculty Development*, *21*, 5–22.

Daumiller, M., & Dresel, M. (2018). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education*, *42*, 1–16.

Davis, E. A. (2003). Prompting middle school science students for productive reflection: Generic and directed prompts. *The Journal of the Learning Sciences*, *12*(1), 91–142. doi:10.1207/S15327809JLS1201_4

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Boston: DC Heath.

Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. (2015, April). CourseMIRROR: Enhancing large classroom instructor-student interactions via mobile interfaces and natural language processing. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1473–1478). ACM.

Fan, X., Luo, W., Menekse, M., Litman, D., & Wang, J. (2017, March). Scaling reflection prompts in large classrooms via mobile interfaces and natural language processing. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (pp. 363–374). ACM.

Gama, C. (2001). *Metacognition and reflection in ITS: Increasing awareness to improve learning.* In AIED'2001: 10th International Conference on Artificial Intelligence in Education (pp. 492–495).

Hirsch, P. L., & McKenna, A. F. (2008). Using reflection to promote teamwork understanding in engineering design education. *International Journal of Engineering Education*, *24*(2), 377–385.

Katz, S., O'Donnell, G., & Kay, H. (2000). An approach to analyzing the role and structure of reflective dialogue. *International Journal of Artificial Intelligence in Education*, *11*(3), 320–343.

Kokkelenberg, E. C., Dillon, M., & Christy, S. M. (2008). The effects of class size on student grades at a public university. *Economics of Education Review*, *27*(2), 221–233. doi:10.1016/j.econedurev.2006.09.011

Lee, A., & Hutchison, L. (1998). Improving learning from examples through reflection. *Journal of Experimental Psychology Applied*, *4*(3), 187–210. doi:10.1037//1076-898X.4.3.187

Luo, W., Fan, X., Menekse, M., Wang, J., & Litman, D. (2015). Enhancing instructor-student and student-student interactions with mobile interfaces and summarization. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 16–20).

Luo, W., & Litman, D. (2015). Summarizing student responses to reflection prompts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* (pp. 1955–1960).

McGrath, A. L. (2014). Just checking in: The effect of an office hour meeting and learning reflection in an introductory statistics course. *Teaching of Psychology*, *41*(1), 83–87. doi:10.1177/0098628313514186

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2011). The effectiveness of students' daily reflections on learning in engineering context. In *ASEE Annual Conference and Exposition, Conference Proceedings*.

Menekse, M., Stump, G. S., Krause, S., & Chi, M. T. H. (2013). Differentiated overt learning activities for effective instruction in engineering classrooms. *Journal of Engineering Education*, *102*(3), 346–374.

Mervis, J. (2013). Grand challenges in science education: Transformation is possible if a university really cares. *Science*, *340*(6130), 292–296. doi:10.1126/science.340.6130.292

Moreno, R., & Mayer, R. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, *97*(1), 117–128. doi:10.1037/0022-0663.97.1.117

Mosteller, F. (1989). The 'muddiest point in the lecture' as a feedback device. *On Teaching and Learning: The Journal of the Harvard-Danforth Center*, *3*, 10–21.

Murphy, J. (2004). Using focused reflection and articulation to promote clinical reasoning: An evidence-based teaching strategy. *Nursing Education Perspectives*, *25*(5), 226.

National Research Council. (2012). *A framework for k-12 science education: practices, crosscutting concepts, and core ideas. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education*. Washington, DC: The National Academies Press.

Overdijk, M., van Diggelen, W., Kirschner, P. A., & Baker, M. (2012). Connecting agents and artifacts in CSCL: Towards a rationale of mutual shaping. *International Journal of Computer-Supported Collaborative Learning*, *7*(2), 193–210. doi:10.1007/s11412-012-9143-2

Peters, E. E., & Kitsantas, A. (2010). Self-regulation of student epistemic thinking in science: The role of metacognitive prompts. *Educational Psychology*, *30*(1), 27–52. doi:10.1080/01443410903353294

Quinton, S., & Smallbone, T. (2010). Feeding forward: using feedback to promote student reflection and learning–a teaching model. *Innovations in Education and Teaching International*, *47*(1), 125–135. doi:10.1080/14703290903525911

Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and assessment in education*. Texas: Pearson Education International.

Rogers, R. (2001). Reflection in higher education: A concept analysis. *Innovative Higher Education*, *26*(1), 37–57. doi:10.1023/A:1010986404527

Sabag, N., Trotskovsky, E., & Waks, S. (2014). Engineering design projects as a reflection promoter. *European Journal of Engineering Education*, *39*(3), 309–324. doi:10.1080/03043797.2013.867312

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*, *5126*. New York: Basic Books.

Shekar, A. (2007). Active learning and reflection in product development engineering education. *European Journal of Engineering Education*, *32*(2), 125–133. doi:10.1080/03043790601118705

Thomas, L. D., Orand, M., Shroyer, K. E., Turns, J. A., & Atman, C. J. (2016). *Tips & tricks for successful implementation of reflection activities in engineering education.* In 123rd American Society for Engineering Education Annual Conference and Exposition, New Orleans, LA.

Tuckman, B. W., & Kennedy, G. J. (2011). Teaching learning strategies to increase success of first-term college students. *The Journal of Experimental Education*, *79*(4), 478–504. doi:10.1080/00220973.2010.512318

Turns, J. A., Sattler, B., Thomas, L. D., Atman, C. J., Bankhead, R. B., III, Carberry, A. R., … Yasuhara, K. (2015). *Reflecting on reflection: How educators experience the opportunity to talk about supporting student reflection*. In

Proceedings of 122nd American Society for Engineering Education Annual Conference and Exposition, Seattle, Washington.

Veenman, M. V., Kok, R., & Blöte, A. W. (2005). The relation between intellectual and metacognitive skills in early adolescence. *Instructional Science*, *33*(3), 193–211. doi:10.1007/s11251-004-2274-8

Von Wright, J. (1992). Reflections on reflection. *Learning and Instruction*, *2*(1), 59–68. doi:10.1016/0959-4752(92)90005-7

Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: integrating active learning into a large lecture course. *CBE—Life Sciences Education*, *7*(4), 361–367. doi:10.1187/cbe.08-02-0004

Zimmerman, B. (1986). Development of self-regulated learning: Which are the key subprocesses. *Contemporary Educational Psychology*, *16*, 307–313. doi:10.1016/0361-476X(86)90027-5

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–71. doi:10.1207/s15430421tip4102_2

Zimmerman, B. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, *48*(3), 135–147. doi:10.1080/00461520.2013.794676

Zimmerman, B. J., & Kitsantas, A. (2005). Students' perceived responsibility and completion of homework: The role of self-regulatory beliefs and processes. *Contemporary Educational Psychology*, *30*(4), 397–417. doi:10.1016/j.cedpsych.2005.05.003