International Journal of Engineering Education Vol. 34, No. 2(A), pp. 304–315, 2018
Printed in Great Britain
0949-149X/91 $3.00+0.00
© 2018 TEMPUS Publications.

Teaching Concepts Using Inquiry-Based Instruction: How Well Does Learning Stick?*

MICHAEL PRINCE and MARGOT VIGEANT
Department of Chemical Engineering, Bucknell University, Lewisburg, PA 17837, USA. E-mail: prince@bucknell.edu, mvigeant@bucknell.edu

KATHARYN NOTTIS
Department of Education, Bucknell University, Lewisburg, PA 17837, USA. E-mail: knottis@bucknell.edu

GARY NOTTIS
Bucknell University, Lewisburg, PA 17837, USA. E-mail gnottis@bucknell.edu

This work examines the impact of inquiry-based learning activities (IBLAs) on students' conceptual understanding of four critical heat transfer concepts. Previous research has shown that IBLAs create significant learning gains compared to traditional instruction as assessed by the Heat and Energy Concept Inventory (HECI). A typical end point for assessing student learning is at the end of the activity, where students demonstrate significantly improved understanding relative to before the activity. This paper investigates how durable that understanding is, and how well this understanding transfers to solving new problems in the same conceptual area. The results show that retention of conceptual learning is generally good, with little drop off in most measures of students' conceptual understanding after several weeks. However, the durability of the learning is affected by several factors, including the concept area itself, the difficulty of the concept, and the degree of transfer required. In cases where the learning does not "stick", it is found that students' initial preconceptions continue to hold some sway on their thinking.

Keywords: inquiry-based activities; conceptual learning; transfer

1. Introduction

Developing students' understanding of fundamental concepts is an important goal of engineering and science education. One of the key differences between novices and experts is not only that experts know more, but that they organize what they know around key concepts while novices view what they know as discrete facts. This expert organization facilitates retrieval and use of that information. For these and other reasons, it is important that educators know how to *teach* concepts.

Unfortunately, a body of educational research in the sciences and engineering shows that traditional classes generally do little to promote conceptual change. At the same time, an extensive literature in the sciences and a growing literature in engineering education demonstrate that inquiry-based instruction is significantly more effective than traditional lectures for teaching concepts [1–5]. There is, however, less research examining more nuanced questions, such as the effectiveness of inquiry-based methods for promoting long-term retention of conceptual learning or on how well such activities develop students' ability to transfer what they have learned through the activities to new contexts.

In addition to gaining conceptual understanding, retention and transfer of knowledge are critically important learning outcomes. After all, learning that one does not remember or that one cannot use in new contexts is literally useless. As some researchers have stated, "long term retention and transfer are the first and only goals of education" [6]. It is therefore important for researchers to examine the impact of educational innovations not just on short-term learning, but also on long-term retention of that learning and on students' ability to transfer what they have learned to new situations.

This study does just that by examining the immediate and longer-term impact of inquiry-based instruction on learning in four concept areas related to heat transfer. The Heat and Energy Concept Inventory (HECI) was used to assess pre/post learning gains [7]. Questions on the HECI are coded by a panel of subject matter experts as being either drawn directly from the IBLA (and therefore requiring "near transfer") or as relating to a new situation and therefore requiring "far transfer" of the concept to a novel context. The four concept areas assessed by the HECI are: (1—Radiation) the effect of surface properties on thermal radiation, (2—Temp vs. Energy) confusion between temperature and energy, (3—Temp vs Perception) the distinction between what is perceived when touching and its temperature and (4—Rate vs Amount)

* Accepted 20 December 2017.

factors that influence the "rate" of heat transfer vs. the "amount" of heat transfer. Each of these concepts was identified in the literature as being both important to learn and difficult for students to master. Further descriptions of these concept areas, justification for their inclusion and their assessment are provided in [7].

2. Background

Inquiry-based instruction is generally defined as any teaching method that starts by introducing students to a specific question, problem or dataset and using this to drive the required learning. In short, it starts with the specific and builds toward a broader theoretical framework. It is often contrasted with deductive teaching, which instead starts by introducing students to broad theories and then applying those theories to specific cases, questions or problems. Inquiry-based instruction encompasses a broad range of specific teaching techniques such as problem-based learning, case studies and project-based learning [8, 9], in addition to the specific IBLAs which are the focus of this research.

Inquiry-based instruction has generally been shown to be more effective than traditional instruction for a number of learning outcomes, including conceptual understanding and better knowledge retention. Problem-based learning, for example, has been shown to be more effective than traditional lectures for promoting conceptual understanding and retention of learning related to concepts and principles [10, 11]. The longevity of conceptual learning, or what [12] has referred to as "concept life" (p. 126), is of concern to engineering educators who want their students to maintain their understandings past instruction. Known as the durability of learning [12], there are few studies that explore which factors affect it and the extent to which they do so. There is also little published research quantifying the durability of conceptual learning promoted by inductive methods. For the purposes of this study, durability has been operationally defined as the initial learning gain that is retained after some period of time.

To examine durability, we sought studies that measured students' conceptual understanding prior to—and immediately after—an intervention, and then conducted a follow up study to assess student performance on the same measures at a later point in time. In addition, we sought studies that used validated instruments known to focus on conceptual understanding rather than content knowledge or procedural knowledge. Given our focus on engineering education, we ideally sought studies examining a population drawn from higher education rather than K-12 or adult learners and preferably with engineering students. However, we were unable to find any studies that met this last preference.

We found four studies that met the remaining criteria, each of which focused on physics education [13–16]. All use validated instruments to assess students' conceptual understanding. All either measured or estimated pre-intervention performance on a concept inventory as well as post-measures shortly after an intervention and a long-term assessment to measure retention. A quick summary of those studies suggests several findings. First, consistent with the broader literature, inquiry-based instruction resulted in greater conceptual learning gains than traditional instruction. Second, and most relevant for this study, the retention of students' conceptual understanding after inquiry-based instruction was generally high. That is, student performance on long-term assessments generally showed little drop-off with time. One study found that retention was better with inquiry-based compared to traditional instruction [13]. None of these studies examined the extent to which specific elements of instruction impacted durability of learning or how those elements of instruction impacted students' ability to transfer what they learned.

This study does examine the effectiveness of inquiry-based activities for developing students' ability to transfer what they have learned from inquiry-based instruction. Transfer relates to the students' ability to extend conceptual understanding acquired in one situation to new contexts [17]. The broad literature on inquiry-based instruction suggests that inquiry-based instruction is superior to traditional instruction for promoting students' ability to transfer what they have learned to settings outside the classroom. For example, problem-based learning has been shown to be more effective than traditional instruction for promoting medical students' ability to apply what they've learned in clinical settings [10, 11]. However, there is little research that quantifies how the effectiveness of inquiry-based instruction varies based on the degree of transfer required in the assessment of that effectiveness. There is similarly little research that examines the factors that seem to promote students' ability to transfer what they have learned to new contexts.

3. Methodology

This research continues the authors' examination of inquiry-based activities to repair student misconceptions. The inquiry-based activities employed in the study used the classical conceptual change model [18] which utilizes cognitive conflict to encourage students to change existing preconceptions.

The specific model used here drew heavily on the work by Priscilla Laws and colleagues in the Workshop Physics group [19]. A key element of the model was that students were asked to make predictions that would fail based on commonly held misconceptions and then to have the physical world or a computer simulation demonstrate that their predictions were incorrect. At that point, students were walked through a series of questions to help them identify and repair their initial misconception. The approach used was similar to that proposed by others [20, 21] and has extensive empirical support [19, 22] A more detailed description of the specific activities used in this study is provided in our earlier work [7, 23–26].

A one-group pre-test-post-post-test design was used to assess the retention and transfer of engineering students' conceptual understanding of the targeted heat transfer concepts. The overall testing and activities timeline is illustrated in Fig. 1. The pre-test was taken very early in the semester, the immediate post-test was taken within a week of doing the inquiry-based activities and the retention-test was taken at the end of the semester. Conceptual comprehension was assessed using a sub-set of questions, referred to as a "mini-test", from the relevant concept subscale of *The Heat and Energy Concept Inventory* (HECI, [7]). The HECI is one of several concept inventories—validated multiple-choice instruments designed to assess conceptual understanding rather than factual information or problem-solving skills—developed for engineering topics. It has 36 questions covering the four targeted concept areas examined in this study. The instrument has demonstrated acceptable levels of internal consistency reliability and content validity in previous research [7]. In the pre- and retention-tests, students complete the full HECI on all concept areas. Immediately following inquiry-based activities, students only complete the mini-test relevant to the concept area of the inquiry-based activities. It should be noted that the timing on the mini-test varies by concept area, since concepts are introduced throughout the course. While the timing of the mini-tests with respect to their activities is consistent across concepts and participating institutions, the time between activities and the retention-test necessarily varies.

Questions on the mini-tests were characterized as measuring either "near" or "far" transfer. Because "near" and "far" transfer have been defined in multiple ways in the literature [27], it is important to note how these terms were defined in the current study. Near transfer was operationally defined as a question identical to a situation experienced in the inquiry-based activities as determined by a panel of three content-experts. Far transfer was operationally defined as a question that went beyond the scenario seen in the inquiry-based activities but that used the same core concept. While the questions classified as far transfer do not meet the multiple dimensions suggested by Barnett and Ceci [27], we hypothesized that this distinction in the degree of transfer required between questions might impact retention of students' learning and sought to test that hypothesis through this study.

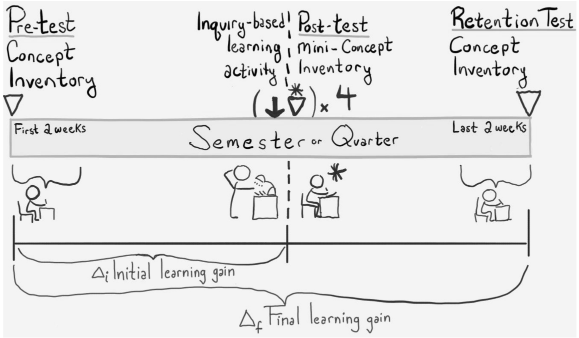


Fig. 1. Timeline of concept inventories and IBLAs.

Questions were also characterized with respect to difficulty, defined as the percentage students getting a specific question correct on the initial pre-test. Descriptive statistics were employed to examine changes in knowledge, as measured by the overall scores of the questions in each concept area at each point of data collection. Paired sample *t*-tests were utilized to determine the statistical significance of changes in knowledge from pre- to post-test, pre- to retention-test, and post to retention test. Analysis of Covariance (ANCOVA) was used to examine the differences between near and far transfer questions when pre-test scores were controlled. Analysis of Covariance (ANCOVA) is used to "partial out" variables believed to be impacting the results. It is thought to be a more appropriate analysis for settings where individuals have not been randomly assigned [28] or randomly selected. Cohen's d, utilized for *t*-tests, was employed to determine the effect size when the dependent *t*-tests were computed. According to Fraenkel, Wallen, and Hyun [29], any effect size of 0.50 or larger [for Cohen's d], "is an important finding" (p. 248). With the Analysis of Covariance, partial Eta Squared was used to calculate effect sizes, since this is what is used with Analysis of Variance Models [30]. Eta squared, ". . . is commonly viewed as the proportion of variance in the dependent variable explained by the independent variable(s) in the sample" [28], p. 76.

A sample of convenience was used in the current study. Participating courses were all undergraduate level engineering courses focused on heat transfer. The number of participating institutions and students for each concept and assessment is summar-

Table 1. Summary of Experimental Design

	Radiation (7 institutions)	Rate vs. Amount (6 institutions)	Temperature vs. Energy (8 institutions)	Temperature vs. Perception (11 institutions)
Pre-Test n, full HECI	236	372	215	327
Immediate Post-Test Mini Concept Inventory n	264	301	215	327
Retention-Test n, full HECI	228	317	215	327

ized in Table 1. Three of the schools assessed students in multiple semesters. A preliminary version of a subset of this work was shared at a conference and appears in [26].

4. Results

Previous research has demonstrated that the inquiry-based activities used in this study increase students' conceptual understanding relative to traditional instruction [7, 23–26]. This study focuses on how the activities impact both the retention of learning and ability to transfer what is learned using the activities. We begin with an overview of the impact of the activities on students' initial and long-term learning. We include an analysis of how learning varies by concept area. We then examine how additional factors influence students' learning and retention. Specifically, we look at the effect of initial question difficulty on short-term and long-term learning. We then examine how the degree of "transfer" required impacts our results. "Transfer" is defined in terms of how much the assessment questions differed from the specific context of the inquiry-based activity. Finally, we examine the extent to which students' initial preconception holds sway for those students for whom learning does not "stick".

4.1 Overview of primary results

Table 2 provides the question by question data and shows several items relevant to the current study. First, it indicates how many questions were used to assess understanding in each concept area along with whether that question required near or far transfer of learning, as determined by the panel of experts. Results are presented in Table 2 for students score on the initial pre-test during the first week of the semester, the immediate post-test after the inquiry-based activity and on the retention-test at the end of the semester. Sample questions from each mini-test are shown in the Appendix.

A visual inspection of the data in Table 2 shows that there is significant variation in the question difficulty (% correct), the initial learning gain from pre- to post-test, and with the durability of learning as measured by a change from the post-test to retention test. Results from Table 2 aggregated by concept area are shown in Table 3 to illustrate how student performance varied for each concept.

In Table 3 we have deliberately separated the "Rate vs. Amount" questions into either "Amount" questions (questions requiring energy *balances*) and "Rate" questions, (questions which focused on how *fast* energy was being transferred) since we noticed a clear distinction in student

Table 2. Learning Gains and Retention by Concept Question

Question	Transfer (Near or Far)	% Correct on Pre-test	% Correct on Post-test	% Correct on Retention-test
Radiation 1	Near	30.9	93.2	87.3
Radiation 2	**Near**	**13.6**	65.5	60.5
Radiation 3	Far	42.4	59.5	61.4
Radiation 4	Far	34.3	37.5	43.4
Rate vs. Amount 1	Near	31.3	90.8	78.3
Rate vs. Amount 2	Near	40.6	89.4	86.2
Rate vs. Amount 3	**Near**	**12.0**	71.4	37.8
Rate vs. Amount 4	**Near**	**18.0**	72.8	44.2
Rate vs. Amount 5	Near	47.9	85.7	81.6
Rate vs. Amount 6	Near	48.8	83.9	80.6
Rate vs. Amount 7	Far	51.2	84.8	78.3
Rate vs. Amount 8	Far	28.1	57.6	40.6
Temperature vs. Energy 1	Near	28.8	77.7	58.6
Temperature vs. Energy 2	**Far**	**24.2**	51.2	47.0
Temperature vs. Energy 3	*Far*	*68.4*	76.7	77.7
Temperature vs. Perceptions 1	*Far*	*60.6*	90.5	86.2
Temperature vs. Perceptions 2	*Far*	*74.6*	90.5	90.2
Temperature vs. Perceptions 3	*Far*	*60.9*	95.5	90.5

Questions defined as "easy" (more than 60% correct) are shown in *italics*, while questions defined as "difficult" (less than 25% correct) are shown in **bold**.

Table 3. Learning Gains and Retention by Concept Area

Concept Area	Mean % Correct on Pre-test	Mean % Correct on Post-test	Mean % Correct on retention-test	Retention (% of initial gain)
Radiation	30.3	63.9	63.2	98%
Amount	44.0	86.9	81	86%
Rate	19.3	67.3	40.9	45%
T vs. Energy	40.5	68.5	61.1	74%
T vs. Perception	65.3	91.8	89.0	89%
Total	38.4	75.2	66.2	78%

performance between these two categories of questions, as illustrated in the table. Table 3 shows that the student performance increased significantly in all concept areas, both in the immediate post-test and the final retention test. Retention rates, or how well students remember what they learned from pre- to post- tests, are shown in the final column of Table 3. Retention was defined as the "final" learning gain (Δ_f assessed at the end of the semester) divided by the "initial" learning gain, (Δ_i, assessed immediately after the activity), as defined in Fig. 1. The final learning gain was calculated as the difference between the mean % correct on the retention-test and the pre-test. Similarly, the initial learning gain was calculated as the difference between mean % correct on the post-test and pre-test. A retention score of 50%, therefore, means that the learning gains measured at the end of the semester were half of the gains measured immediately after the activity. A summary of the results from significance testing using paired samples *t*-tests is provided in Table 4, followed by detailed statistics; questions on a single concept or sub-concept are referred to as a "mini-test".

As in previous research utilizing the entire HECI, data collected with the mini-tests shows that the inquiry-based activities produced significant immediate learning gains for all concept areas as measured by Δ_i. More relevant to this specific study, the data show that the durability of that learning varied significantly by concept area. While all areas showed an improvement from the pre- to the retention test (Δ_f), there were drops in understanding from the post to the retention tests for four of the areas with only Radiation showing no significant decrease in student performance. There was a significant drop with a small effect size for Temperature vs. Perceptions of Hot/Cold and for Amount questions, a significant drop with a medium effect size for Temperature vs. Energy and a significant drop with a moderate to large effect size for the Rate questions.

4.2 Detailed primary results

When the Radiation questions were examined collectively as a mini-test, a paired samples *t*-test showed that there was a significant difference between the pre- to post-test scores with a very large effect size; $t(235) = -16.15$, $p < 0.01$, $d = 1.05$. There was no significant difference in the mean scores between the post- and retention tests. However, there was a significant difference between the pre- and the retention test with a large effect size; $t(208) = -13.81$, $p < 0.01$, $d = 0.96$.

Collective examination of Amount questions using a paired samples *t*-test showed a significant increase from pre- to near post-test with a very large effect size; $t(295) = -21.71$, $p < 0.01$, $d = 1.26$. A dependent *t*-test revealed that there was a significant decrease from the near to the far-post-test with a small effect size; $t(252) = 4.31$, $p < 0.01$, $d = 0.27$. Finally, there was a significant difference (increase) from pre- to retention test with a large effect size; $t(311) = -18.73$, $p < 0.01$, $d = 1.06$.

Similarly, the Rate questions collectively show a significant increase from pre- to near post-test with a

Table 4. Summary of *t*-test Results by Concept Area

Concept Area	Δ_i—Pre-to-Post Test Differences	Post-to-Retention-Test Differences	Δ_f Pre-to-Retention-Test Differences
Radiation	Significant increase with very large effect size.	No significant difference.	Significant increase with large effect size.
Amount	Significant increase with very large effect size.	Significant decrease with a small effect size.	Significant increase with large effect size.
Rate	Significant increase with large effect size.	Significant decrease with moderate to large effect size	Significant increase with moderate effect size.
Temperature vs. Energy	Significant increase with moderately large effect size.	Significant decrease with small effect size.	Significant increase with moderate effect size.
Temperature vs. Perceptions of Hot/Cold	Significant increase with moderately large effect size.	Significant decrease with small effect size.	Significant increase with moderate effect size.

large effect size; $t(295) = -19.47$, $p < .01$, d = 1.13. There was a significant decrease from the near to the far post-test with a moderate to large effect size; $t(252) = 9.65$, $p < 0.01$, $d = 0.61$. If we look at the gains from pre- to far post-test, there is a significant increase with a moderate/medium effect size; $t(311) = -8.59$, $p < 0.01$, $d = 0.49$.

When the Temperature versus Energy questions were collectively examined, a dependent *t*-test revealed a significant increase between the mean scores of the pre- and post-test with a moderately large effect size; $t\ (214) = -11.32$, $p < 0.01$, $d = 0.77$. A paired-samples *t*-test showed that there was a significant decrease in the mean score from the post to the retention test with a small effect size; $t(214) = 3.33$, $p < 0.01$, $d = 0.23$. Lastly, there was a significant increase from pre- to retention test with a medium effect size; $t(214) = -9.32$, $p < 0.01$, $d = 0.64$.

Examining the Temperature vs. Perceptions of hot/cold collectively, a dependent *t*-test found a significant increase in the mean scores from pre- to post-test with a moderately large effect size; $t\ (326) = -14.09$, $p < 0.01$, $d = 0.78$. A paired samples *t*-test showed a significant decrease in the mean score from the post to retention test with a small effect size; $t\ (326) = 2.44$, $p < 0.05$, $d = 0.14$. Finally, there was a significant increase from pre- to retention test with a medium effect size; $t(326) = -12.09$, $p < 0.01$, $d = 0.67$.

As in previous research utilizing the entire HECI, data collected with the mini-tests shows that the inquiry-based activities produced significant immediate learning gains as measured by the difference in pre- and post-test results. More relevant to this specific study, the data show that the durability of that learning varied significantly by concept area. There was no significant decrease in student performance for Radiation, a significant drop with a small effect size for Temperature vs. Perceptions and for Amount questions, a significant drop with a medium effect size for Temperature vs. Energy and a significant drop with a moderate to large effect size for the Rate questions.

4.3 The effect of question difficulty on student learning and retention of learning

We also explored whether learning and retention were influenced by the initial difficulty of the questions. Table 5 examines the performance of students on both difficult questions (measured by their performance on the pre-test) compared to easy questions. Easy questions were operationally defined as those questions where over 60% had the correct answer on the pre-test. There were four questions on the pre-test that met this criterion. Difficult questions were operationally defined as those questions where less than 25% of the participants had the correct answer on the pre-test. Paired samples *t*-tests found that all of the differences between the difficult and easy questions at each assessment time were significant. Mean pre-test scores for the easy questions were significantly higher, with a very large effect size, than the mean pre-test scores for the hard questions; $t(252) = 23.29$, $p < 0.01$, $d = 1.46$. At the post-test, after instruction, the mean scores for the easy questions were significantly higher than the mean scores for the difficult questions with a large effect size; $t(263) = 13.70$, $p < 0.01$, $d = 0.84$. The same pattern was seen with the retention test with the mean scores for the easy questions significantly higher than for the difficult questions with a very large effect size; $t(263) = 20.92$, $p < 0.01$, $d = 1.29$. Looking at this table, we see a great deal of improvement for the more difficult questions after instruction but a bigger drop on the retention test when compared to the easy questions.

A paired samples *t*-test was used to examine the differences between students' performance on easy and difficult questions at each assessment time. For easy questions, there was a significant increase from pre- to post-test with a large effect size; $t(326) = -13.78$, $p < 0.01$, $d = 0.76$. While students' mean scores on easy questions decreased from the post to retention tests, the difference was not significant. For hard questions, there was a significant increase from pre- to post-test with a very large effect size; $t(252) = -21.26$, $p < 0.01$, $d = 1.34$. However, there was a significant decrease with a moderate effect size from post to retention tests for difficult questions; $t(263) = 8.68$, $p < 0.01$, $d = 0.53$.

4.4 The effect of "transfer" on students' initial and long-term learning

Recall that questions were defined as requiring either near or far transfer. Questions were characterized as near transfer if they matched the specific situation employed in the activity. Similarly, the

Table 5. Student Performance on Questions by Difficulty

	Mean % Correct on Pre-test	**Mean % Correct on Post-test**	**Mean % Correct on Retention-test**
Difficult Questions	22.3	75.2	58.1
Easy Questions	44.2	73.5	71.9

questions were labelled "far transfer" if the concept was employed in a different context on the assessment than the context students experienced during the activity. Table 6 shows student performance on questions requiring near transfer to those requiring far-transfer. Again, there are significant short term and long-term improvements on student performance after the targeted activities. There are, however, differences in the patterns observed between the post-test and retention-test for each category of questions. One can see that the inquiry-based activities initially produced larger improvements on questions requiring near transfer than for those requiring far transfer. However, the drop off in student scores between the immediate post-test and the end of semester exam was significantly less for questions requiring far transfer. That is, while student scores on questions requiring far transfer improved less initially, those gains were more stable over time.

A paired samples t-test was used to look at the difference between students' performance on the pre-test when compared with the post-test and the post-test when compared with the retention-test for both near and far transfer questions. For near transfer questions, there was a significant increase from pre- to post-tests with a very large effect size [$t(252) = -25.37$, $p < 0.01$, $d = 1.60$] but, a significant decrease from post to retention tests with a medium to large effect size [$t(263) = 9.79$, $p < 0.01$, $d = 0.60$]. For far transfer questions, there was a significant increase from pre- to post-test with a large effect size [$t(326) = -16.84$, $p < 0.01$, $d = 0.93$ and a significant decrease from post to retention tests with a small-to-low medium effect size [$t(326) = 4.34$, $p < 0.01$, $d = 0.24$].

A one-way analysis of variance (ANOVA) with pre-test as the dependent variable and question type (near vs. far transfer) as the independent variable confirmed that there was a significant difference between the two question types on the pre-test with a medium effect size; $F(1, 578) = 64.39$, $p < 0.01$, partial $\eta^2 = 0.10$. The mean score on the far transfer questions was significantly higher than the mean score on the near transfer questions. Given this finding, a one-way analysis of covariance (ANCOVA) was done to determine whether there were significant differences between the two kinds of questions (near/far transfer) on the post- and retention tests when pre-test scores were controlled. There was a significant difference with a large effect size for question type on the mean post-test scores after controlling for pre-test scores: $F(1, 577) = 158.24$, $p < 0.01$, partial $\eta^2 = 0.16$. The mean score for the near transfer questions was significantly higher than the mean score for the far transfer questions on the post-test when pre-test scores were partialed out. There was also a significant difference between question type (near/far transfer) with a small effect size on the mean retention test scores after controlling for pre-test scores: $F(1, 577) = 37.08$, $p < 0.01$, $\eta^2 = 0.06$. The mean score on the near transfer questions was significantly higher than the mean score for the far transfer questions on the retention test when the pre-test scores were controlled.

4.5 The effect of students' initial preconception when learning doesn't "stick"

Table 7 shows more detailed information on what is happening over time with regard to retention of learning. In particular, we are interested in understanding what is happening when learning doesn't "stick". We've defined "failing to stick" as situations students went from being incorrect on the pre-test to correct on the post-test (suggesting that they learned the concept correctly) and then reverting to giving a wrong answer on the retention test (suggesting that they did not remember what they had apparently learned). The percentage of times this happened for each question is shown in the second column in Table 7. We also tracked the percentage of times that students who chose a wrong answer reverted to their initial answer given on the pre-test. This is shown in column 3. Column 4 simply shows the odds of this happening by random chance, given the number of possible responses provided on each question.

Two patterns emerge from this data. First, this failure of learning to persist happened most frequently for questions focused on understanding factors promoting the rate of energy transfer. Clearly, the inquiry-based activities were least effective in promoting more permanent learning gains for this concept area. Secondly, we looked at how often students who failed to retain what they learned reverted to their original preconception along with how often one might expect them to revert to this response due to random chance. The data show that revisions to the original preconception occurred

Table 6. Mean Student Performance on Questions Requiring Near vs. Far Transfer

	Mean % Correct On Pre-test	**Mean % Correct On Post-test**	**Mean % Correct On Retention-test**
Near-Transfer	26.26	69.78	58.86
Far-Transfer	40.54	59.16	55.66

Table 7. Learning that doesn't "stick" and reversions to original preconceptions (percentage)

Concept Area	Wrong-Right-Wrong	Reversion to Original Preconception	Random Chance of Reversion to Preconception
Radiation (mean)	8.4	68.4	25.8
Rate vs. Amount (aggregate mean)	24.4	57.5	29.6
Rate (mean)	37.5	63.2	30.5
Amount (mean)	16.6	54.1	30.0
T vs. Energy (mean)	18.1	46.3	24.5
T vs. Perception (mean)	10	51.6	27.1

about twice as often as could be attributed to random chance.

5. Discussion

The limited literature on how well students retain what they learn through active-engagement or inquiry-based instruction suggests that, on average, these methods are effective for promoting long-term changes in how students think. On balance, our results suggest similar findings. The inquiry-based activities examined in this study promoted significant learning gains that lasted several weeks after the activity. A more nuanced examination of the data, however, allows for a richer examination of student learning.

One observation is that both initial learning and retention of what is learned varies by concept area. Given the study's focus on retention, it is interesting to note that student performance on the Radiation questions shows very little drop off between the post-test and the retention-test. In fact, two of the Radiation questions actually show an increase in student performance between the time of the intervention and the end of the semester. One possible explanation for this is that Radiation as a topic is generally covered towards the end of the semester whereas instruction related to the Rate vs. Amount questions would typically come earlier in the semester. There are two relevant consequences for our study. First, there is likely to be a smaller period of time for students to "forget" what they learned in the Radiation activity, which probably helps to explain this observation. In addition, there may be some direct instruction in radiation still occurring between the time of the post-test and the end of semester retention-test, which might explain the improvement in student performance on some radiation questions. While we do not know with certainty that the timing of instruction influenced these findings, it is one possible explanation.

Timing of when topics are introduced within the heat transfer course, even if it is a significant factor in some cases, cannot explain all of the observed trends. For example, the difference in student performance on questions focused on energy balances (Amount) vs. Rate questions is unlikely to be due to when topics are covered in the course since both are typically covered together, early in the semester. However, there may be timing issues at work on a different scale. For example, results show that while the inquiry-based activities produced greater immediate improvements on Rate questions than on Amount questions, the long-term impact was significantly less. This may have to do with when concepts are introduced in the curriculum. The concept of "rates" is likely to be new to students when taking a heat transfer course. In many curricula, heat transfer is the first course which discusses and models "rates" of energy transfer. On the other hand, energy balances are frequently covered in earlier courses, so that students typically begin the course with a stronger foundation in this area. That is suggested by the higher score on the pre-test for "amount" questions. Therefore, the timing of when the concept is introduced in the curricula might influence how well students retain what is learned through the activity.

Regardless of timing issues, data collected in this study suggests that retention is influenced by the initial difficulty of the concept. There is a larger drop off in student performance over time on questions that are initially more difficult than on questions students initially found easier. This suggests that difficult concepts may be more generally resistant to long term change and require a greater degree of intervention to affect more lasting learning.

Another trend worth noting is that between questions requiring near- vs. far-transfer of what is learned in the activity. We see larger effect sizes for improvement on questions requiring near-transfer, although we want to be cautious using large effect sizes as a key measure of change, given that sample size can impact effect sizes [31]. This result for near-transfer is not surprising, since one would expect students to find it easier to recognize the correct answer in situations where that question closely mimics what students observed in the activity. Perhaps more surprising is the improved retention of student performance on questions requiring far-transfer. One can argue that questions requiring far transfer are in fact testing "deep learning" and that deep learning would hold up better over time. This is consistent with Bacon and Stewart's [32] interpretation of some of the data in that study.

Finally, some previous research [7, 33] has suggested that traditional instruction does little to change students' initial preconceptions. The finding that strongly held preconceptions may continue to exert some influence on students' long-term thinking is also suggested by some of the data presented here (Table 7). Specifically, we have found that when learning doesn't "stick", students are significantly more likely to revert to their original preconception than to randomly change their views. The phenomenon of students reverting to their original preconceptions has also been observed in other studies [34]. In our study, we found that learning "stuck" least well for the Rate questions or questions that required students to understand factors which influenced the rate at which energy would transfer in specific scenarios. This may be due to something about the concept itself or the lack of background understanding students bring into the course in this area (as mentioned previously). We also note that Rate questions were among the most difficult on the instrument and, as noted previously, we have observed that retention of learning seems to be negatively influenced by the initial difficulty of the concept, at least within our dataset.

There are several implications of these findings for both educational researchers and classroom instructors. Instructors looking to adopt techniques such as inquiry-based activities need to know if the positive results found for the immediate impact of the activity produce long term gains or lead to students' ability to apply what they have learned to new concepts. In general, this study provides some confidence about the benefits of the instruction for both outcomes. While there is a general drop-off in performance over time, the drop-off is (on average) slight. And while short term learning on questions related to near-transfer show the greatest improvement, there are statistically significant findings for both long-term retention of learning and for learning as assessed by questions requiring far-transfer. Learning gains of both types have been found to be greater using inquiry-based instruction than those found using traditional instruction, as reported in earlier studies by the authors [2, 35–37].

The results are equally relevant for educational researchers. The results show that several factors moderate the measured short-term and long-term effectiveness of inquiry-based instruction. The impact of the intervention is influenced by several variables, including the concept area itself, initial difficulty, the degree of transfer required and possibly issues related to when a concept is introduced either in the course or in the overall curriculum. This suggests that investigators seeking to understand the relative effectiveness of an intervention will need to keep all these factors in mind when drawing comparisons.

5.1 Limitations

There are a couple of limitations in the current study. One is the use of an intact, sample of convenience rather than a random sample. Even though care was taken to obtain participants from a number of institutions and locations, samples of convenience still remain unrepresentative of the population [29]. Second, estimates of internal consistency reliability of the radiation questions were low. Future research should work to increase the internal reliability of this assessment, possibly by adding more questions from the HECI.

6. Conclusions

The study examined the impact of inquiry-based instruction on students' learning and retention of that learning for several concepts in heat transfer. Engineering students' retention of concepts taught using inquiry-based activities was generally high (78% retention on average) from the post-test immediately after the activity to the retention-test at the end of the semester. This high retention is consistent with a limited number of similar studies found in physics education. However, the retention level varied significantly by concept area. Specifically, while retention rates from post-test to retention test were high for some areas such as Amount (86% retention) and Radiation (98% retention), they were significantly lower (45% retention) for questions related to factors that influenced that rate of heat transfer. Retention rates for Radiation might be attributable to the fact that radiation is generally taught later in the semester. Higher retention rates on Amount or energy balance questions might be due to the fact that students begin the heat transfer course with a pre-existing foundation in material balances, making these questions easier for them.

While inquiry-based instruction produced larger initial gains on questions requiring near-transfer than those requiring far-transfer, the *retention* of what was learned was actually higher for far-transfer questions. This might indicate that "far transfer" questions measure deep learning, which is retained longer.

Students seem to retain more of what they learned based on the initial difficulty of the question. Specifically, the retention of learning was better for easier questions while the retention of what was learned dropped more significantly for questions students initially found more difficult. When learning is not retained, students in this study reverted to their initial misconceptions more frequently than can be accounted by random chance. This suggests that their initial preconception continues to exert some influence on their thinking, even after it appears that the misconception has been repaired through inquiry-based instruction as measured by short term assessments.

Acknowledgements—The authors would like to acknowledge funding from NSF DUE#0717536 and DUE#0442234.

References

1. S. Suping, Conceptual Change Among Students in Science, *ERIC Digest ED482723*, 2003.
2. M. Prince, M. Vigeant and K. Nottis, Using inquiry-based activities to repair student misconceptions related to heat, energy, and temperature, *Frontiers in Education*, Seattle, WA, 2012.
3. S. Avsec and S. Kocijancic, Effectiveness of Inquiry-Based Learning: How Do Middle School Students Learn to Maximise the Efficacy of a Water Turbine, *International Journal of Engineering Education*, **30**(6), 2014, pp. 1436–1449.
4. M. Atilhan, F. Eljack, H. Alfadala, J. Froyd, M. El-Halwagi and V. Mahalec, Inquiry Guided Learning in a Chemical Engineering Core Curriculum: General Instructional Approach and Specific Application to the Fluid Mechanics Case, *International Journal of Engineering Education*, **30**(6), 2014, pp. 1450–1460.
5. F.-K. Chiang, S. Diao, H. Ma and Y. Wang, Effects of Hands-on Inquiry-Based Learning Using Lego1 Materials on the Learning of Eighth-Grade Physics Students, *International Journal of Engineering Education*, **33**(3), 2017, pp. 1098–1103.
6. D. F. Halpern and M. D. Hakel, Applying the Science of Learning to University Teaching and Beyond, *New directions for teaching and learning*, **89**, 2002.
7. M. Prince, M. Vigeant and K. Nottis, Assessing the Prevalence and Persistence of Engineering Students' Misconceptions in Heat Transfer, *Journal of Engineering Education*, **101**(3), 2012, pp. 412–438.
8. M. Prince and R. Felder, Inductive Teaching and Learning Methods: Definitions, Comparisons, and Research Bases, *Journal of Engineering Education*, **95**(2), 2006, pp. 123–138.
9. M. Prince and R. Felder, The Many Faces of Inductive Teaching and Learning, *Journal of College Science Teaching*, **36**(5), 2007, 14.
10. J. Strobel and A. Van Barneveld, When is PBL More Effective? A Meta-Synthesis of Meta-Analyses Comparing Pbl to Conventional Classrooms, *Interdisciplinary Journal of Problem-based Learning*, **3**(1), 2009, 4.
11. F. Dochy, M. Segers, P. Van Den Bossche and D. Gijbels, Effects of Problem-Based Learning: A Meta-Analysis, *Learning and Instruction*, **13**(5), 2003, pp. 533–568.
12. P. Georghiades, Beyond Conceptual Change Learning in Science Education: Focusing on Transfer, Durability and Metacognition, *Educational Research*, **42**(2), 2000, pp. 119–139.
13. J. Bernhard, Does active engagement curricula give long-lived conceptual understanding? In Eds R. P. and S. Surinach, Physics Teacher Education Beyond 2000, Paris: Elsevier, 2001, pp. 749–752.
14. L. Deslauriers, E. Schelew and C. Wieman, Improved Learning in a Large-Enrollment Physics Class, *Science*, **332**(13 May), 2011, pp. 862–864.
15. G. E. Francis, J. P. Adams and E. J. Noonan, Do They Stay Fixed?, *Physics Teacher*, **36**, 1988, p. 488.
16. S. J. Pollock, Longitudinal Study of Student Conceptual Understanding in Electricity and Magnetism, *Physical Review Special Topics—Physics Education Research*, **5**, 2009, p. 110.
17. J. P. Byrnes, *Cognitive development and learning in instructional contexts, 2nd ed.*, Allyn & Bacon, Needham Heights, MA, 2001.
18. S. Vosniadou, C. Ioannides, A. Dimitrakopoulou and E. Papademetriou, Designing Learning Environments to Promote Conceptual Change in Science, *Learning and Instruction*, **11**(4–5), 2001, pp. 381–419.
19. P. Laws, D. Sokoloff and R. Thornton, Promoting Active Learning Using the Results of Physics Education Research, *UniServe Science News*, **13**, 1999.
20. S. Hausfather, It's Time for a Conceptual Change, *Science and Children*, Nov/Dec, 1992, pp. 22–23.
21. M. F. Thomas, I. Malaquas, M. Valente and M. Antunes, An Attempt to Overcome Alternative Conceptions Related to Heat and Temperature., *Physics Education*, **30**(1), 1995, pp. 19–26.
22. B. Thacker, K. Trefz and S. Lea, Comparing Problem Solving Performance of Physics Students in Inquiry-Based and Traditional Introductory Physics Courses, *American Journal of Physics*, **62**(7), 1994, p. 627.
23. M. Prince, M. Vigeant and K. Nottis, Preliminary Study on the Effectiveness of Inquiry-Based Activities for Addressing Misconceptions of Undergraduate Engineering Students, *Education for Chemical Engineers*, **4**(2), 2009, pp. 29–41.
24. M. Vigeant, M. Prince and K. Nottis, Engineering undergraduates' conceptual understanding of thermodynamics: Assessment and change after normal instruction, *Hawaii International Conference on Education*, Honolulu, HI, 2011.
25. M. Vigeant, M. Prince and K. Nottis, Teaching and Learning of Energy in K-12 Education: Repairing Engineering Students' Misconceptions About Energy and Thermodynamics, 2014.
26. M. Prince, M. Vigeant, K. Nottis, G. Nottis and M. Koretsky, Impact of inquiry-based learning activities on the retention and transfer of conceptual learning in heat transfer, *Research in engineering Education Symposium*, Dublin, Ireland, 2015.
27. S. Barnett and M. Ceci, When and Where Do We Apply What We Learn?: A Taxonomy for Far Transfer?, *Psychological-Bulletin*, **128**(4), 2002, pp. 612–637.
28. C. A. Mertler and R. A. Vannatta, *Advanced and multivariate statistical methods*, Pyrczak Publishing, Glendale, CA, 2005.
29. J. R. Fraenkel, N. E. Wallen and H. H. Hyun, *How to design and evaluate research in education*, McGraw Hill, New York, 2012.
30. B. C. Cronk, *How to use SPSS: A step-by-step guide to analysis and interpretation*, Routledge, New York, 2017.
31. A. C. K. Cheung, and R. E. Slavin, How Methodological Features Affect Effect Sizes in Education, *Educational Researcher*, **45**(5), 2016.
32. D. R. Bacon and K. A. Stewart, How Fast Do Students Forget What They Learn in Consumer Behavior? a Longitudinal Study, *Journal of Marketing Education*, **28**, 2006, p. 181.
33. R. Hake, Lessons From the Physics Education Reform Effort, *Conservation Ecology*, **5**(2), 2002, p. 28.
34. A. Learner, Minds of Our Own: Can We Believe Our Eyes?, *Math and Science Collection*, **1**, 1997.
35. M. Vigeant, M. Prince, K. Nottis, M. Koretsky, E. Bent, R. Cincotta and K. MacDougall, Why not just run this as a demo? Differences in students' conceptual understanding after experiments or demonstrations, *American Association for Engineering Education*, Columbus, OH, 2017.
36. K. Nottis, M. Prince and M. Vigeant, Undergraduate Engineering Students' Understanding of Heat, Temperature, and Energy: An Examination By Gender and Major, *US-China Education Review A*, **7**(4), 2017.
37. M. Prince, M. Vigeant and K. Nottis, The Impact of Inquiry-Based Learning Activities on the Retention and Transfer of Conceptual Learning in Heat Transfer, *Chemical Engineering Education*, **49**(1), 2016.

Appendix: Sample Concept Questions from each Mini-Test: (Radiation, Rate vs. Amount, T vs. Energy and T vs. Perceptions)

Sample Radiation questions

Steam at 100°C is fed into 2 metal pipes, otherwise identical except that one is painted with a flat black paint while the other pipe has a shiny copper finish. You may assume that radiation is a significant fraction of the total heat flux from the pipes to the surrounding room and that the rate of heat loss from the pipes determines how fast the steam condenses.

Question 1: In which pipe will steam condense at a faster rate?

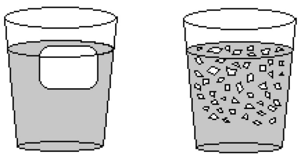
a. Steam will condense more quickly in the shiny copper pipe
b. Steam will condense more quickly in the painted black pipe
c. Steam will condense in both pipes at equal rates

Question 2: Because. . .

a. The paint acts as an insulator
b. The black paint absorbs and holds in the heat better
c. The polished surface will reflect heat better
d. Black paint has a higher emissivity
e. The exterior color does not matter

Sample Rate vs. Amount Questions: Q1 is an "amount" question, Q2 is a "rate" question

You would like to cool a beverage in an insulated cup either by adding 1 large ice cube or the same mass of finely chipped ice.



Q1. Which option will cool the beverage to a lower temperature?

a. The large ice cubes will cool the beverage to a lower temperature.
b. The finely chipped ice will cool the beverage to a lower temperature.
c. Either option will ultimately cool the beverage to the same temperature.
d. There is not enough information provided.

Q2. Which option will cool the beverage more quickly?

a. The large ice cubes will cool the beverage more quickly.
b. The finely chipped ice will cool the beverage more quickly.
c. Either option will cool the beverage at the same rate.
d. There is not enough information provided.

Sample Temperature vs. Energy Question

Pouring either 10 ml of boiling water at 100°C or 100 ml of ice cold water at 0°C into a container of liquid nitrogen at its boiling point of –200°C will cause some of the liquid nitrogen to evaporate. Assume that the container is perfectly insulated and that there is initially a very large volume of liquid nitrogen so that *some* remains at the end of the process.

Which situation will cause *more* liquid nitrogen to evaporate?

a. Adding 100 ml of ice cold water because it has a higher surface area
b. Adding 100 ml of ice cold water because it has a higher mass
c. Adding 10 ml of boiling water because it has a higher energy level
d. Adding 10 ml of boiling water because of the greater temperature difference with the liquid nitrogen
e. Either will cause the same amount of liquid nitrogen to evaporate

Sample Temperature vs. Perceptions of Hot and Cold Question

An engineering student walking barefoot (without shoes or socks) from a tile floor onto a carpeted floor notices that the tile feels cooler than the carpet. Which of the following explanations seems like the most plausible way to explain this observation?

a. The carpet has a slightly higher temperature because air trapped in the carpet retains energy from the room better.
b. The rate of heat transfer into the room by convection (air movement) is different for tile and carpet surfaces.
c. The carpet has a slightly higher temperature because air trapped in the carpet slows down the rate of energy transfer through the carpet into the floor
d. The tile conducts energy better than the carpet, so energy moves away from the student's foot faster on tile than carpet.

Michael Prince is a professor of chemical engineering at Bucknell University in Lewisburg, Pennsylvania. He is also co-Director of the National Effective Teaching Institute and ''How to Engineer Engineering Education'', both well-established workshops that attract a national audience of engineering faculty each year. Dr. Prince is a Fellow of the American Society for Engineering Education. His current research examines how to assess and repair student misconceptions and how to increase the diffusion of educational theory into educational practice.

Margot Vigeant is a professor of chemical engineering and Rooke Professor of engineering at Bucknell University. She is interested in assessing the effectiveness of active learning methods, including inquiry-based activities, making, and PBL, and promoting their wider adoption.

Katharyn Nottis is professor emeritus of education at Bucknell University. She has worked extensively with colleagues in engineering, physics, and geosciences on development and assessment of educational innovations. After her recent retirement, she plans to continue her scholarly work in this area on a consulting basis while taking more time to enjoy the company of friends, family, and sunshine.

Gary Nottis is a graduate, with honours, from the Department of Geology, Bucknell University. He has worked on earthquake hazard assessments with the New York State Geological Survey, as a Teaching Assistant for Quantitative Research Methods in the Education Department of Bucknell University, and as a consultant to the Chemical Engineering Department of Bucknell on statistical analyses. He is currently involved in historical earthquake documentation and related statistical analyses.