International Journal of Engineering Education Vol. 38, No. 4, pp. 892–904, 2022
Printed in Great Britain
0949-149X/91 $3.00+0.00

Repeated Use of Adaptive Comparative Judgment to Develop Student Understanding of Artificial Intelligence in Problem Based Learning Assignments*

LISA BOSMAN[1], AASAKIRAN MADAMANCHI[2], SCOTT BARTHOLOMEW[3] and VETRIA BYRD[1]
[1] Purdue University, Purdue Polytechnic Institute, West Lafayette, IN, USA. E-mail: lbosman@purdue.edu, vbyrd@purdue.edu
[2] University of Michigan, School of Information, Ann Arbor, MI, USA. E-mail: amadaman@umich.edu
[3] Brigham Young University, School of Technology, Provo, UT, USA. E-mail: scottbartholomew@byu.edu

Artificial intelligence (AI) is a rapidly developing field with growing importance in engineering, in particular, it serves as a means to better understand and manipulate big data. As educators look to develop more T-shaped engineers, where students have both a breadth and a depth of knowledge and skills, understanding artificial intelligence (AI) applications is extremely important due to its versatility. However, the literature is sparse in how to educate engineers on the use of AI applications. In this paper, the researchers examine the utility of a problem-based learning approach with the well-known supply chain management 'Beer Game' using adaptive comparative judgment (ACJ). ACJ is a mechanism for 'learning by evaluating' through formative iterative comparisons as students develop their understanding of AI applications in supply chain management. The guiding research question was as follows: *Does repeated use of adaptive comparative judgment (as a 'learn by evaluation' tool) lead to enhanced student understanding of artificial intelligence?* Findings provide evidence towards the effectiveness of the 5-week module to improve student perceptions and learning outcomes related to the intersection between supply chain management (SCM) and AI, but only when the treatment and control subgroups were "engaged" students who completed all module requirements. In other words, the use of ACJ 'learning by evaluation' was only found to be statistically significant for students who participated 100%; it was not found to be statistically significant for students who only partially participated. This is a novel finding that extends our understanding of the effectiveness of 'learning by evaluation' for problem-based learning assignments.

Keywords: Problem-based learning (PBL); artificial intelligence; adaptive comparative judgment

1. Introduction

The literature is sparse in how to educate engineers on the use of AI applications. In this paper, the researchers examine the utility of a problem-based learning approach using adaptive comparative judgment (ACJ). ACJ is an approach to evaluation through comparison [1]. Research has shown that repetitive pairwise comparisons, through ACJ, can facilitate student learning [2] and, in this research, we investigated this approach for stimulating student learning and understanding of AI applications in supply chain management. We posited that a 'learning by evaluating' approach may be suitable for this context as it has shown promise in other similar fields [3]. Motivation for integrating AI into the industrial engineering classroom was driven by the desire to better prepare students to enter the Industry 4.0 workforce (manufacturing with a focus on automation, machine learning, real-time data, big data, and interconnectivity).

Adaptive comparative judgment (ACJ) – used in this research as an approach for stimulating student learning through repeated comparative evaluations – is an assessment approach in which a set of items are ranked through a series of holistic comparative judgments between two items at a time [1]. ACJ was originally developed as a summative assessment tool alternative to rubric-based assessment [1, 4–7]. ACJ has proven to be a valid and reliable assessment tool in a variety of disciplinary contexts including writing, design, human development, math, and social studies [5]. More recently, researchers have identified the potential for the use of ACJ as a learning tool for providing formative feedback to students [8–10]. The use of ACJ for 'learning by evaluation' is supported by considerable research within learning sciences which suggests that the act of comparison itself may promote learning by prompting students to identify similarities and differences [11–16]. However, it remains unclear whether 'learning by evaluation' through the iterative comparative judgments, prompted by ACJ, can be formative for students' understanding of a complex technology topic such as AI applications in SCM. Thus, the purpose of this study is to evaluate the use of ACJ as a 'learning by evaluation' tool in this context. The guiding research question was as follows:

Does repeated use of adaptive comparative judgment (as a 'learn by evaluation' tool) lead to enhanced student understanding of artificial intelligence?

* Accepted 2 January 2022.

2. Literature Review

2.1 Problem Based Learning

Problem-based learning (PBL) is an increasingly popular pedagogical approach in which learning is fostered through student-driven development of solutions to real-world problems [17, 18]. In PBL students are presented with an authentic problem that requires the application and development of both domain knowledge and critical thinking skills. Teachers act primarily as facilitators and guides for active, student-directed learning [19]. Traditionally, students have engaged with PBL curricula by working in small groups, although individualized PBL curricula also exist [20]. Originally developed in the context of medical education, PBL is now used in diverse educational contexts ranging from elementary to higher education, and in a broad array of disciplines [21].

Engineering, a discipline focused on real-world applications, has been a natural fit for PBL. Over the past 40 years, a diverse set of PBL practices and forms have been employed within engineering education [22]. PBL can be implemented as a module within a course [23], as an entire course [21], or as a multi-course collaboration [24]. PBL modules have been used to teach topics including industrial statistics [25], lean manufacturing [26], and robotics [24]. Traditionally PBL modules have been deployed in traditional face-to-face classrooms but instructors have also made use of different modalities including web-based deployments and project management workflows [27, 28].

There are several key challenges associated with the use of PBL curricula within engineering education [22]. Foremost among these is fostering appropriate student learning behaviors [29]. Metacognition, or self-reflection on the learning experience is essential to the benefits of PBL, and the learning behaviors that students use in traditional didactic settings are not the same as those needed for PBL [30, 31]. A related issue is the inexperience or lack of training in effective PBL facilitation by most instructors [32]. Finally, a third issue is the challenge of assessing learning within PBL. Because PBL instruction focuses on broad communication, problem-solving, and self-directed learning skills [33] rather than on specific domain knowledge, assessment of student learning gains can be challenging [34, 35]. Often poorly matched traditional assessment methods such as quizzes and exams are used; more recently new assessment approaches including peer review or self-assessment have been employed, but these methods are largely unvalidated [36].

Despite these challenges, there is widespread belief about the benefits of PBL in engineering education [37]. Studies indicate that participation in PBL has led to self-reported increases in interest in engineering, motivation to pursue engineering as a career, and most crucially in engineering skills [38, 39]. These self-reported gains align with data that suggest that participation in PBL is associated with improved academic performance, as well as retention in engineering [40, 41]. Notably, these results were most pronounced in traditionally underrepresented student groups [42].

2.2 Problem Based Learning + Supply Chain Management

The use of problem-based learning (PBL) to teach supply chain management is prolific in the literature. One study implemented PBL into a supply chain management course using five modules with a focus on the newspaper industry [43]. The authors found that the integration of the PBL approach in the SCM course was able to develop students' critical thinking, in-depth technical knowledge, and problem-solving and team working skills. Another study used PBL to introduce lean six sigma concepts in the supply chain classroom [44]. The instructors used PBL projects which included a robust data set that can design distinct problem-scenarios of a complex business problem for specific lean six sigma phases, whereby the team of students was given a process improvement project to identify and address the issue of consumer complaints and decreased revenue. The outcomes imply increased learning outcomes and increased teaching satisfaction with respect to consistency and quality of learning. In a different paper, researchers implemented PBL within a supply chain management class through the development of A3 reports (e.g., a one-page A3 printer-sized document used for communication progress reporting and decision-making) to solve logistics problems [45]. The findings show improvements in learning outcomes, problem-solving skills, and communication.

One of the most popular ways instructors and researchers, alike, have incorporated PBL into the supply chain management classroom has been through the beer distribution game [46, 47] which is discussed next.

2.3 Beer Distribution Game + Supply Chain Management

The Beer Distribution Game is an educational role-playing exercise that has been a staple of supply chain education for decades [48, 49]. In the game, a simplified four-member beer distribution supply chain consisting only of a beer factory, beer wholesaler, beer distributor, and beer retailer is used to illustrate the importance of information sharing, coordination, and scientific inventory management techniques [48]. Studies have demonstrated that naïve gameplay, even with total information transparency among participants, often results in a bullwhip effect – an emergent, complex phenomenon in which progressively larger shortages and surpluses propagate through the supply chain [50-53].

Over the years several educators have developed computer-based [54], phone-based [55], or web-based versions [56–58] of this game. At the same time, researchers investigating AI applications in supply chain management (SCM) have used reinforcement learning approaches to develop algorithmic solutions for playing the beer game while minimizing the bullwhip effects [59, 60]. More recently, Opex Analytics has published a free online Beer Game that allows individuals to play as humans, as AIs, or in a combination [61]. Although there has been research on the effectiveness of teaching supply chain principles through the Beer Game [56, 62, 63], the authors are not aware of research available on the effectiveness of teaching AI or data-guided decision-making to industrial engineering students using the Beer Game.

2.4 Problem Based Learning + Adaptive Comparative Judgment

Adaptive comparative judgment (ACJ) is an assessment approach in which a set of items are ranked through a series of holistic comparative judgments between two items at a time [1]. ACJ was originally developed as a summative assessment tool alternative to rubric-based assessment [1, 4–7]. ACJ has proven to be a valid and reliable assessment tool in a variety of disciplinary contexts including writing, design, human development, math, and social studies [5].

More recently, researchers have identified the potential for the use of ACJ as a learning tool for providing formative feedback to students [8–10]. Further, the use of ACJ as an intentional learning tool for students – referred to as 'learning by evaluation' – has been supported [3]. This approach to engaging students in learning through evaluative comparisons aligns well with other research in learning sciences which suggests that the act of comparison itself may promote learning by prompting students to identify similarities and differences [11–16]. However, the majority of the research into ACJ and learning by evaluating has revolved around essay writing or engineering design; it remains unclear whether 'learning by evaluation' through the iterative comparative judgments, prompted by ACJ, will be impactful for students' understanding of a complex technology topic such as AI applications in SCM.

ACJ is often used as a complement to problem-based learning (PBL). One study investigated the utility of ACJ as a method for informing the teaching and practice of engineering design [64]. The study included qualitative and quantitative methods with 110 undergraduate engineering students from higher education institutes in the United States who were divided into 29 groups to solve an industry-driven, open-ended engineering design challenge. The study findings were; that involving ACJ can provide better awareness to engineering students for informing their design process through peer feedback and peer work comparison, and similarity and differences between the design projects judgments of educators, students, and practice engineers. Another study evaluated the use of ACJ using three panels of judges, from various countries, to evaluate design values [65]. The study included six teachers from the United States who implemented a predetermined design activity and 706 students who were divided into groups to complete an open-ended challenge to design a prototype of a new container for distributing pills. Findings show that "good design" does not come free from cultural contexts, and more care should be placed in stating design criteria requirements. A different paper assessed the use of ACJ to evaluate middle school students learning, engagement, and experience with an open-ended assignment in a technology and engineering education course [66]. The research consisted of 706 middle school students who worked in small groups on a two-week design challenge to complete a design portfolio and produce a solution to an open-ended engineering design challenge. The finding of the study demonstrates the effectiveness of ACJ in grading the students' projects, eliciting the judges' understanding of students' solution process, and reliability and validity of ACJ to assess student learning outcomes in STEM education. Similar to these previous studies, the purpose of this study is to evaluate the use of ACJ as a 'learning by evaluation' tool to support PBL using the beer game application.

3. Methods

3.1 Participants

The study took place at a research-intensive public university in the Midwest United States. Participants were sophomore-level industrial engineering technology students enrolled in a three-credit Supply Chain Management (SCM) Technology course. The course was taught in a hybrid manner where students attend one credit hour of lecture each week and engaged in two credit hours online. The lecture component was split into two sections (60 students per section). All student participants engaged in one central online environment

Fig. 1. Summary of 5 Key Learning Experiences.

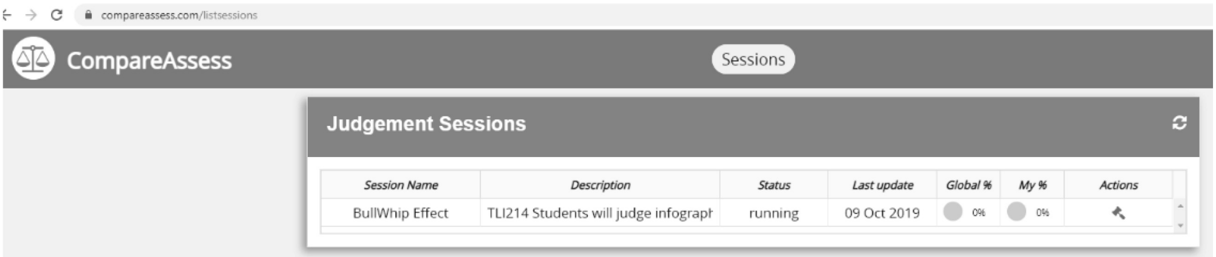

Fig. 2. Compare Assess Entry Screen.

together. A total of 120 students participated in the study.

3.2 Supply Chain Management Contextual Focus

The participants completed a five-week teaching intervention including five key learning experiences, as summarized in Fig. 1. The module included three weeks of the free traditional classic online beer game and two weeks of the free online artificial intelligence (AI) enhanced beer game. For each class, students were assigned homework to develop an infographic that summarizes and communicates the specified AI supply chain concept.

The course was split into two sections. One section (e.g., treatment group) participated in the ACJ sessions, the other section (e.g., control group) used traditional lecture methods to evaluate projects. Both the treatment and control groups consisted of 60 students each.

3.3 Adaptive Comparative Judgment – Assessment Portal

The ACJ assessment was completed through the online portal, www.compareassess.com. Each student participant was provided with their own individual login details. Upon entering the portal, students viewed the screen, and clicked on the gavel located in the Actions column, as shown in Fig. 2, to get started. The next screen provided the viewer with a comparison of two different infographics. The infographics displayed were selected through an embedded algorithm in the ACJ software; this algorithm initially selects items randomly and then, over time, selects pairs of items adaptively based on the win-loss record of each item to refine the resulting rank order of all items. An example is provided in Fig. 3. In each case, the students were prompted to judge which infographic was better based on its ability to fulfill that week's assignment. Each student viewed multiple pairs of items and, in each case, was prompted to explain why they like one infographic over the other by typing in a text window. This process was repeated by all participating students until the judgments were completed, requiring approximately 10–15 minutes for each student.

3.4 Quasi-Experimental Design

This study deployed a quasi-experimental design, which did not include random assignment but instead was based on student enrollment into two different course sections. One section (e.g., treat-

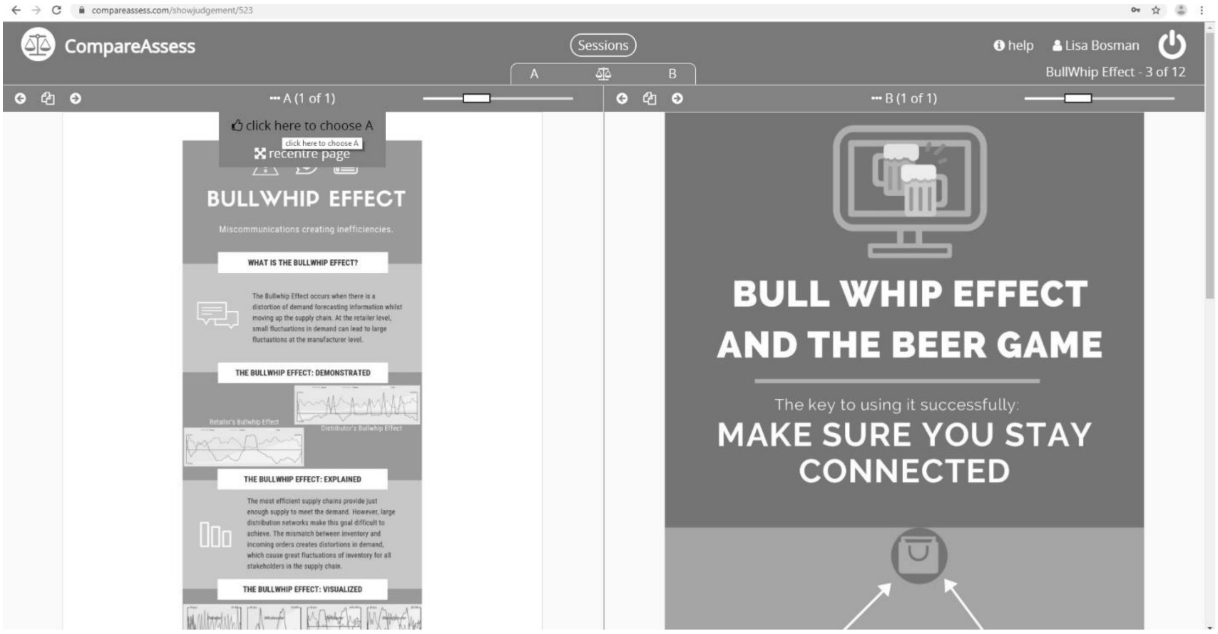

Fig. 3. Example Comparison.

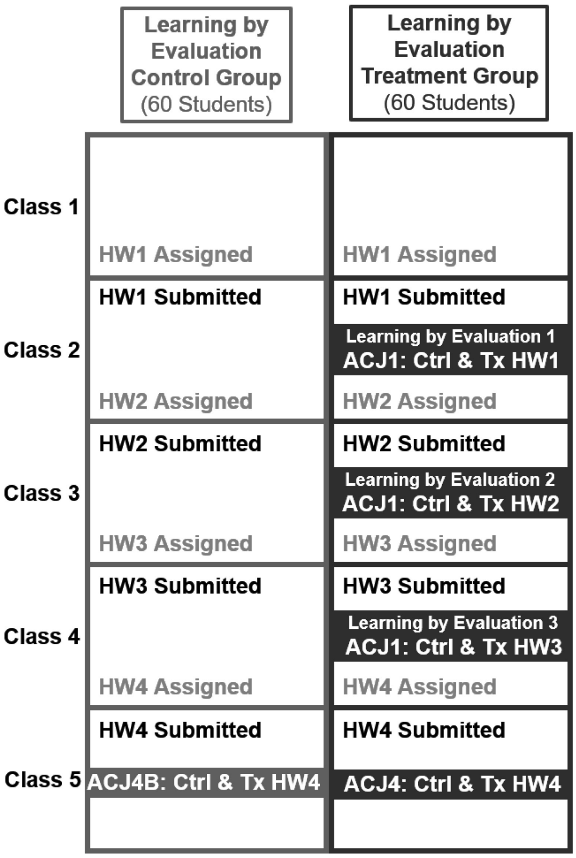

Fig. 4. ACJ Experimental Design.

ment group) participated in the ACJ sessions, the other section (e.g., control group) used traditional lecture methods to evaluate projects. Both the treatment and control groups consisted of 60 students each. The effectiveness of 'learning by evaluation,' through adaptive comparative judgment (ACJ) was assessed and compared. The experimental design is depicted in Fig. 4.

In classes one through four, each section was assigned a homework assignment to develop an infographic that summarizes and communicates an AI supply chain concept. In classes 2 through 4, the Treatment section participated in 'learning by evaluation' through ACJ on the homework assignments (from both treatment and control) from the week before. The control group participated in a traditional lecture-based classroom review during that time.

In addition to our proposed 'learning by evaluation' function, ACJ has an established and validated assessment function. Therefore, in class 5, both the treatment and control groups independently assessed HW4 submissions from both groups through ACJ. The relative rank of the treatment versus control group HW submissions were then compared between HW1 and HW4 to investigate whether 'learning by evaluation' corresponded with an improvement in the treatment group's ability to visually communicate AI Concepts.

3.5 Data Collection

Quantitative data were collected upon student completion of the ACJ sessions using Compare Assess.com. Inter-rater reliability data were analyzed to evaluate the efficacy of the ACJ session. Low inter-rater reliability would imply disagreement (with respect to assignment quality) among the participants, whereas high inter-rater reliability would imply agreement (with respect to assignment quality) among the participants.

4. Results and Discussion

4.1 Reliability Analysis: ACJ Algorithm Results

Table 1 provides the inter-rater reliability for the four ACJ sessions. Cohen [67] suggests interpreting the inter-rater reliability according to the following acceptability scale: scores ≤ 0 as no agreement, scores 0.01 to 0.20 as slight agreement, scores 0.21 to 0.4 as fair agreement, scores 0.41 to 0.6 as moderate agreement, scores 0.61 to 0.8 as substantial agreement, and scores 0.81 to 1 as almost perfect agreement. It is important to note that Cohen's acceptability scale is based on two raters. Since this study had 60 participants for each section, the researchers feel comfortable interpreting using this scale. That being said, two things should be pointed out. First, the high inter-rater reliability score for Session 3 implies a substantial agreement among participants, and the relatively high inter-rater reliability scores for Session 1 and 4 imply a moderate agreement among participants. Second, the relatively low inter-rater reliability score for Session 2 implies a fair agreement among participants. The Session 2 scores were surprising and unexpected. Specific to Session 2, the study researchers (of which includes two teaching team members) brainstormed plausible causes. It was realized that Session 2 was completed at the end of the class with limited time to spare. Although Session 1 was also completed at the end of the class, more time was allotted for students to complete the assessment. As such, the researchers hypothesized that students may have been more interested in leaving class quickly, and as a result, may have randomly clicked through the system instead of evaluating the artifacts effectively. For the remaining Session 3 and Session 4, time was dedicated at the beginning of class to encourage students to respond more thoughtfully within the *CompareAssess* platform. This approach allowed correction of the inter-rater reliability for these remaining sessions.

4.2 ACJ 'Learning by Evaluation': Control vs. Treatment Group

The resulting data, collected from the ACJ sessions, was used in conjunction with our stated research question. An analysis of the data from both Assignment 1 and 4, comparing the control group to the treatment group is included below (see Fig. 7). Additionally, as the assignment participation was varied, a breakdown of group participation is provided in Table 2.

Fig. 5 shows ACJ standardized scores for Assignment 1 and Assignment 4. On the left side of Fig. 5, visual A shows the notch box plots of ACJ standardized scores for Assignment 1. On the right side of Fig. 5, visual B shows the notch box plots of ACJ standardized scores for Assignment 4. The treatment group represents the course section that participated in three rounds of the ACJ 'learning by evaluation' exercise after each of the first three assignments. In contrast, the control group, did not participate in 'learning by evaluation' and instead spent an equivalent amount of course time in an instructor-led discussion. As explained in the Methods, both the treatment and control group participated in ACJ sessions to assess Assignment 4 at the end of the module – these results were used to then identify the differences, if any existed, in the rankings of students following participation in the treatment with those of their peers that did not.

In Fig. 5A we see that the 'learning by evaluation' treatment group started with a lower average student standardized score in comparison to the control group. In Fig. 7B we see that the treatment group, although still with a lower average standardized score in comparison to the control group, has improved relative to the control group by assignment 4.

Fig. 6 includes four visuals, A, B, C, and D. On the top of Fig. 6, visuals A and B display the ACJ-generated standardized scores, along with standard errors, for assignments 1 and 4. The scores are depicted in rank order to allow comparison across the range of student scores. On the bottom of Fig. 6, visuals C and D show density plots showing the distribution of ACJ standardized scores for each student assignment. Fig. 6A and Fig. 6C show that

Table 1. ACJ Results: Inter-rater Reliability

ACJ Session	Inter-rater Reliability	# of rounds
Assessment of HW1: Bullwhip Effect	0.78	19
Assessment of HW2: Benefit of Intelligence	0.31	19
Assessment of HW3: Human-Enabled Decision-Making Algorithm	0.81	18
Assessment of HW4: AI-Enabled Decision-Making Algorithm	0.64	15

Table 2. Breakdown of Group Participation

	Control Group (N = 60)	Treatment Group (N = 60)
Assignment 1	n = 30	n = 35
Assignment 2	n = 42	n = 38
Assignment 3	n = 38	n = 39
Assignment 4	n = 37	n = 36

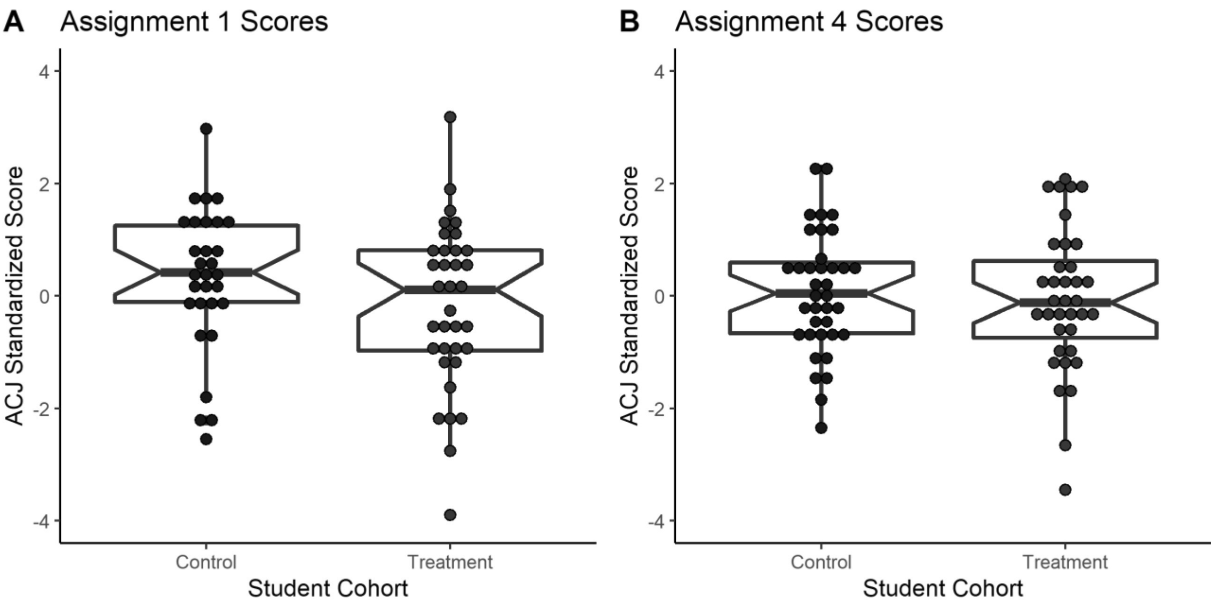

Fig. 5. ACJ Standardized Scores – 'Learning by Evaluation': Notched Boxplot.

the standardized scores were lower for the treatment group in comparison to the control group. In contrast, Fig. 6B and Fig. 6D show that the standardized scores were higher for the treatment group in comparison to the control group. Specifically, Fig. 6C and Fig. 6D showcase a shift to the right for the treatment group, albeit a modest shift, in comparison to the control group. This suggests the 'learning by evaluation' intervention may have resulted in meaningful gains within the treatment group.

In addition to the representations of the data, statistical analysis was completed to test for a significant difference between the control group and treatment on Assignment 4. Since the treatment and group groups were not normally distributed, the non-parametric Mann-Whitney U test (also known as the Wilcoxon test in the R software)

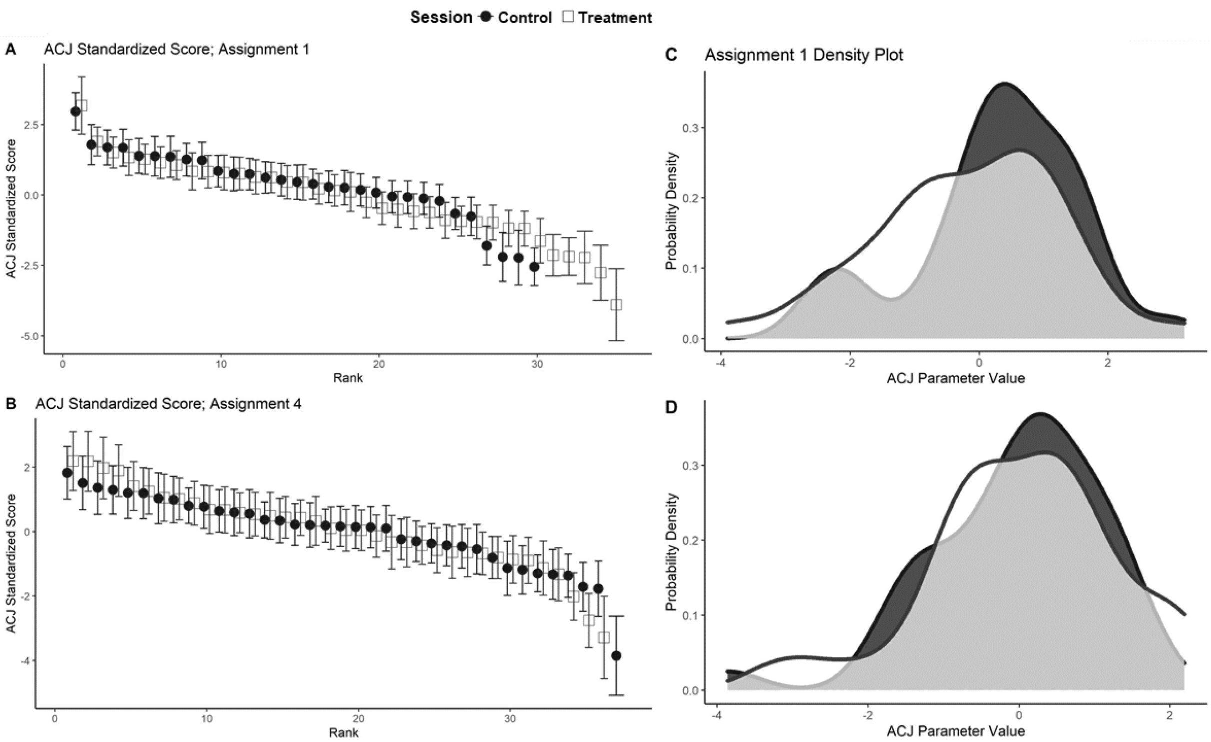

Fig. 6. ACJ Standardized Scores – 'Learning by Evaluation': Ranked Plots & Density Plots.

```
> wilcox.test(AllStudentData$Parameter_Value ~ AllStudentData$Group)

        Wilcoxon rank sum test with continuity correction

data:  AllStudentData$Parameter_Value by AllStudentData$Group
W = 2474, p-value = 0.654
alternative hypothesis: true location shift is not equal to 0
```

Fig. 7. R-Software Output – Test for Difference in Means.

Table 3. Breakdown of "Engaged" Group Participation

	Control Group (N = 60)	**Treatment Group (N = 60)**
Assignment 1	n = 22	n = 21
Assignment 2	n = 22	n = 21
Assignment 3	n = 22	n = 21
Assignment 4	n = 22	n = 21

was used to test for a difference between the treatment group to the control group. Moreover, the Mann-Whitney U test is more rigorous and has fewer false positives than the parametric Student T-test [68]. Fig. 7 shows the resulting statistical significance p-value level was 0.654. Thus, at a 0.05 alpha level, the difference between the groups is not statistically significant.

4.3 ACJ 'Learning by Evaluation': "Engaged" Control vs. Treatment Group

Although not part of the initial research question, the researchers noticed that a subsample of each group (control and treatment) completed all module requirements, including (1) lecture attendance, (2) completion of online discussion, and (3) ACJ assessment or participation in the instructor-led discussion. The completion of module requirements indicates a greater level of engagement with the artificial intelligence module. Thus, the researchers decided to evaluate whether engaged students represented a subgroup who experienced differential gains through the 'learning by evaluation' treatment. A breakdown of this subsample group participation is provided in Table 3. Similar to the previous section, the same analysis was completed with this smaller subset of students.

Recognizing the potential for a subset of data to have issues with validity, the ACJ standardized scores in this subgroup were checked before further investigation in two ways. First, the ACJ rank from this subgroup was compared with the rank produced by all the students (both engaged and others); the resulting correlation was high (Spearman correlation = 0.63). Additionally, the ACJ output from the engaged group was compared with the scoring assigned through traditional classroom assessment (i.e., instructor scores in the grade book). The ACJ standardized scores and the instructor scores were strongly correlated (Spearman correlation = 0.73). After confirming the strong correlations in both of these tests we proceeded with our analysis.

Fig. 8 includes two visuals, A and B. On the left side of Fig. 8, visual A shows the notch box plots of "engaged" ACJ standardized scores for Assignments 1 and 4, comparing the "engaged" subsample of the treatment and control groups. On the right side of Fig. 8, visual B shows the notch box plots of change in ACJ standardized score between assignments 1 and 4 for the "engaged" subsample of treatment and control group.

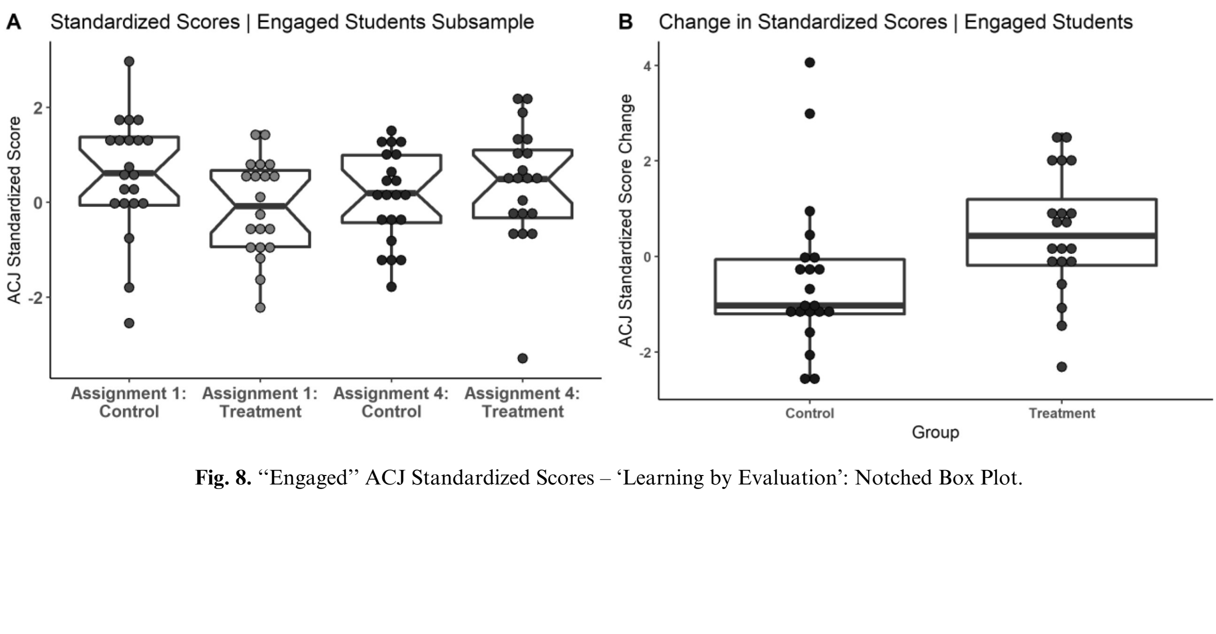

Fig. 8. "Engaged" ACJ Standardized Scores – 'Learning by Evaluation': Notched Box Plot.

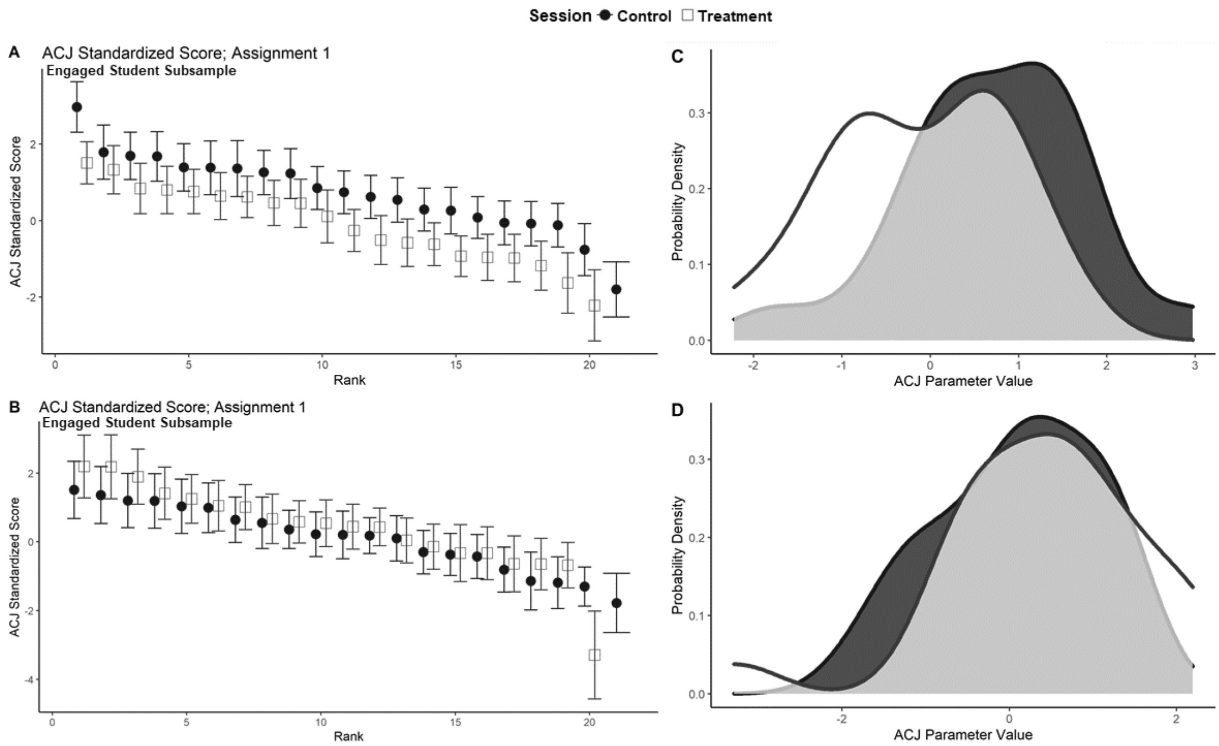

Fig. 9. ''Engaged'' ACJ Standardized Scores – 'Learning by Evaluation': Rank & Density Charts.

In Fig. 8A, similar to Fig. 5A, the ''engaged'' subsample of the treatment group has lower standardized scores for Assignment 1 than the ''engaged'' subsample of the control group. However, after three rounds of 'learning by evaluation', the ''engaged'' subsample of the treatment group scored higher than their counterparts in the control group. In Fig. 8B the change in standardized scores between Assignment 1 and Assignment 4 is displayed for the ''engaged'' sample of the treatment and control groups. The treatment group has a substantially higher change in standardized scores.

Fig. 9 includes four visuals: A, B, C, and D. The rank-ordered standardized scores and standard errors for the ''engaged'' subsample of the treatment and control group are shown at the top of Fig. 9 in visuals A and B. Density plots depicting the distribution of ACJ standardized scores for the ''engaged'' subsample of the treatment and control group are shown at the bottom of Fig. 9 in visuals C and D.

In Fig. 9A the ''engaged'' subsample of the treatment group is shown to have lower scores than the ''engaged'' subsample of the control group across the assignment rank. In Fig. 9B this difference is gone, and in fact, the highest-ranked students in the ''engaged'' subsample of the treatment group scored higher than the similarly ranked ''engaged'' students of the control group. This comparison is confirmed by analysis of the distribution of students' standardized scores in Fig. 9C and Fig. 9D; the distribution of the ''engaged'' subsample of the control groups is largely similar between assignment 1 and assignment 4, but there is a

```
> wilcox.test(TLIdiff$Diff ~ TLIdiff$Group)

        Wilcoxon rank sum test with continuity correction

data:  TLIdiff$Diff by TLIdiff$Group
W = 113.5, p-value = 0.01227
alternative hypothesis: true location shift is not equal to 0

> library(rcompanion)
Warning message:
package 'rcompanion' was built under R version 3.6.2
> wilcoxonR(TLIdiff$Diff, TLIdiff$Group)
     r
 0.394
```

Fig. 10. R-Software Output - Test for Difference in Means and Effect Size for ''Engaged'' Groups.

substantial rightward shift, indicating improved scores, in the distribution for the "engaged" subsample of the treatment group. Together this data indicates that 'learning by evaluation' led to significant gains for the "engaged" students in the treatment group.

Statistical analysis was completed to test for a significant difference between the two groups (see Fig. 10). Because the "engaged" subsample of the treatment and control group was not a normally distributed sample, the non-parametric Mann-Whitney U test (also known as a Wilcox test in the R software) was used to compare the treatment group to the control group. A Mann-Whitney U test yielded a p-value of 0.01227, indicating that the growth in ACJ standardized scores was significantly different in the 'learning by evaluation' treatment group than in the control. Additionally, the effect size was calculated using the R-companion package. The result effect size was r = 0.394, which falls within the medium range for effect sizes.

5. Conclusion

5.1 Practical Implications and Contribution to the Problem-Based Learning Literature

The use of problem-based learning (PBL) to teach supply chain management is prolific in the literature. One of the most popular ways instructors and researchers, alike, have incorporated PBL into the supply chain management classroom has been through the beer distribution game [46, 47]. This study provided one approach for learning about artificial intelligence through a five-week module comparing and contrasting outputs from two different beer distribution games. The module included three weeks of the free traditional classic online beer game and two weeks of the free online artificial intelligence (AI) enhanced beer game. Each week, students were assigned homework to develop an infographic that summarizes and communicates the specified AI supply chain concept. Findings provide evidence towards the effectiveness of the five-week module to improve student perceptions and learning outcomes related to the intersection between supply chain management (SCM) and AI, but only when the treatment and control subgroups were "engaged" students who completed all module requirements. In other words, the use of ACJ 'learning by evaluation' was only found to be statistically significant for students who participated 100%; it was not found to be statistically significant for students who only partially participated. This is a novel finding that extends our understanding of the effectiveness of 'learning by evaluation' for PBL assignments (such as the beer distribution game).

5.2 Limitations

This study has three major limitations. First, the study employed a quantitative approach, where the ACJ process was validated quantitatively. Given the research question and a large number of student participants, an explanatory study is appropriate. However, the study was limited in richness and depth commonly found with a qualitative approach. Second, the study was limited to one Industrial Engineering course (e.g., Supply Chain Management) and one semester. More could have been learned by using different types of engineering courses, with varying sample sizes, diverse student demographics, at different higher education institutions. Third, although the PBL module intended to build student interest in artificial intelligence and to encourage students to enroll in higher-level decision science-based coursework, the study was limited to one semester. The study could have benefited from a longitudinal analysis, where students are tracked to see if the module actually influenced their decision to enroll in additional artificial intelligence or decision science coursework.

5.3 Future Research

Findings of this work support recommendations for future research. First, future research would benefit by utilizing various assessment methods to evaluate broader knowledge gain and understanding of AI, which go beyond data analytics. Second, future research would benefit from considering different aspects of AI. For example, future research could promote student understanding concerning the strengths and weaknesses of AI so students can acquire a more nuanced understanding of the role of AI in SCM. In addition, future research could integrate real-world problems where students can apply AI understanding. Third, future research should consider how to use ACJ and curriculum design as a way to engage students in the learning process. Finally, future research would benefit from replication. This could be done by increasing the number of participants and/or expanding the study longitudinally over a few academic semesters.

References

1. A. Pollitt, The method of adaptive comparative judgement, *Assessment in Education: principles, policy & practice*, **19**, pp. 281–300, 2012.

2. S. R. Bartholomew, N. Mentzer, M. Jones, D. Sherman and S. Baniya, Learning by evaluating (LbE) through adaptive comparative judgment, *International Journal of Technology and Design Education*, 2020/11/21 2020.
3. S. R. Bartholomew and M. D. Jones, A systematized review of research with adaptive comparative judgment (ACJ) in higher education, *International Journal of Technology and Design Education*, pp. 1–32, 2021.
4. S. R. Bartholomew, L. Zhang, E. Garcia Bravo and G. J. Strimel, A Tool for Formative Assessment and Learning in a Graphics Design Course: Adaptive Comparative Judgement, *The Design Journal*, **22**, pp. 73–95, 2019.
5. S. R. Bartholomew and E. Yoshikawa, A systematic review of research around Adaptive Comparative Judgment (ACJ) in K-16 education, *Council on Technology an Engineering Teacher Education: Research Monograph Series*, 2018.
6. R. Kimbell, The origins and underpinning principles of e-scape, *International Journal of Technology and Design Education,* **22**, pp. 123–134, 2012.
7. A. Pollitt, Let's stop marking exams, in *IAEA Conference, Philadelphia*, 2004.
8. N. Seery and D. Canty, Assessment and learning: The proximal and distal effects of comparative judgment, *Handbook of technology education*, pp. 1–14, 2017.
9. N. Seery, D. Canty and P. Phelan, The validity and value of peer assessment using adaptive comparative judgement in design driven practical education, *International Journal of Technology and Design Education,* **22**, pp. 205–226, 2012/05/01 2012.
10. S. R. Bartholomew, G. J. Strimel and A. Jackson, A comparison of traditional and adaptive comparative judgment assessment techniques for freshmen engineering design projects, *International Journal of Engineering Education,* **34**, pp. 20–33, 2018.
11. M. L. Gick and K. Paterson, Do contrasting examples facilitate schema acquisition and analogical transfer?, *Canadian Journal of Psychology/Revue canadienne de psychologie,* **46**, p. 539, 1992.
12. D. L. Schwartz, C. C. Chase, M. A. Oppezzo and D. B. Chin, Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer, *Journal of Educational Psychology*, **103**, p. 759, 2011.
13. P. G. Sidney, S. Hattikudur and M. W. Alibali, How do contrasting cases and self-explanation promote learning? Evidence from fraction division, *Learning and Instruction*, **40**, pp. 29–38, 2015.
14. F. Marton, Sameness and difference in transfer, *The Journal of the Learning Sciences*, **15**, pp. 499–535, 2006.
15. R. Gagné and J. Gibson, Research on the recognition of aircraft, *Motion picture training and research*, pp. 113–168, 1947.
16. D. D. Cummins, Role of analogical reasoning in the induction of problem categories, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18**, p. 1103, 1992.
17. H. S. Barrows, Problem-based learning in medicine and beyond: A brief overview, *New directions for teaching and learning*, **1996**, pp. 3–12, 1996.
18. J. R. Savery and T. M. Duffy, Problem based learning: An instructional model and its constructivist framework, *Educational Technology,* **35**, pp. 31–38, 1995.
19. A. Masek and S. Yamin, The effect of problem based learning on critical thinking ability: a theoretical and empirical review, *International Review of Social Sciences and Humanities*, **2**, pp. 215–221, 2011.
20. A. Burgess, J. Bleasel, I. Haq, C. Roberts, R. Garsia, T. Robertson and C. Mellis, Team-based learning (TBL) in the medical curriculum: better than PBL?, *BMC medical education*, **17**, pp. 1–11, 2017.
21. J. C. Perrenet, P. A. Bouhuijs and J. G. Smits, The suitability of problem-based learning for engineering education: theory and practice, *Teaching in higher education*, **5**, pp. 345–358, 2000.
22. J. Chen, A. Kolmos and X. Du, Forms of implementation and challenges of PBL in engineering education: A review of literature, *European Journal of Engineering Education*, **46**, pp. 90–115, 2021.
23. I. Gratchev and D.-S. Jeng, Introducing a project-based assignment in a traditionally taught engineering course, *European Journal of Engineering Education*, **43**, pp. 788–799, 2018.
24. I. Calvo, I. Cabanes, J. Quesada and O. Barambones, A multidisciplinary PBL approach for teaching industrial informatics and robotics in engineering, *IEEE Transactions on Education*, **61**, pp. 21–28, 2017.
25. M. Darmawan and N. Hidayah, Application of scl-pbl method to increase quality learning of industrial statistics course in department of industrial engineering pancasila university, in *IOP Conference Series: Materials Science and Engineering*, p. 012037, 2017.
26. F. Badurdeen, P. Marksberry, A. Hall and B. Gregory, Teaching lean manufacturing with simulations and games: A survey and future directions, *Simulation & Gaming*, **41**, pp. 465–486, 2010.
27. E. Tambouris, E. Panopoulou, K. Tarabanis, T. Ryberg, L. Buus, V. Peristeras, D. Lee and L. Porwol, Enabling problem based learning through web 2.0 technologies: PBL 2.0, *Educational Technology and Society*, **15**, pp. 238–251, 2012.
28. S. Kizaki, Y. Tahara and A. Ohsuga, Software development PBL focusing on communication using scrum, in *2014 IIAI 3rd International Conference on Advanced Applied Informatics*, pp. 662–669, 2014.
29. S. M. Loyens, J. Magda and R. M. Rikers, Self-directed learning in problem-based learning and its relationships with self-regulated learning, *Educational Psychology Review*, **20**, pp. 411–427, 2008.
30. K. Downing, T. Kwong, S.-W. Chan, T.-F. Lam and W.-K. Downing, Problem-based learning and the development of metacognition, *Higher Education,* **57**, pp. 609–621, 2009.
31. C. Stefanou, J. D. Stolk, M. Prince, J. C. Chen and S. M. Lord, Self-regulation and autonomy in problem-and project-based learning environments, *Active Learning in Higher Education*, **14**, pp. 109–122, 2013.
32. F. D. Salinitri, S. M. Wilhelm and B. L. Crabtree, Facilitating facilitators: Enhancing PBL through a structured facilitator development program, *Interdisciplinary Journal of Problem-Based Learning*, **9**, p. 11, 2015.
33. K. Edström and A. Kolmos, PBL and CDIO: complementary models for engineering education development, *European Journal of Engineering Education*, **39**, pp. 539–555, 2014.
34. P. J. Giabbanelli and A. A. Tawfik, Overcoming the PBL assessment challenge: Design and development of the incremental thesaurus for assessing causal maps (ITACM), *Technology, Knowledge and Learning*, **24**, pp. 161–168, 2019.
35. E. Marti, A. Gurguí, D. Gil, A. Hernández-Sabaté, J. Rocarías and F. Poveda, PBL On Line: A proposal for the organization, part-time monitoring and assessment of PBL group activities, *Journal of Technology and Science Education,* **5**, pp. 87–96, 2015.
36. J. L. M. Nunez, E. T. Caro and J. R. H. Gonzalez, From higher education to open education: Challenges in the transformation of an online traditional course, *IEEE Transactions on Education*, **60**, pp. 134–142, 2016.

37. B. Johnson, R. Ulseth, C. Smith and D. Fox, The impacts of project based learning on self-directed learning and professional skill attainment: A comparison of project based learning to traditional engineering education, in *2015 IEEE Frontiers in Education Conference (FIE)*, pp. 1–5, 2015.
38. L. A. Meadows, R. Fowler, and E. S. Hildinger, Empowering students with choice in the first year, in *American Society for Engineering Education*, 2012.
39. J. E. Mills and D. F. Treagust, Engineering education – Is problem-based or project-based learning the answer, *Australasian Journal of Engineering Education*, **3**, pp. 2–16, 2003.
40. M. Hoit and M. Ohland, The impact of a discipline-based introduction to engineering course on improving retention, *Journal of Engineering Education*, **87**, pp. 79–85, 1998.
41. D. W. Knight, L. E. Carlson and J. F. Sullivan, Staying in engineering: Impact of a hands-on, team-based, first-year projects course on student retention, *Age*, **8**, p. 1, 2003.
42. H. Nguyen, L. Wu, C. Fischer, G. Washington and M. Warschauer, Increasing success in college: Examining the impact of a project-based introductory engineering course, *Journal of Engineering Education,* **109**, pp. 384–401, 2020.
43. W. Sutopo and E. F. Aqidawati, Learning a Supply Chain Management Course by Problem Based Learning: Case Studies in the Newspaper Industry, in *Proceedings of the International Conference on Industrial Engineering and Operations Management, Bangkok, Thailand*, pp. 5–7, 2019.
44. K. E. Miller, C. Hill and A. R. Miller, Bringing Lean Six Sigma to the supply chain classroom: A problem-based learning case, *Decision Sciences Journal of Innovative Education,* **14**, pp. 382–411, 2016.
45. O. S. Silva Filho and R. Calado, Learning supply chain management by pbl with A3 report support, *IFAC Proceedings Volumes*, **46**, pp. 471–477, 2013.
46. J. R. Macdonald, I. D. Frommer and I. Z. Karaesmen, Decision making in the beer game and supply chain performance, *Operations Management Research*, **6**, pp. 119–126, 2013.
47. G. Ravid and S. Rafaeli, Multi Player, Internet and Java-Based Simulation Games: Learning and Research in Implementing a Computerized Version of the ''Beer-Distribution Supply Chain Game'', *Simulation Series*, **32**, pp. 15–22, 2000.
48. J. S. Goodwin and S. G. Franklin, The beer distribution game: using simulation to teach systems thinking, *Journal of Management Development*, **13**, pp. 7–15, 1994.
49. F. Pasin and H. Giroux, The impact of a simulation game on operations management education, *Computers & Education*, **57**, pp. 1240–1254, 2011.
50. J. Nienhaus, A. Ziegenbein and P. Schönsleben, How human behaviour amplifies the bullwhip effect. A study based on the beer distribution game online, *Production Planning & Control,* **17**, pp. 547–557, 2006.
51. H. L. Lee, V. Padmanabhan and S. Whang, The bullwhip effect in supply chains, *Sloan Management Review,* **38**, pp. 93–102, 1997.
52. M. Coppini, C. Rossignoli, T. Rossi and F. Strozzi, Bullwhip effect and inventory oscillations analysis using the beer game model, *International Journal of Production Research*, **48**, pp. 3943–3956, 2010.
53. J. D. Sterman, Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment, *Management Science*, **35**, pp. 321–339, 1989.
54. J. R. Coakley, J. A. Drexler Jr, E. W. Larson and A. E. Kircher, Using a computer-based version of the beer game: Lessons learned, *Journal of Management Education*, **22**, pp. 416–424, 1998.
55. J. M. Day and M. Kumar, Using SMS text messaging to create individualized and interactive experiences in large classes: A beer game example, *Decision Sciences Journal of Innovative Education*, **8**, pp. 129–136, 2010.
56. S. Sarkar and S. Kumar, Demonstrating the effect of supply chain disruptions through an online beer distribution game, *Decision Sciences Journal of Innovative Education*, **14**, pp. 25–35, 2016.
57. J. A. Machuca and R. del Pozo Barajas, A computerized network version of the Beer Game via the Internet, *System Dynamics Review: The Journal of the System Dynamics Society*, **13**, pp. 323–340, 1997.
58. F. R. Jacobs, Playing the beer distribution game over the internet, *Production and Operations Management*, **9**, pp. 31–39, 2000.
59. A. Oroojlooyjadid, M. Nazari, L. Snyder and M. Takáč, A Deep Q-Network for the Beer Game: A Reinforcement Learning Algorithm to Solve Inventory Optimization Problems, *arXiv preprint arXiv:1708.05924*, 2017.
60. S. K. Chaharsooghi, J. Heydari and S. H. Zegordi, A reinforcement learning model for supply chain ordering management: An application to the beer game, *Decision Support Systems,* **45**, pp. 949–959, 2008.
61. O. Analytics, 5 Reasons We Created the Opex Analytics Beer Game, in *Opex-Analytics*, ed: @opexanalytics, 2019.
62. L. William, L. Wu and R. de Souza, Effectiveness of Supply Chain Games in Problem-Based Learning Environment, in *Game-Based Assessment Revisited*, ed: Springer, pp. 257–280, 2019.
63. B. Tundys and A. Rzeczycki, The effectiveness of using the logistics decision games in the educational process – The analysis of the empirical studies, *Research Journal of the University of Gdańsk. Transport Economics and Logistics,* **66**, pp. 43–53, 2017.
64. G. J. Strimel, S. R. Bartholomew, S. Purzer, L. Zhang and E. Y. Ruesch, Informing engineering design through adaptive comparative judgment, *European Journal of Engineering Education*, **46**, pp. 227–246, 2021.
65. S. R. Bartholomew, E. Y. Ruesch, E. Hartell and G. J. Strimel, Identifying design values across countries through adaptive comparative judgment, *International Journal of Technology & Design Education*, **30**, 2020.
66. S. R. Bartholomew, L. S. Nadelson, W. H. Goodridge and E. M. Reeve, Adaptive comparative judgment as a tool for assessing open-ended design problems and model eliciting activities, *Educational Assessment*, **23**, pp. 85–101, 2018.
67. J. Cohen, A coefficient of agreement for nominal scales, *Educational and Psychological Measurement,* **20**, pp. 37–46, 1960.
68. H. M. Samawi and R. Vogel, Notes on two sample tests for partially correlated (paired) data, *Journal of Applied Statistics,* **41**, pp. 109–117, 2014.

Lisa B. Bosman, PhD in Industrial Engineering, is an Assistant Professor in the Department of Technology Leadership and Innovation at Purdue University. Her desire to increase STEM education accessibility and attainment have resulted in her founding of the Purdue University iAGREE Labs (www.iagree.org). Dr. Bosman has authored over 60 publications, obtained over $2M USD in research funding, has been an invited speaker and workshop facilitator for over twenty

national and international engagements, and currently serves as an engineering councilor for the Council of Undergraduate Research (3-year commitment, 2019–2022), and as a division officer for the American Society of Engineering Education (5-year commitment, 2018–2023). In addition, she has recently authored two texts: "Teaching the Entrepreneurial Mindset Across the University: An Integrative Approach" (Springer-Verlag GmbH, 2021) and "Teaching the Entrepreneurial Mindset to engineers (Springer-Verlag GmbH, 2018).

Aasakiran "Bobby" Madamanchi, PhD is a Lecturer III in Data Science within the School of Information at the University of Michigan. His teaching and his research focus on computational workforce development. His interests include democratizing data science and accessible computational workflows, tools and educational methods. Prior to his role at the University of Michigan, he was an independent interdisciplinary postdoctoral fellow in the Future of Work & Learning in the Polytechnic Institute at Purdue University. His PhD is in Cancer Biology from Vanderbilt University.

Scott R. Bartholomew, PhD is an assistant professor of Technology and Engineering Studies at Brigham Young University where he works with aspiring secondary education students. Bartholomew is a former middle school technology teacher who lived in the Philippines for two years as a religious service missionary and, prior to his role at BYU, Dr. Bartholomew was an assistant professor of Engineering/Technology Teacher Education at Purdue University. Bartholomew is involved in several international STEM education efforts including a position as a Fulbright Scholar (Philippines) and a US-Embassy STEM Grant representative (Turkey). His research revolves around teacher professional development, open-ended design assessment, and international STEM education. Bartholomew's hobbies include volleyball, fly-fishing, and he and his wife are the proud parents of four children.

Vetria L. Byrd, PhD is an assistant professor of Computer Graphics Technology in the Polytechnic Institute at Purdue University in West Lafayette, Indiana. Dr. Byrd is the Director of the Byrd Data Visualization Lab where she does research on data visualization pedagogy and develops data visualization curriculum for undergraduate and graduate level courses. Dr. Byrd is the founder and organizer of BPViz: Broadening Participation in Visualization (BPViz) Workshops designed to introduce data visualization to persons interested in building data visualization capacity skills. Dr. Byrd has given numerous invited talks on visualization and has been featured in HPC Wire online magazine (2014), and numerous workshops nationally and internationally. Dr. Byrd received graduate and undergraduate degrees at the University of Alabama at Birmingham, in Birmingham, Alabama which include: Ph.D. in Computer and Information Sciences, Master's degrees in Computer Science and Biomedical Engineering and a Bachelor's degree in Computer Science. Dr. Byrd's research interests include: data visualization, data visualization capacity building, high performance visualization, big data, and collaborative visualization. Dr. Byrd's research utilizes data visualization as a catalyst for communication, as a conduit for collaboration, as a pathway to STEM and as a mechanism for broadening participation and inclusion. http://orcid.org/0000-0002-0733-2062.