Higher Education Quality Council of Ontario

An agency of the Government of Ontario

Evaluating the Effectiveness of Modified Peer Instruction in Large Introductory Physics Classes

Tetyana Antimirova, Aleksandra Kulesza, Andrea Noack, Mary Stewart, Ryerson University

Published by

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada, M5E 1E5

Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Cite this publication in the following format:

Antimirova, T., Kulesza, A., Noack, A., & Stewart, M.* (2015). *Evaluating the Effectiveness of Modified Peer Instruction in Large Introductory Physics Classes.* Toronto: Higher Education Quality Council of Ontario.

* The use of alphabetical order reflects equal authorship.

Table of Contents

List of Figures

List of Tables

Executive Summary

This report presents the results of research into the use of collaborative, multiple-choice format question-writing activities as a supplement to standard peer instruction (PI) methods in a large introductory physics course.

The standard PI method includes posing questions for student reflection and challenging students to identify gaps in their own understanding. A typical PI class session consists of brief lecture segments interposed with short quizzes consisting of conceptual questions, mostly in a multiple-choice format. In large classes, students’ responses are usually collected with the aid of a personal response system. After a first poll of quiz results, students spend a few minutes discussing their choices with two to four peers, attempting to agree on the correct answer. Students in classes that use PI show significant gains in conceptual understanding, as measured by standardized tests. Moreover, the gain in conceptual understanding that results from PI translates into better problem solving skills than for students in traditional lecture-based classes. However, PI pedagogy relies heavily on multiple-choice format questions (MCFQs). Therefore, despite the effectiveness of the PI method, the advantages and disadvantages of using MCFQs for both teaching and evaluation have been a topic of ongoing debate. Although many of the critiques of MCFQs can be surmounted, one fundamental limitation of MCFQs is hard to dismiss. Even proponents of PI acknowledge that using MCFQs means that students do not learn to formulate or articulate their own ideas, and instead select from among the provided responses.

In response to this limitation, we developed and tested a modified form of PI that includes a supplementary activity prompting students to formulate and articulate their own ideas clearly. The new activity has students work collaboratively in small groups to write MCFQs similar to those used for PI. These activities took place both in class and online. The objective of this study was to evaluate whether the introduction of these MCFQ-writing activities enhanced students’ understanding of fundamental concepts in physics, students’ attitudes towards physics and the degree of student engagement, as compared to the standard PI pedagogy. This study was implemented in PCS120 (Physics I), a required introductory physics course for undergraduate science program students at Ryerson University in Toronto, Ontario. PI supported by a personal response system (i.e., clickers) has been used in the course since 2006 and has had a documented positive effect on students’ learning outcomes.

The effect of incorporating MCFQ-writing activities into the course was assessed using both quantitative and qualitative approaches. The quantitative measures included an assessment of the normalized gain in students' conceptual knowledge, as measured by a pre- and post-test administration of the Force Concept Inventory (FCI). In addition, change in student attitudes was measured by a pre- and post-test administration of the Physics version of the Colorado Learning Attitudes about Science Survey (CLASS). A socio-demographic survey asking students' about their personal and educational background and preferences was also administered. After completing the course, some students participated in semi-structured, in-depth interviews probing their perception of the new activity and their learning in the course overall.

The results show that the introduction of collaborative MCFQ-writing activities had inconsistent effects on students' learning. In the 2012 cohort, student characteristics appeared to have little effect on normalized learning gains, whereas participating in the MCFQ-writing activities was associated with a significant improvement in conceptual learning. This effect was not replicated in the 2013 cohort. In contrast, in the 2013 cohort, students' personal characteristics were strongly associated with normalized learning gains, but participating in the MCFQ-writing activities appeared to have little effect on conceptual learning. When the results of the two cohorts are pooled, gender, English-language proficiency and attendance were each associated with larger normalized gains in conceptual learning than participating in the MCFQ-writing activities. With the exception of attendance, these attributes are typically not influenced by the teaching practices of introductory physics instructors. These results suggest that the use of MCFQ-writing activities in large, introductory physics courses has some beneficial effect, but that this effect is not necessarily consistent across all groups of students.

CLASS is used to distinguish between novice-like and expert-like attitudes towards science overall and in eight empirically generated categories. Introductory physics courses commonly lead students to adopt less-expert attitudes. In this study, a positive shift in students' attitudes was detected in the areas of problem solving sophistication, conceptual understanding and applied conceptual understanding. The positive attitude shift in these three areas related to conceptual understanding reflects the success of PI pedagogy in promoting students' engagement and conceptual learning in physics and cannot be attributed specifically to the MCFQ-writing activities. A negative shift was however apparent in two areas: personal interest in physics and sense-making effort (feeling like the effort needed to make sense of the material is worthwhile). Compared to students who only participated in standard PI activities, the use of MCFQ-writing activities did not appear to have an effect on student learning. In other words, adding MCFQ-writing activities on the top of the standard PI activities does not seem to further improve students' attitudes toward physics.

Semi-structured interviews were used to probe students' perception of the pedagogy and technology tools used in the course, and prompted them to reflect on their own learning. Interviewees often did not recognize the MCFQ-writing activities as a unique component of the course and many did not specifically remember participating in them. This omission is notable, as about 60% of the class received credit for participating in the online MCFQ-writing activities and an even larger proportion participated in the in-class activities. Students who did recall the MCFQ-writing activities provided positive feedback, recognizing them as an opportunity to reflect on material, monitor their own understanding and try "getting into the mindset of the professor."

Given the results of the study, the researchers recommend that instructors consider the cost, time and resources needed to implement these small-group, collaborative MCFQ-writing activities. Compared to standard PI techniques, this new activity requires a significant amount of preparation, as well as time-consuming monitoring of student online activities, without demonstrating consistent effects on student learning. This intervention may have a different effect in a course with fewer student learning supports already in place. We encourage instructors in large classes to consider alternative interventions to increase student learning, while also remaining mindful of the potential for resource overload among students. Although the new MCFQ-writing activities proved to be less effective than expected, interviewees expressed overwhelming support and appreciation for the diverse opportunities for active learning in the course.

1. Introduction

1.1 The Challenges of Large Introductory Physics Courses

As a result of both fiscal and demographic pressures, university class sizes in Ontario have been growing, especially at the first-year level. Common negative effects of large class sizes include reduced student-faculty interactions and individual feedback to students, and a lower overall level of student engagement (Cuseo, 2007; Iaria & Hubball, 2008). On its own, lecture-style instruction promotes a banking method of learning and has the potential to isolate students with different learning needs and diverse educational experiences. The traditional lecture format encourages students to focus on factual materials and superficial details instead of thinking about the concepts, underlying principles and major ideas (McCarthy & Anderson, 2000). In addition, students' previous academic backgrounds and expectations can vary dramatically in large classes. For example, as is the case in the present study, some introductory physics courses in Ontario universities combine students with and without senior-level high school physics experience. Moreover, the diversity within the Ontario student population in relation to immigration experiences, visible minority status, Aboriginal identity, students with disabilities, first-generation students and second-career students means that in order to be effective, instructors must adopt strategies to reach out to a wide range of learners.

Among all the science disciplines, physics continues to be perceived as a particularly challenging subject. Students' attitudes toward and pre-conceptions about physics often negatively affect their learning and pose an additional challenge for instructors (Redish, Saul & Steinberg, 1998; Gray, Adams, Wieman & Perkins, 2008). Students often feel disconnected from the material taught in the physics classrooms and, as a result, gradually lose motivation and interest in the course. This effect is particularly pronounced in large introductory classes, where the abstract nature of the subject matter and the lack of hands-on experience can pose significant barriers to students' learning. This contributes to ongoing problems with student retention. In response to these challenges, many physics instructors have adopted new pedagogical techniques to increase students' engagement with class material in an effort to improve grades and course completion rates.

There is ample evidence from science education research that in instructor-centered, lecture-style classes students learn only a fraction of what they could have learned; students learn much more in classes that employ enquiry and offer a high degree of interactive engagement and social interaction (Hestenes, Wells & Swackhamer, 1992; Hake, 1998; Thornton & Sokoloff, 1998; Savinainen & Scott, 2002; Thornton, Kuhl, Cummings & Marx, 2009). In large introductory physics courses where the instructor used a didactic one-directional lecture approach, learning gains as low as 10-15% have been reported (Hake, 1998). Traditional lectures often reinforce students' erroneous beliefs that the rote memorization of unrelated facts and random examples represent meaningful learning. Activity-based, technology-enhanced learning is one approach with the potential for shifting the classroom focus from professor-centered to student-centered, thus improving student learning outcomes. Dramatic evidence of the effectiveness of interactive engagement compared to the one-directional lecture has recently emerged from a comparative study at the University of British Columbia (Deslauriers, Schelew & Wieman, 2011).

One meaningful measure of successful learning in science is the development of scientific reasoning and the mastery of key scientific concepts, which combine to promote better problem-solving skills and critical data analysis (Hestenes et al., 1992; Thornton & Sokoloff, 1998; Savinainen & Scott, 2002; Thornton et al., 2009). Improving students' attitudes towards science is very important, since it can also have a positive effect on student learning (Perkins, Adams, Finkelstein, Pollock & Wieman, 2004; Pollock, 2004; Perkins, Gratny, Adams, Finkelstein & Weiman, 2005; Adams et al., 2006). In addition to the substantive course content, helping physics learners to acquire more expert-like behaviours and attitudes has long been considered an important goal of physics teaching. However, this goal is particularly hard to achieve in a large lecture-style course. Changing students' attitudes requires engagement strategies that encourage students to be active participants in the classroom and outside.

Since instructors have a wide variety of PI and active learning strategies available to them, it is important to assess which strategies students find the most useful and which lead to the best learning outcomes in various learning environments. Several common student engagement techniques are described below.

1.2 Using Peer Instruction and Personal Response Systems to Improve Student Learning and Engagement

Peer Instruction (PI) is a well-documented strategy for increasing students' classroom engagement (Mazur, 1997). The use of PI in science classes has led to significant positive effects on students' performance in the many institutions that use it (Deslauriers et al., 2011).There are many variations of PI pedagogy: with or without personal response systems, with or without requiring prior reading, and in combination with just-in-time teaching or used independently. Students in classes that use PI show significant gains in conceptual understanding, as measured by standardized tests (Crouch & Mazur, 2001; Fagen, Crouch & Mazur, 2002; Singh, 2005; Smith et al., 2009). Moreover, the gain in conceptual understanding that results from PI translates into better problem-solving skills than for students in traditional classrooms, even though PI activities do not usually explicitly teach students how to solve traditional numerical problems.

Activity-based 'learning by doing' is rooted in the ideas of constructivism, a learning theory that is particularly relevant for science education (Fensham, Gunstone & White, 1994). Constructivist approaches treat learning as a process through which individuals develop new concepts by building upon their prior knowledge and incorporating new information. In this framework, mobilizing students' prior knowledge is absolutely crucial for successful learning to take place. In physics education, however, the ideas that students bring with them to class often include misconceptions, which often persist despite formal instruction (McDermott & Redish, 1999). For example, it is common for particular misconceptions about Newtonian mechanics to persist among graduate students, even though these topics are typically covered in introductory, first-year undergraduate physics courses (McDermott, 1990; McDermott & Redish, 1999).

The success of PI relies heavily on using meaningful conceptual questions that allow peers and instructors to elicit, confront and resolve students' misconceptions. Often, these conceptual questions take on a multiple-choice format, with each incorrect answer option representing a specific, identifiable misconception or mistake. As an additional benefit, frequently practicing multiple-choice format questions (MCFQs) gives students increased confidence in completing MCFQ-format formal evaluations. In large classes, students' answers to the MCFQs used for PI are often collected in real-time using electronic response systems (clickers). Although clickers were initially developed to combat the challenges of large classes, they have become effective technology tools used to support PI pedagogy in classes of all sizes. Popular clicker technology consists of a radio-frequency receiver for the instructor, individual student clickers (voting devices) and supporting software for the analysis of student responses. A more recent trend is web-based clicker-like voting systems, such as, for example, Web-Clicker iClicker (https://webclicker.iclicker.com/), or TOP HAT (https://tophat.com/), which do not require a separate physical device for either the instructor or students. In either form, personal response systems can be exceptionally useful for providing course instructors with an immediate snapshot of the prevailing ideas and misconceptions among students in the class so that they can adjust their teaching accordingly.

The availability of good-quality MCFQs that students can answer in real-time using personal response systems is central to many forms of PI. Fortunately, science educators have produced extensive research-based materials to help instructors implement PI and other interactive teaching methods. For example, many science textbook publishers now provide extensive multiple-choice question banks with almost every text, though the quality of the questions varies substantially. Since physicists were among the first to start using personal response systems, many physics textbooks have particularly good question banks to support PI pedagogy. In addition, some instructors create and share their own questions with colleagues; there are a growing number of online databases dedicated to allowing instructors to share and compile effective MCFQs for PI.

Despite the success of PI that uses MCFQs in improving learning outcomes in science classes, the advantages and disadvantages of using MCFQs for both teaching and evaluation remains a topic of ongoing debate. One serious concern raised by critics of PI using MCFQs is that this pedagogy prompts students to select from among pre-established answers and thus students do not learn to formulate their own statements and express their own ideas (McDonald, 2001; Brown, Race & Smith, 2004). Some instructors argue that using MCFQs for evaluation allows students to answer questions by guessing, without necessarily figuring out the solutions (Cunningham, 2005). Depending on the number of options for each question (normally four or five), guessing will result in 20 to 25% false positives. Some students also perceive MCFQs to be unfair compared to regular full-solution questions because they do not provide partial marks. Students may believe that the main reason why the instructors use MCFQs is to save time on marking. Another widespread misconception among students is that the marks for such MCFQ tests are lower than for long-answer tests (which is not true, provided that the tests are properly designed). Our motivation for implementing MCFQ-writing activities in addition to PI was to counter some of these concerns by providing students with opportunities to formulate their own ideas and showing students how MCFQs are carefully structured to identify common misconceptions. Previous research on the use of student-generated MCFQs is scarce. Bottomley and Denny (2011) reported some success using this strategy with biomedical sciences students, using the PeerWise software for students' online collaboration. These researchers reported that students were eager to participate and produced a large repository of relevant, good-quality MCFQs.

1.3 Modified Peer Instruction in Large Introductory Physics Classes at Ryerson University

Ryerson University is a mid-to-large-size institution with approximately 38,000 undergraduate students, located in downtown Toronto. It attracts an ethnically and linguistically diverse student body. Like other similar-sized universities, Ryerson University offers several large introductory science classes, many of which have more than 200 students per lecture section and more than 400 students per course. Students enrolled in these courses come from a variety of different programs (specializations) with varying admission requirements and thus also have diverse backgrounds in terms of their previous science and math education. In order to capture the diversity of student histories and understand their relation to learning outcomes, the researchers administered a socio-demographic survey at the same time as the other measures used in this study.

Since 2007, instructors in many of the large introductory science courses at Ryerson University have promoted active learning through the use of PI, aided by personal response systems. The PI method used includes posing questions to foster student reflection and to challenge them to identify gaps in their understanding. Typical PI classes consist of brief (typically 7-15 minutes) lecture segments interspersed with short quizzes consisting of conceptual questions, mostly in a multiple-choice format. Each student first answers the question individually. After a first poll of their responses, students spend a few minutes debating their choices with two to four peers, attempting to agree on the correct answer. In the process of PI, students learn to focus on concepts rather than memorizing facts. Previous research shows that these PI activities have had a positive effect on students' performance and retention in this context (Antimirova, Noack & Milner-Bolotin, 2009; Milner-Bolotin et al., 2011).

In 2011, one instructor introduced an additional activity that aimed to overcome some of the problems associated with the reliance of PI pedagogy on multiple-choice format questions. The goal of this additional intervention was to enhance student learning through the creation of new MCFQs, increase students' engagement and collaboration through small-group activities, provide the students opportunity to formulate their own ideas and promote students' overall sense of responsibility for their own learning and evaluation. The new activity extends the existing PI in a large, introductory physics course by requiring students to collaborate in small groups to create their own MCFQs based on various physics scenarios. The students were prompted to write their own conceptual questions in a multiple-choice format that included the correct answer as well as distractor answers that represented possible, common misconceptions. The additional activity was based on earlier work by the course instructor that replaced some PI activities with stand-alone collaborative group activities based on cognitive conflict (Kalman, Milner-Bolotin & Antimirova, 2010). Although the results of these collaborative group activities were encouraging in terms of improving students' conceptual understanding of the topics addressed, they were too time-consuming to be used in the classroom on a regular basis.

The innovative elements in this intervention were:

a) Student-generated multiple-choice format questions (MCFQs)
b) The creation of MCFQs incorporated as small-group collaborative activities during class time and online
c) If successful, the incorporation of student-generated MCFQs into PI activities during class time and evaluations

1.4 Research Objectives

The purpose of this study was to evaluate the effectiveness of peer instruction supplemented with the new small-group collaborative MCFQ-writing activity, as compared to the more standard peer instruction without the additional activity.

This study aimed to answer three main research questions:

a) What is the effect of our model of interactive engagement (modified peer instruction supplemented by small-group collaborative activities that resulted in the development of student-generated multiple-choice questions) as compared to the standard PI model in terms of students' gain in conceptual learning?
b) What is the effect of our model of interactive engagement as compared to the standard PI model in changing the students' attitudes towards physics from 'novice-like' to 'expert-like'?
c) What are students' perceptions of their learning in our model of interactive engagement?

We conclude by summarizing the results of these three questions in order to make a general assessment of the relationship between our model of interactive engagement and conceptual learning in introductory science classes.

2. Class Pedagogy and Practices

2.1 Class Content, Structure, and Learning Supports

This study was conducted during the Fall 2012 and Fall 2013 semesters in the course 'PCS120: Physics I', a mandatory introductory physics course for first-year Faculty of Science undergraduate students at Ryerson University. Physics 1 is described in the university course calendar as "a calculus based course covering fundamental physics concepts: units, vectors, linear motion, circular motion, force and motion, work and energy, collisions, gravitation, electrostatics, capacitance, and simple DC circuits." This course is typically taught in two or three parallel lecture sections. Although the different sections are taught by different instructors, they have identical syllabi, common online assignments, mandatory labs and tutorials, and a common grading scheme for major evaluations (the midterm test and final exam). Enrolment in this course has been growing steadily since the inception of the science programs at Ryerson in the Fall 2005 semester and currently there can be more than 500 students registered each semester.

The course consists of three hours of lecture each week (usually in one two-hour session and one one-hour session), one hour of tutorial each week, and a two-hour laboratory session every other week. Tutorials and laboratory sessions are led by teaching assistants, who are normally graduate students in physics (or occasionally engineering). The number of students per tutorial and lab sections does not exceed 25 students.

In the Fall 2012 semester, the course was offered in two sections, each taught by a different instructor. In the Fall 2013 semester, the course was offered in three sections, each taught by a different instructor. In all five iterations of the course, the instructors used PI techniques that enabled students to work through material collaboratively. The majority of the questions used for PI were conceptual multiple-choice questions that did not involve numerical calculations. In each case the PI pedagogy was supported by the use of clickers. In both years, students were required to read the textbook prior to attending lecture. In the Fall 2012 semester, the instructors did not monitor whether the readings were completed; in the Fall 2013 semester, students were required to complete short, online pre-lecture quizzes based on the reading prior to covering the same material in lectures. In addition, the course instructors use tablets for their lectures and rely on Tablet PCs' pen technology to conduct derivations in front of the class instead of presenting static slides; these instructor-annotated materials can be further annotated by students.

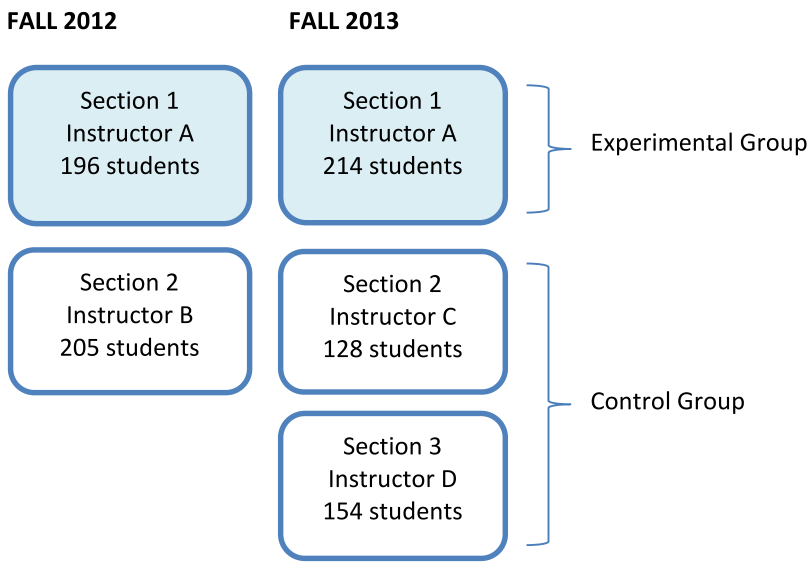
Students enrolled in the course are taught in state-of-the-art undergraduate laboratories, which provide access to probes, sensors and the LoggerPro software from Vernier Software & Technology (http://www.vernier.com/) that allow learners to collect data in real time and to store and manipulate it. Video analysis tools, which involve recording videos and retrieving the experimental data from the recordings, are also available. These new technologies for data collection, streaming and storage mean that students can focus their attention on concepts, implement multiple experimental scenarios and efficiently manipulate, visualize and analyze the collected data.

In addition, students had access to substantial learning supports outside of the classroom and laboratory sessions. Lecture notes and additional materials, including solutions to tutorial questions, examples of past exams, and practice problems, were made available to students throughout the duration of the course using the university's online course management system (Blackboard). Students were also required to use the online self-tutoring and homework program, 'MasteringPhysics' (http://www.pearsonmylabandmastering.com/northamerica/masteringphysics/), which allows students to progress through problem-solving at their own pace to complete several comprehensive homework assignments during the term. After each chapter, students were asked to complete online homework assignments consisting of skill-building tutorial items and traditional physics problems requiring derivations and numerical answers. Following a just-in-time teaching approach (Simkins & Maier, 2010), students were encouraged to send their questions to the instructor after finishing the assigned reading and prior to the lecture. To the best of our knowledge, not many students used this opportunity. Finally, students had access to regular, drop-in peer facilitated study sessions and instructor office hours.

2.2 The Multiple Choice Format Question-Writing Activities

One section of the course in the Fall 2012 semester and one section of the course in the Fall 2013 semester incorporated the new modified form of PI, in which students were asked to collaborate in small groups to create their own MCFQs. The two course sections that used the new activities were taught by the same instructor (a research team member), whereas the three course sections that used traditional PI were each taught by different instructors. Students in the two sections that used the new activities (one in the Fall 2012 semester and one in the Fall 2013 semester) are referred to as the 'experimental' group, whereas students in the remaining sections are referred to as the 'control' group.

Figure 1: Research Design



During the first half of the course, students in the experimental group were taught to develop MCFQs in class. Often, students were presented with open-ended questions during the lecture and then asked to collaborate in small groups (three to five students) in order to produce several plausible answer options for a MCFQ that was subsequently discussed by the entire class. This collaborative activity required more class time than the use of standard PI techniques. Standard PI activities often include three or four questions that can be discussed together in 15 to 20 minutes. In contrast, the collaborative activity of discussing an open-ended question, proposing plausible incorrect answer options related to common misconceptions and formulating a correct answer option takes at least 20 minutes for a single question. Because of the amount of time this activity required, it could not be used every class or even every week. There is some risk associated with introducing time-consuming class activities too often, in case they do not substantially contribute to students' learning and reduce the amount of time spent on more effective strategies. Students did not receive any additional credit for participating in the in-class MCFQ-writing activities. It was simply impractical to track students' participation in class activities in a large-class environment. The main purpose of the in-class activities was to introduce the students to writing MCFQs and model the process they could use to write questions on their own.

During the second half of the course, students were prompted to participate in creating MCFQs using the online course discussion board within the course management system. Students received credit for completing one of the following activities: posting a physics scenario that could be used to create a multiple choice question(s), posting the entire multiple-choice question, posting an answer option (a correct answer or an incorrect answer which is a plausible distractor) for a scenario posted by other students, or providing meaningful comments or constructive critique to posts by other students. Students were reminded that the critique of their peers' contributions should be constructive and aimed at the improvement of the questions. In order to obtain credit for this activity, students were required to address qualitative or semi-quantitative questions. Purely numerical questions requiring calculations were not credited. In addition, the problems posted had to have some element of originality.

Students enrolled in the experimental group were able to earn 2% in bonus marks for voluntarily participating in creating MCFQs and posting their work on the online course discussion board. About 60% of the students took advantage of this opportunity. When students began submitting work, it became clear that some students did not take the requirement of originality seriously and they tried to post existing questions and problems taken directly from the textbooks, websites and test banks. For the most part this seemed to be an unintentional mistake of students not understanding the meaning of 'original' and instead posting the questions that they liked to share with the rest of the class. This unintended outcome provided an opportunity to engage students in a conversation about plagiarism. It also made it obvious that students need very clear and detailed instructions about what this activity requires, as many students have never before generated their own MCFQs. Although the intention was to incorporate student-generated MCFQs into formal course evaluations, the relatively low quality of the questions that were generated led to the abandonment of this plan.

3. Methodology

The impact of the new MCFQ-writing activities on students' course outcomes was assessed using a combination of quantitative and qualitative approaches.

3.1 Quantitative Measures

Robust methods for evaluating the effectiveness of instruction are well established in the discipline of physics. Pre- and post-instruction testing using reliable diagnostic instruments such as concept inventories and attitude assessments were developed in the mid-1990s and a tremendous amount of data have been collected from a variety of physics classes in institutions worldwide. This study relies primarily on two commonly used, validated measures in physics education research: the Force Concept Inventory (FCI) (Hake, 1998) and the Colorado Learning Attitudes about Science Survey (CLASS) (Adams et al., 2006). The Force Concept Inventory (FCI) is a well-established diagnostic instrument for measuring students' understanding of the basic Newtonian mechanics that have traditionally been the focus of introductory physics courses (Hestenes et al., 1992; Huffman & Heller, 1995; Thornton & Sokoloff, 1998; Thornton et al., 2009). The FCI measures conceptual learning through 30 multiple-choice questions that ask students to apply scientific concepts to understand and solve 'real-world' problems. Although some researchers have raised concerns about the limited validity of the tool, the vast majority of data from the application of the FCI over the past twenty years suggests that student performance on this instrument provides a good measure of conceptual learning and correlates positively with other diagnostic instruments (Thornton & Sokoloff, 1998; Thornton et al., 2009). The FCI is designed to be administered to students at two points in time: the beginning and end of an instructional unit (such as a course). The difference between the pre-instruction and post-instruction scores represents students' absolute gain in conceptual understanding. More typically, researchers using the FCI focus on students' normalized gain in conceptual understanding, also known as Hake's gain. Hake's gain is a ratio that captures the amount of new material students learned in the course relative to the maximum amount of material they could have learned (the previously unlearned material). Hake's gain (g) is calculated as follows:

$$< g >= \frac{FCI_{t_2} - FCI_{t_1}}{FCI_{max} - FCI_{t_1}}, \text{where } FCI_{max} = 30$$

In this report, Hake's gain is used as the primary measure of students' improvement (or decline) in conceptual understanding of the material covered in Physics 120.

Students' attitudes towards science were measured using the Colorado Learning Attitudes about Science Survey (CLASS), which evaluates students' attitudes in eight sub-areas and distinguishes between novice- and expert-like attitudes about science (Winter, Lemons, Bookman & Hoese, 2001; Adams et al., 2004; Perkins et al., 2005; Adams et al., 2006). The CLASS contains 42 Likert-style statements grouped into eight categories: personal interest, real-world connections, conceptual connections, sense-making/effort, problem-solving sophistication, problem-solving confidence, problem-solving in general and applied conceptual understanding. These categories were empirically developed based on students' responses to the survey and demonstrate that students have relatively consistent ideas about learning physics and problem solving. Like FCI, the CLASS is a widely used and validated tool that is designed to be administered at the beginning and end of an instructional session (such as a course), in order to evaluate changes in students' attitudes. There are detailed, standardized methods for the analysis of CLASS data, including tests for response set bias and a 'specified response' question located near the end of the survey to assess whether students are reading each question carefully.

In addition to these two validated measures, the research team developed a socio-demographic survey to gather more information to supplement the analysis. The survey collected information about students' demographic characteristics (such as age, sex, visible minority status, English-language proficiency, immigration status, disability, living situation), their educational background (completion of previous science/math courses), their additional time commitments during the year (working for pay, studying, commuting), as well as some information about their parents (level of education, occupation and interest in science). Since previous research has shown that students' socio-demographic characteristics affect their learning outcomes (Noack, Antimirova & Milner-Bolotin, 2009; Milner-Bolotin et al., 2011), select characteristics were used as control variables in the regression model predicting the effect of the intervention in this study.

Official course grades are not used as a measure of success in this study, since it was not feasible to collect this information from participating students prior to their completion of the course.

3.1.1 Quantitative Data Collection and Analysis

In each of the Fall 2012 and Fall 2013 semesters, data were collected during week two of instruction (the pre-test) and week twelve of instruction (the post-test) in each section of Physics 120. In week two of the Fall 2012 term, students were asked to complete three paper surveys (the FCI, the CLASS, and the demographic survey) during two separate fifty-minute sessions; in week two of the Fall 2013 term, students were asked to complete the three surveys in a single eighty-minute session. Students were informed ahead of time of the research study, of the voluntary nature of participation and were also asked to complete an informed consent form that outlined in detail the confidentiality of their responses. All students were given course participation credits for completing the FCI, regardless of whether they consented to have their data included in this study or not. In order to avoid any real or perceived bias, the instructor who was a member of the research team was not present in the classroom when the study was introduced to students, nor when consent forms and socio-demographic surveys were being administered. Instructors were not informed which students consented to have their data included in this study until after the course was completed and final grades had been submitted. In week twelve of each term, students completed the FCI and the CLASS in a single one-hour session and answered three additional questions about their class attendance and their use of learning supports during the term. Again, students were given course participation credits for the completion of the FCI, regardless of whether they consented to have their data included in this study or not. This research protocol was approved by the Ryerson University Research Ethics Board.

The FCI and the CLASS were administered using Scantron forms; these data were processed electronically and imported into SPSS for analysis. Data from the socio-demographic survey were entered manually into SPSS. Data were matched using university-assigned student identification numbers. Only students who i) consented to participate in the study and ii) completed at least one pre-test/post-test pair of measures (FCI or CLASS) are included in the analyses below. Students whose CLASS results showed response set bias or a failure to read each question were removed from the analysis of student attitudes. Bivariate analysis was performed to assess whether there were statistically significant demographic differences between the experimental and control groups, and to answer the first two research questions identified above: the effect of introducing MCFQ-writing activities on students' gain in conceptual learning, and the effect on students' attitudes towards physics.

3.1.2 Description of Respondents

In total, 465 out of 896 students consented to participate in the research and completed at least one pre-test/post-test pair of measures, for a partial response rate of 52% (see Table 1). Of those, 388 students completed all of the components of the research used in the analysis below (pre- and post- test FCI, pre-and post- test CLASS, and the socio-demographic survey), for a full response rate of 43%. This discrepancy is primarily due to students who did not complete the CLASS at one or both data collection sessions, since they were not allotted course participation credits for doing so. As shown in Table 1, response rates were slightly higher in 2012 than in 2013. This may be because the presentation of the research and the distribution and collection of the tests, surveys and consent forms was less rushed in 2012, since it was spread across two class sessions and students had more time to ask questions. In 2013, some instructors were only willing to allot a single class session to data collection at the beginning of the term and we strove to use the same schedule for all of the course sections in order to ensure consistency. As a result, the presentation of the research and the distribution and collection of the research material were substantially more rushed in 2013.

Table 1: Response Rates, by Year and Section

Course	Students enrolled in Physics 120	Students who consented and completed at least one post-test	Students who consented and completed all measures	Partial response rate	Full response rate
Fall 2012					
Experimental Section (Instructor A)	196	127	107	64.8%	54.6%
Control Section (Instructor B)	205	113	99	55.1%	48.3%
2012 Total	*401*	*240*	*206*	*59.9%*	*51.4%*
Fall 2013					
Experimental Section (Instructor A)	214	93	76	43.5%	35.5%
Control Section 1 (Instructor C)	128	65	56	50.8%	43.8%
Control Section 2 (Instructor D)	154	67	50	43.5%	32.5%
2013 Total	*496*	*225*	*182*	*45.4%*	*36.7%*
Total	897	465	388	51.8%	43.3%

Table 2 shows the distribution of respondents on key demographic and educational characteristics overall and for the experimental and control groups separately. Notably, there are few significant differences between the composition of the control group and the experimental group, suggesting that differences in learning outcomes between the two groups are related to instructional practices and not pre-existing student differences.

Overall, the majority of students who participated in this research were women (55%) and were aged 18 or younger (62%). Reflecting Ryerson University’s diversity as an urban university, the majority of students (60%) reported that they were members of a visible minority group and 40% of students were born outside of Canada. Almost half of students (47%) report a mother tongue other than English and about one-third of students (35%) report that English is not the primary language that they use at home.

Three out of five students who participated in this research (62%) report that they successfully completed a grade 12 physics course, with most receiving an ‘A’ or ‘B’ grade. About two-thirds of students (65%) say that they prefer to study alone, which stands in contrast to the collaborative small-group format of the new activity that was introduced.

The vast majority of students in both the control and the experimental groups report attending “all” or “most” of the lectures and tutorials each term (attending 9 to 13 weeks). This high attendance reporting may be influenced by the fact that data collection took place during class time, thus students who missed many classes are less likely to be included in this study.

Table 2: Distribution of Student Characteristics, by Group (n=449)

Characteristic		Experimental (%)	Control (%)	Total (%)
Gender	Women	57.2	54.2	55.6
	Men	42.8	45.8	44.4
Age Group	18 years old or younger	62.6	63.1	62.9
	19-22 years old	25.2	29.6	27.6
	23 years or older	12.1	7.3	9.6
Visible Minority Status	Visible minority	64.0	57.3	60.4
	Not a visible minority	36.0	42.7	39.6
Immigrant Status	Immigrated less than 10 years ago	20.1	21.9	21.1
	Immigrated 10 or more years ago	21.1	18.0	19.5
	Born in Canada	58.8	60.1	59.5
Mother Tongue	Another language	45.4	47.9	46.7
	English	54.6	52.1	53.3
Primary Language Spoken at Home	Another language	35.6	33.6	34.5
	English	64.4	66.4	65.5
Completed Grade 12 Physics	No	38.3	38.0	38.1
	Yes	61.7	62.0	61.9
Grade in Grade 12 Physics	A	37.2	35.9	36.5
	B	35.1	40.1	37.7
	C	14.1	13.8	14.0
	D	6.3	2.8	4.4
	Don't know	7.3	7.4	7.4
Program of Study*	Biology	42.1	29.6	35.4
	Biomedical science[1]	8.4	22.3	15.9
	Chemistry	14.9	13.3	14.0
	Contemporary science	9.4	5.2	7.1
	Mathematics and its applications	6.4	9.4	8.0
	Medical physics	8.4	12.9	10.8
	Undeclared science	10.4	7.3	8.7
Preferred Study Method	Studying alone	63.6	65.8	64.7
	Studying with one other person	23.3	20.7	22.0
	Studying with two or more people	13.1	13.5	13.3

Characteristic		Experimental (%)	Control (%)	Total (%)
Lecture Attendance	All or most (9-13 weeks)	87.6	89.5	88.6
	Fewer (0-8 weeks)	12.4	10.5	11.4
Tutorial Attendance	All or most (9-13 weeks)	90.8	91.0	90.9
	Fewer (0-8 weeks)	9.2	9.0	9.1

[1] Fall 2013 was the first year of the biomedical science program at Ryerson University, and thus students from this program only participated in the second year of data collection.
* Indicates a statistically significant difference between groups at the p < .05 level using chi-square tests

The only statistically significant difference between the experimental and control groups is related to the program that students were enrolled in (x^2=25.26, *df*=6, p<0.000). This discrepancy likely relates to patterns in the scheduling of the required first-term courses for students in the different degree programs. Students in the experimental group were more likely to be enrolled in biology, contemporary science or undeclared science (a general entry program). In contrast, students in the control group were more likely to be enrolled in biomedical science, mathematics or medical physics. Students enrolled in different programs might have different motivations to learn physics. Notably, despite the varied admissions requirements for each degree program, students in both groups were equally likely to have completed grade 12 physics.

3.2 Qualitative Data Collection and Respondents

In addition to the quantitative measures described above, the research team felt that it was important to understand students' subjective experiences with developing MCFQs in small groups. We were interested in collecting students' perceptions about the usefulness of the new activity, understanding whether they felt that it contributed to their learning compared to other active learning elements used in the class, and learning about any difficulties they encountered. The consent form included a question asking students whether they were willing to be interviewed by a member of the research team after the completion of the course. Interviews were not conducted until the Winter 2013 and Winter 2014 semesters (respectively) in order to avoid the course instructors becoming aware of which students had agreed to participate in the research during the course delivery. In total, 81 students from the experimental group agreed to be interviewed (32 in 2012 and 49 in 2014). Each January, students who volunteered to participate in an interview were sent a recruitment email reminding them about the study and describing the next steps for scheduling an interview. Few students replied to these emails; in total, only 18 students were interviewed, even after multiple follow-up efforts. Students who completed an interview were given a $20 honorarium.

Each interviewer used a semi-standardized interview guide and interviews lasted from 25 to 50 minutes. The students were asked questions probing their perception of the pedagogy and technology tools used in the course and were prompted to reflect on their own learning. In particular, they were asked about how each of the different components of the course contributed to their learning. Each interviewee was asked to sign an informed consent form prior to the interview and reminded about their right to decline to answer any question or stop participating at any time. Interviews were both audio- and videotaped and selectively transcribed. The interview data were analyzed for key themes by the research team. Overall, 11 women and seven men were interviewed. The majority of interviewees were in their first year of study. Interviewees were enrolled in a wide range of programs: six were in biology, three were in chemistry, two were in mathematics, two were in contemporary science, two were in medical physics, two were in biomedical science and one was unknown (likely a non-certificate student).

4. Key Findings

4.1 Research Question 1: The Effect of MCFQ-Writing Activities on Students' Gains in Conceptual Learning

What is the effect of our model of interactive engagement (modified peer instruction supplemented by small-group collaborative activities that resulted in the development of student-generated multiple-choice questions) as compared to the standard PI model in terms of students' gain in conceptual learning?

As described above, students' conceptual learning was measured using the 30-question Force Concept Inventory – Physics. Table 3 shows the absolute and normalized (Hake's) gain in FCI scores between week 2 and week 12 for students overall, by year and by group. The overall average Hake's gain of 31% was in the expected range for a course using interactive instruction techniques, though it fell at the lower end of the range, as might be expected for larger classes (Hake, 1998). There was no statistically significant difference in the absolute or normalized gains in conceptual learning between the 2012 and the 2013 study cohorts (absolute: F=0.124, *df*=1, 416, p=0.725; normalized: F=0.762, *df*=1, 415, p=0.383). There was, however, a statistically significant difference in both absolute and normalized gains between the experimental and control groups (absolute: F=8.225, *df*=1, 416, p=0.004; normalized: F=5.886, *df*=1, 415, p=0.016). Students in the experimental group had the largest average normalized learning gain (33%), suggesting that the MFCQ-writing activities promoted students' learning of previously unknown physics knowledge. The average normalized learning gain for the control group was 6 percentage points lower. Students in the experimental group also had the largest absolute learning gain, getting an average of 6.6 more questions correct on the FCI post-test than on the pre-test. In contrast, the average absolute gain for the control group was 5.4 questions.

Table 3: Normalized and Absolute Mean Gains in Student Learning, by Group and Year (n=418)

	Overall Average	Average Outcome by Group		Average Outcome by Year	
		Experimental	Control	2012	2013
Hake's Gain (normalized)	30.7% (*SD*=25.4)	33.9%* (*SD* =23.5)	27.9%* (*SD*=26.8)	29.7% (*SD*=27.7)	31.9% (*SD*=22.6)
Absolute Gain	6.0 (*SD*=4.3)	6.6* (*SD*=4.7)	5.4* (*SD*=3.8)	6.0 (*SD*=4.6)	5.9 (*SD*=3.9)
FCI Pre-Test Score	10.1 (*SD*=5.4)	9.8 (*SD*=5.6)	10.3 (*SD*=5.6)	9.9 (*SD*=5.1)	10.2 (*SD*=5.7)
FCI Post-Test Score	16.0 (*SD*=5.8)	16.4 (*SD*=5.8)	15.7 (*SD*=5.8)	15.9 (*SD*=5.6)	16.1 (*SD*=5.9)

*Indicates a statistically significant difference between groups at the p < .05 level using one-way ANOVA tests

An analysis of the relationships between normalized (Hake's) learning gains and students' demographic and educational characteristics, as shown in Table 4, shows some clear trends. Similar to the results of previous research (Noack, Antimirova & Milner-Bolotin, 2009), women have significantly lower average normalized learning gains than men (F=8.463, *df*=1, 397, p=0.004). In this study, women's average normalized gain was 7.4 percentage points lower than men's. Even though females outnumber males overall at the undergraduate level, women are still much less likely than men to major in mathematics, engineering and science or to choose a profession or pursue graduate education in these fields. This outcome often is considered to be one of the effects of negative sex-based stereotypes. Our results suggest that gender stereotyping (Guimond & Roussel, 2001) continues to affect women's physics learning even if they choose to study science, and even given the presence of women as role models – in this study, women taught three of the five course sections. Women's average normalized gains did not differ significantly in relation to the gender of the course instructor (F=1.542, *df*=1, 220, p=0.216). Female students may be disadvantaged due to stereotypical perceptions that women are primarily interested in "caring" or "soft" sciences (such as nursing and midwifery) which carry different academic expectations in comparison to "research" or "hard" sciences. Female students may experience disadvantages in classroom interactions as a result of gendered assumptions by other students, teaching assistants or instructors, impacting their willingness to speak up, ask questions or challenge ideas. Finally, women might feel uncomfortable in or be disadvantaged by the masculinized culture of science overall (Kelly, 1985).

In addition, the results show that mature students had significantly higher average normalized gains than students in the typical 18-22 year-old age range (F=3.319, *df*=2, 391, p=0.037). Mature students often have different motivations for learning than their younger counterparts and may be more invested in improving their level of conceptual understanding. Intrinsic motivation is what often distinguishes mature students from their younger counterparts (Murphy & Roopchand, 2003).

Though visible minority status and immigrant status have no clear effect on normalized learning gains (visible minority status: F=0.386, *df*= 1, 390, p=0.535; immigrant status: F=3.592, *df*=1, 396, p=0.059), it is clear that English-language proficiency does. Students who speak English as a mother tongue and/or primarily speak English at home have significantly higher average normalized gains than those who do not (mother tongue: F=4.888, *df*=1, 396, p=0.028; home language: F=6.941, *df*=1, 393, p<0.000). Since English is the primary language of the classroom at Ryerson University, it may be that these students gain less from their classroom instructional time and from participating in small-group activities. Students who are less confident in their English skills – for whatever reason – might also be less able to decipher effectively the MCFQs used in the FCI, and thus this measure may be a less accurate reflection of their conceptual learning. This result mirrors those of other researchers; for example, researchers who systematically administered the FCI to students at the Khalifa University of Science, Technology and Research (KU) in the United Arab Emirates found that normalized gains were strongly modulated by language proficiency (Hitt et al., 2014).

Although there is some variation in students' average normalized gains by program of study, with mathematics students having the highest average normalized gain and contemporary science students having the lowest, these differences are not statistically significant (F=0.748, *df*=6, 384, p=0.611). This suggests that the unequal distribution of students from different programs in the experimental and control groups is not a substantial source of concern.

As expected, students who report high lecture and tutorial attendance have significantly higher normalized gains, on average, than those who report attending fewer sessions (lecture: F=4.692, *df*=1, 356, p=0.031; tutorial: F=5.805, *df*=1, 356, p=0.016). Students who report attending all or most lectures have an average normalized learning gain that is 9.4 percentage points higher than those with lower attendance; similarly, students who report attending all or most tutorial have an average normalized learning gain that is 11.2 percentage points higher than those with lower attendance. High attendance exposes students to course concepts multiple times and in a variety of formats, offers opportunities for question-asking and suggests that the student is engaged in their learning overall.

Table 4: Mean Learning Gain, by Students' Demographic and Educational Characteristics (n=417)

Characteristic		Average Hakes Gain (%)
Sex *	Female	27.1
	Male	34.5
Age Group*	18 years old or younger	29.7
	19-22 years old	27.9
	23 years or older	40.2
Visible Minority Status	Visible minority	30.6
	Not a visible minority	32.0
Immigrant Status	Immigrated less than 10 years ago	25.0
	Immigrated 10 or more years ago	29.8
	Born in Canada	32.7
Mother Tongue*	Another language	27.3
	English	33.0
Primary Language Spoken at Home*	Another language	25.6
	English	32.7
Completed Grade 12 Physics	No	29.6
	Yes	30.6
Grade Achieved in Grade 12 Physics	A	32.9
	B	27.7
	C	31.0
	D	24.6
	Don’t know	28.8
Program of Study	Biology	29.4
	Biomedical science	32.0
	Chemistry	31.4
	Contemporary science	24.9
	Mathematics and its applications	36.4
	Medical physics	28.8

Characteristic		Average Hakes Gain (%)
	Undeclared science	26.7
Preferred Study Method	Studying alone	30.5
	Studying with one other person	31.1
	Studying with two or more people	29.2
Lecture Attendance*	All or most (9-13 weeks)	31.4
	Fewer (0-8 weeks)	22.0
Tutorial Attendance*	All or most (9-13 weeks)	31.4
	Fewer (0-8 weeks)	20.2

*Indicates a statistically significant difference at the $p < .05$ level using one-way ANOVA tests

Multivariate linear regression analysis allows the development of a model that isolates the independent effects of students' personal characteristics as well as the effect of the new MCFQ-writing activities in predicting students' normalized conceptual learning gains. Students' personal characteristics that had a statistically significant relationship with normalized learning gains (from the bivariate analysis shown in Table 4) were included in the model. Students' age was included in the regression model as a continuous variable in an effort to increase the explanatory value of this predictor. The high correlation between the two indicators of English-language proficiency (mother tongue and primary home language) made it difficult to estimate the independent effect of each, so these two characteristics were combined into a single predictor indicating whether a student speaks a language other than English as a mother tongue *and* at home. Similarly, the high correlation between the two attendance indicators (lecture attendance and tutorial attendance) necessitated combining these two characteristics into a single predictor indicating whether a student reported low lecture *or* low tutorial attendance (or both).

The resulting ordinary least squares regression model shows how much of the variation in normalized learning gains can be explained by students' personal characteristics and participating in the MCFQ-writing activities. Table 5 shows the best-fitting model for the sample overall and for each year's study cohort separately. The adjusted R^2 value provides an overall measure of model fit by estimating how much of the variation in normalized learning gain is explained by the predictors included in the model. In the overall model, students' personal characteristics explain about 6.3% of the variation in normalized learning gains; the explanatory power of the model increases to 7.1% when the experimental intervention is included as a predictor. Interestingly, the model is more effective at predicting normalized learning gains for the 2013 study cohort than for the 2012 cohort. For the 2012 cohort, students' personal characteristics explain only 2.5% of the variation in normalized learning gains, though this increases to 4.8% when the experimental intervention is included as a predictor. For the 2013 cohort, students' personal characteristics have much more explanatory power: they account for 16.3% of the variation in normalized learning gains. For the 2013 cohort, however, the experimental intervention does not explain any further variation – in fact, the predictive power of the model decreases, as the calculation of the adjusted R^2 value implements a penalty for including non-explanatory predictors in the model. Overall, students' personal characteristics, along with the experimental intervention, allow us to predict only a relatively small proportion of the variation in normalized learning gains. It is likely that other, unmeasured (or unmeasurable) factors also have a substantial influence on students' learning outcomes in introductory physics courses. These might include, but are not limited to: satisfaction with program of study, commitment to academics overall, commitment to the specific course, attitudes towards the instructor or classroom environment, conflicting familial or personal obligations, the effect of peer interactions and pressures, and the quality of previous mathematics/physics instruction.

The ‘constant’ value for each model shown in Table 5 represents the predicted value of the normalized learning gain for a hypothetical 18 year-old, male student in the control group, who speaks English as a first language or at home and who attended all or most course lectures and tutorials. For the overall model, this hypothetical student is predicted to have a normalized learning gain of 35%. Students in the course sections that included the small-group MCFQ-writing activities are predicted to have a normalized learning gain that is 5 percentage points higher, even after taking students’ personal characteristics into account. An analysis by cohort, however, shows that the effect of the small-group MCFQ-writing activities is most pronounced in the 2012 cohort and is much smaller in the 2013 cohort. We do not have an obvious explanation for this result. In terms of the class activities, students in the 2013 cohort differed in that they were required to complete online, pre-lecture quizzes to assess their understanding of the textbook readings. The 2013 cohort also included students enrolled in the new biomedical science program (which attracted students with relatively high entrance grades). However, these differences applied to both the experimental and control groups. The smaller predicted difference in normalized learning gains as a result of the MCFQ-writing activities in the 2013 cohort may be the result of differences in the PI activities used by the control group instructors in 2013 compared to 2012 (while the experimental group was taught by the same instructor, the control group instructors differed each year).

Table 5: Predicted Normalized Student Learning Gains, Overall and by Year, based on OLS Regression

		Year	
	Overall Model (n=344)	**2012 (n=172)**	**2013 (n=172)**
Constant	35.1*	27.5*	44.2*
Sex (Female)	-8.0*	-1.3	-16.6*
Age[1]	1.0	1.3	0.1
Less English[2]	-9.1*	-9.4	-8.4*
Low Attendance[3]	-10.5*	-9.4	-10.7*
MCFQ-Writing Intervention	***5.4****	***9.8****	***1.4***
Adj. R Squared (R^2)	7.1%	4.8%	15.9%

[1] Age in years (not grouped) centered around age 18

[2] Students who did not learn English as a first language nor is English the primary language spoken at home

[3] Students who reported low attendance at either lectures or tutorials (or both)

* Indicates a statistically significant difference at the $p < .05$ level

In addition to showing the inconsistent effect of the MCFQ-writing activities on students' normalized learning gains, this model also shows that students' personal characteristics can independently have just as much impact on students' conceptual learning. While characteristics such as gender and age are fixed, students' attendance can potentially be influenced by an instructor's motivational strategies. Although it may be possible for physics instructors to influence students' English-language proficiency, there is typically limited time and few resources available to support students who speak English as an additional language in large classes.

Women are predicted to have normalized learning gains that are 8 percentage points lower than men. Women had significantly lower FCI pre-test scores than men (an average of 8.1 compared to 12.7) and thus had more previously unlearned material to master during the term. So, while men's and women's absolute learning gains were roughly equivalent (women had an average absolute gain of 6.0, compared to 5.7 for men), women's normalized learning gain was lower since they had more content to learn. The effect of gender was particularly pronounced among the 2013 study cohort; there is no clear explanation for this discrepancy. After other factors are controlled for, however, student age is no longer predicted to have a significant effect on normalized learning gains.

Students who are less familiar with the English language are predicted to have a normalized learning gain that is 9 percentage points lower than those who speak English as a mother tongue or as their primary home language. This result is consistent across both cohorts. Although students with less English-language proficiency had similar average FCI scores at the beginning of the course, they had lower absolute and normalized learning gains at the end of the course. This might demonstrate that these students experience more difficulties in understanding the classroom instruction; in particular, small-group and peer-oriented activities might have fewer learning benefits for students with limited English-language skills.

Unsurprisingly, students who attend most or all lectures and tutorials are predicted to have a normalized learning gain that is 10.5 percentage points higher than students who report attending less. This result was consistent across both cohorts. Though this finding is predictable, it reinforces the value of motivating students to attend lectures and tutorials through engaging activities, participation grades and other methods. Presence in lecture and tutorial increases students' exposure to course material, allows them to participate in peer-instruction activities and engages them in problem-solving in a supported environment.

Overall, it appears that the MCFQ-writing activities introduced in this study are positively associated with normalized learning gains, but that the effects are inconsistent. In the 2012 cohort, student characteristics appear to have little effect on normalized learning gains, whereas the MCFQ-writing activities were associated with a significant positive improvement in conceptual learning. In contrast, in the 2013 cohort, students' personal characteristics are strongly associated with normalized learning gains, but the MCFQ-writing activities appeared to have little effect on conceptual learning. In the overall model, gender, English-language proficiency and attendance are all associated with larger normalized gains than the MCFQ-writing activities. Taken together, these results suggest that the use of MCFQ-writing activities in large, introductory physics courses has some beneficial effect, but that this effect is not necessarily consistent across groups of students.

4.2 Research Question 2: The Effect of MCFQ-Writing Activities on Students' Attitudes towards Physics

What is the effect of our model of interactive engagement as compared to the standard PI model in changing the students' attitudes towards physics from 'novice-like' to 'expert-like'?

In this study, students' attitudes towards physics were measured using the CLASS tool. This tool distinguishes between novice-like and expert-like attitudes towards science overall and in eight empirically generated categories. We anticipated that participating in the small-group MCFQ-writing activities might encourage students to adopt more expert-like attitudes towards physics. Being put in the 'expert' and 'teacher-like' role of writing multiple-choice questions – some of which were intended to be used in evaluations – might have prompted some students to adopt more expert-like attitudes to science in general. Table 6 shows the average percentage of favourable responses (i.e., agreement with experts) among study participants, overall and in each of the eight CLASS categories. Of particular interest is the shift in favourable attitudes between the beginning and end of the course. A substantial body of research shows that participating in introductory physics courses actually prompts *less* expert-like attitudes among students (Perkins et al., 2004; Perkins et al., 2005; Pollock, 2004). In this study, students do not exhibit a significant overall negative shift in attitudes. A negative shift is, however, apparent in two areas: personal interest in physics (paired $t=3.050$, $df=369$, $p=0.002$) and sense-making/effort (feeling like the effort needed to make sense of the material is worthwhile; paired $t=4.325$, $df=365$, $p<0.001$).

In contrast, there is a statistically significant positive shift in students' attitudes in three areas: problem solving sophistication (paired $t=2.417$, $df=369$, $p=0.016$), conceptual understanding (paired $t=5.342$, $df=369$, $p<0.000$) and applied conceptual understanding (paired $t=5.095$, $df=369$, $p<0.000$). In the CLASS, 'conceptual understanding' captures whether a student understands physics as a coherent field with connections between topics, whereas 'applied conceptual understanding' reflects whether a student feels that they can apply a conceptual approach and reasoning when problem-solving, as opposed to simply memorizing or mimicking other solutions. The positive attitude shift in these two areas for all of the students included in this study reflects the success of PI and active learning in general in promoting students' engagement and conceptual learning in physics.

Participating in MCFQ-writing activities did not appear to have a substantial effect on students' attitudes overall or in any of the eight categories, compared to students in the control group. In fact, in the two key areas of 'conceptual understanding' and 'applied conceptual understanding', students who participated in the MCFQ-writing activities had a slightly smaller positive shift than students in the control group. In addition, students who participated in MCFQ-writing activities had a larger negative shift (a shift away from expert-like attitudes) when it comes to sense-making/effort, or feeling like it was worth putting in the effort to understand or solve a problem. The one better outcome for students who participated in the MCFQ-writing activities is in the area of personal interest, where these students simply did not have a negative attitude shift, whereas those in the control group did. This result is tempered by the fact that the average percentage of favourable attitudes in the area of personal interest at the end of the course is very similar for both groups; this outcome is a result of students in the control group having a higher percentage of favourable attitudes in the area of personal interest at the beginning of the course.

Table 6: Week 2 and Week 12 CLASS Results by Sub-categories, Overall and by Group[1]

	Overall (n=370)			**Experimental (n=173)**			**Control (n=197)**		
CLASS Sub-Category	Pre	Post	Shift*	Pre	Post	Shift*	Pre	Post	Shift*
Overall Attitudes	52.6	52.2	--	52.8	51.1	--	52.5	53.1	--
Personal Interest	54.3	50.2	-4.1	53.4	50.4	--	55.1	50.1	-5.0
Real World Connections	60.8	57.3	--	62.4	59.0	--	59.3	55.9	--
Problem Solving General	55.6	55.5	--	55.9	53.5	--	55.3	57.4	--
Problem Solving Confidence	55.8	55.1	--	54.9	53.1	--	56.6	57.0	--
Problem Solving Sophistication	36.1	39.2	3.2	35.6	38.8	--	36.4	39.6	--
Sense-making/Effort	65.7	59.3	-6.3	65.2	56.7	-8.5	66.2	61.7	-4.5
Conceptual Understanding	46.0	53.8	7.8	45.5	52.2	6.7	46.5	55.2	8.6
Applied Conceptual Understanding	34.5	40.9	6.4	34.9	40.9	6.0	34.2	40.9	6.7

[1] Average % of favourable responses (agree with experts) for each category is shown
* Only shifts that are statistically significant at the p<.05 level, using paired t-tests, are shown

Taken together, these results lead us to conclude that participation in small-group MCFQ-writing activities does not have a substantial effect on students' attitudes towards science, nor does participation in these activities make them more 'expert-like' in their approach to physics.

4.3 Research Question 3: Students' Perceptions of Active-Learning Strategies

What are students' perceptions of their learning in our model of interactive engagement?

Semi-structured interviews were used to probe the students' perceptions of the pedagogies and technology tools used to promote learning in this course.

4.3.1 Students' Perceptions of MCFQ-Writing Activities

Many of the interviewees from the experimental group were not able to recall the particular activities that focused on writing multiple-choice format questions, either in class and/or outside of the classroom. This is particularly puzzling, as 60% of students participated in the online activity at least once; it is possible that these students did not participate in interviews or that those students who did participate in the interviews did not identify the MCFQ-writing activities as particularly meaningful. Those interviewees who did recall participating in the MCFQ-writing activities did not rate it as particularly useful for their learning compared to the standard PI pedagogy. However, when interviewees were asked whether they would recommend continuing to use MCFQ-writing activities, the majority said yes.

When interviewees were asked whether they preferred to write MCFQs during class time or online, opinions were divided. Some students preferred the in-class guided activities that allowed them to collaborate with their peers in real time and get clarifications if needed, while others preferred doing the activities outside of class time. Students expressed concerns that small-group activities can use too much class time that could be spent more effectively. One interviewee said that there was “not enough time to create multiple choice questions” during the class and that it would be better to “take more time outside of class, need time to think.” Some students were skeptical about the ability of the students to produce good-quality MCFQs, saying, “Students cannot write questions above (the level) of what they know” or “Students do not have the same knowledge as professors.” Some also felt that the creation of questions is the job of the professor and not the students, and expressed concern about the potential use of the student-generated questions in formal course evaluations such as the midterm and the final exam.

Those interviewees who provided positive feedback about the MCFQ-writing activities recognized them as an opportunity to reflect on material, monitor their own understanding and try “getting into the mindset of the professor.” They stated that the activity forced them to reflect more on the concepts, solidified the material and encouraged them to be creative. One student said, “Creating multiple choice questions is like working backwards, [you] must incorporate multiple concepts, consider response choices, [if] it is right and possible ways to solve it.”

Although the MCFQ-writing activities proved to be less effective than expected, the interviewees expressed substantial support and appreciation for the rest of the active-learning opportunities in the course. These activities were common to all the sections (experimental and control). Below, we outline the prevailing themes that emerged from the interviews.

4.3.2 Students’ Perceptions of Peer Instruction and Personal Response Systems

Overall, interviewees were very positive about using PI with personal response systems (clickers) during the course lectures. Students appreciated the ability to instantly check their understanding of the concepts as they were covered in class. They also noted the value of having the instructor be able “to gain better understanding where the students are struggling” and said that the use of clickers meant that the “professor took time to do examples and not assume students understood.” Many interviewees also said that they enjoyed collaborating with peers, learning from and teaching other students in the class. Interviewees noted that “helping others is an opportunity to develop better understanding of the material” and “…students can discuss with classmates, discuss your methods and learn new methods.” Students felt that the use of MCFQs combined with clickers “prepares students for exams/midterms, for the types of questions.” Some students believed that PI with clickers promoted a better success rate in the course overall. Others mentioned that even using clickers for practice without marks attached would still be an incentive to attend the classes. When interviewees were asked about the number of MCFQ clicker questions used in each class, the majority of the students indicated that the frequency of clicker questions in the lectures (typically 4-6) was good and that the PI activities did not take up too much time or take away from the teaching/lecturing of the content but were used enough to gauge student understanding.

4.3.3 Students' Perceptions of Multiple-Choice Format Questions for Teaching and Evaluations

Some instructors complain that students have a negative attitude towards multiple-choice questions. In this project, interviewees did not express any persistent negative attitudes towards MCFQs, even for testing purposes. When asked about their perceptions of the advantages and disadvantages of MCFQs, interviewees commented that using multiple-choice questions for PI during the semester helped them to feel prepared for the multiple-choice components of the formal evaluations. The majority of interviewees were comfortable with the questions asked in multiple-choice format during the midterm test and final exam and felt aware of concepts and questions that would be covered. Interviewees correctly understood that the multiple choice questions allow instructors to identify areas where students are having difficulty with the material. Some interviewees indicated that MCFQs can be helpful, saying that "if you don't see your answer you can go back and do (the problem) again" and it is "reassuring to see your answer (as one of the options)." Other students stated that the use of the multiple choice format in lectures "makes discussion possible about why other options are not possible, it makes you think more" and provides "immediate feedback [about] right or wrong."

4.3.4 Students' Perceptions of Laboratory and Demonstration Tools

This course involved extensive use of laboratory tools aimed at bringing the experimental aspect of physics into our course. Real-time data acquisition tools (motion detectors and force probes) and motion video analysis supported by LoggerPro software from Vernier Software & Technology (http://www.vernier.com/) were used in the laboratory experiments and, to a lesser extent, for the classroom demonstrations. Interviewees were overwhelmingly in support of the use of these laboratory and demonstration resources. Students commented that using the LoggerPro software allowed them to visualize the concepts and "to see principles in action." The students appreciated the exposure to technology that was not available in high school and found that "putting theory into practice is helpful." Comments from interviewees about these technologies highlight that they helped them to apply concepts: "labs helped [me] to see experimental side of physics, made it a reality," "cool to see the applications," "how variables are measured and controlling for different conditions, how physicists do experiments," "improved the excitement for physics," "see the theory and experience the theories that you learn." In reference to the particularly difficult task of graphing motion, one student commented that "sometimes it is hard to piece together [material but] the technology makes it easier to visualize, it is happening and being graphed at the same time." In addition to the laboratory experiments, the students appreciated using simulations and short videos. They particularly valued the attempts to bridge learning in classroom with the real life phenomena: "Sharing the theory in action, using simulation in class; physics is not just theory [we] must understand what is happening in the world, watching simulations makes it real."

4.3.5 Students' Perceptions of Self-Tutoring, Homework and Reading Activities

Overall, interviewees were very positive regarding the use of the required self-tutoring and online homework provided through the 'MasteringPhysics' platform. Students felt that the system allowed them to practice problem-solving extensively with different types of questions that were not covered in class activities. They appreciated that they could complete the required course work online at their convenience. While many students commented that these activities were demanding and time-consuming, they believed that they were essential for developing the problem-solving skills for numerical problems and long answer questions.

Interviewees were asked their opinion about the requirement to read the text prior coming to the class. The majority of participants indicated that reading the textbook better prepared them for the lectures, made taking notes in the class more efficient and allowed them to identify what they did not understand before attending the class. However, despite clearly recognizing the benefits of pre-lecture reading, survey results show that most students reported completing the readings in advance of the lecture only about half of the time. Although the students had the opportunity to submit questions to the instructors before class, only a few students used this opportunity on a relatively regular basis. The majority of the students used this opportunity only rarely or never. Perhaps this 'just-in-time' teaching strategy would be better utilized if more students completed the readings in advance.

4.3.6 Students' Overall Course Impressions and Suggestions for Improvement

Interviewees were asked to provide their general impressions of the course and areas for improvement. Overall, most students expressed a great degree of satisfaction with the course. The majority of interviewees said that the course met expectations in terms of the grade they received, the knowledge they gained, and that it was a positive experience overall. Several students expressed regrets about not taking physics in high school, while some commented that they were surprised to enjoy the course as they previously believed that physics is “not interesting,” “too difficult” or “not relevant for everyday life.” The students commented on the need to understand the concepts and enjoyed the fact that they could immediately apply principles they learned in the lectures to solve problems, experiments in the lab or think about everyday life situations. Students did not recognize the peer instruction method as an activity or practice that was unique to their physics course.

Students said that their biggest challenge in the course was keeping up with the assignments and readings, time management and staying on top of school tasks. Several interviewees mentioned that with so many course components and evaluations, it was hard to prioritize their course activities. Nonetheless, students’ suggestions for improving the course included requests for more online resources such as extra questions, tutorials, step-by-step instructions on how to solve particularly difficult problems (possibly in the form of screen capture or short video), and more in-class demonstrations.

5. Conclusions

5.1 Study Limitations

This study has several limitations. First, the inquiry was restricted to an introductory physics course at a single institution. Second, because the use of the new activity was untested, the amount of in-class time spent on collaboratively writing MCFQs was limited, so as not to risk negatively affecting students' learning outcomes. Another limitation that may have affected results is that different instructors taught the control group in the different years of the study. The addition of pre-lecture quizzes based on the reading in the second year of the study also limits the comparability between cohorts. Finally, the relatively small number of interviewees also limits the generalizability of the qualitative findings.

5.2 Overall Effectiveness of the MCFQ-Writing Activities

Overall, our study found that incorporating MCFQ-writing activities into PI instruction methods in a large course has some positive, though inconsistent, effects on students' conceptual learning gains. The use of these activities has little effect on promoting more 'expert-like' attitudes about physics. There are several possible explanations for these results. First, this new activity was added to an array of active-learning tools that were already being used in the course. While it is desirable to provide students with multiple ways to participate in a class, the provision of too many tools, tasks and activities could actually have the opposite effect by creating cognitive overload (De Jong, 2009). In fact, many interviewees noted the difficulty of prioritizing between different course activities. In this particular case instructors must be mindful that introductory physics is just one of the six required courses (three science courses, one mathematics course, one computing course, and a university orientation course) that science students take in the first semester of their university career. Secondly, upon examination of the multiple-choice questions that were collaboratively developed by students, it became clear that in general, the class did not produce high-quality MCFQs nor the expected volume of questions. Although about 60% students participated in the online MCFQ-writing activities, many contributions were of poor quality. Overall, the class failed to produce a usable bank of conceptual questions on Newtonian mechanics. This resulted in the abandonment of the plan to use student-generated MCFQs in formal course evaluations. It is possible that the instructor's guidance about how to develop multiple-choice format questions was insufficient for the majority of the students in the class. Alternatively, it is possible that the in-class question-writing practice activities were not as extensive as they needed to be. Finally, because the activity was voluntary and for bonus marks only, it is possible that the students did not take it sufficiently seriously, especially as there were so many mandatory activities included in the course. Given the quality of the MCFQs produced, it is not surprising that participating in the MCFQ-writing activities produced very limited – if any – benefit in conceptual learning gains and no benefit in improving the students' attitudes toward physics.

5.3 Lessons Learned and Recommendations

Overall, the research team recommends that instructors consider the cost, time and resources needed to implement small-group MCFQ-writing activities. This new activity requires a significant amount of preparation prior to use, with inconsistent effects on student learning. In addition, guiding, monitoring and evaluating student activities on the online course discussion board, as well as providing feedback to students and maintaining a record of completion, can be very time-consuming for the instructor. Although participating in MCFQ-writing activities was associated with slightly higher gains in conceptual learning in one cohort, this effect was not consistent across time and smaller than the effect of other student characteristics. This particular intervention also did not appear to be effective in prompting students to adopt more 'expert-like' attitudes towards science and problem-solving. Interviews with students show that they did not recognize the MCFQ-writing activities as a unique experience in their learning and many did not remember the activities taking place. In the context of this course, which already incorporated substantial PI and active-learning components, students may have experienced resource overload and been unable to take advantage of all of the learning opportunities, prioritize or even distinguish between them. This intervention may have a different effect in a course with fewer student learning supports in place.

Overall, the substantial effort invested into this additional activity cannot be justified given the lack of obvious positive outcomes on student learning. We encourage instructors in large classes to consider alternative interventions to improve student learning, while also remaining mindful of the potential for resource overload among students. We also encourage instructors to consider students' preferred study methods. In our study, the majority of students reported that they preferred to study alone, and implementing additional activities that promote small-group, collaborative interaction may have negatively affected these students' participation and learning gains. These effects might be particularly pronounced for students who are less familiar with English. Although in-class, small-group activities provide instructors with an opportunity to teach effective group work practices, they should remain mindful of some students' preference to work independently.

In retrospect, the decision to have students complete some of the MCFQ-writing activities online may also have limited their success. The online course discussion board is not specifically designed to monitor and evaluate the students' contributions to peer collaborations. A more specialized system for peer collaboration may have been more appropriate, but we chose to have students continue working on the same online platform that hosts the rest of the course. Bottomley and Denny's (2011) attempts to use similar MCFQ-writing activities were more successful and used a technological system that was specifically designed for facilitating peer collaboration (PeerWise, https://peerwise.cs.auckland.ac.nz/).We certainly recommend that instructors who wish to incorporate collaborative writing activities in their courses choose their supporting technology with great care.

Since this project was initiated in 2011, there have been substantial changes in the educational technology available to instructors. Most notably, personal response systems are becoming more versatile in regard to the types of the questions that can be asked in addition to the standard multiple-choice questions. For example, "Learning Catalytics" from Pearson (https://learningcatalytics.com/) provides a wide range of question-type options beyond the multiple-choice format.

Although the new MCFQ-writing activity itself did not prove to be effective, the results of our study provided further compelling evidence in support of active learning environments in large introductory physics classes. In both the experimental and control groups of this study, the FCI results suggest that these large lecture classes achieved a relatively high level of conceptual learning and student engagement. The CLASS results indicate that students did not demonstrate the negative shift in attitudes toward physics that is typical of lecture-based introductory physics courses and even demonstrate a positive shift in selected categories. The qualitative data from students' interviews provide overwhelming support for the use of PI, supported by personal response systems, experiential learning in the laboratories, and online self-tutoring and homework.

References

Adams, W., Perkins, K., Podolefsky, N., Dubson, M., Kinfelstein, N., & Wieman, C. (2006). New Instrument for Measuring Student Beliefs about Physics and Learning Physics: The Colorado Learning Attitudes about Science Survey. *Physics Review Special Topics: Physics Education Research, 2*.

Adams, W., Perkins, K., Dubson, M., Finkelstein, N., & Wieman, C. (2004). The Design and Validation of the Colorado Learning Attitudes about Science Survey. *Proceedings from APPT Physics Education Research Conference 2004.* Sacramento, CA.

Antimirova, T., Noack, A., & Milner-Bolotin, M. (2009). The Effect of Classroom Diversity on Conceptual Learning in Physics. *Proceedings from Physics Education Research Conference 2009*. Ann Arbor, MI.

Bottomley, S., & Denny, P. (2011). A Participatory Learning Approach to Biochemistry using Student Authored and Evaluated Multiple-choice Questions. *Biochemistry and Molecular Biology Education, 39*(5), 351-361.

Brown, S., Race, P., & Smith, B. (2004). 500 Tips on Assessments: Routledge.

Cunningham, G. (2005). Assessment In The Classroom: Constructing And Interpreting Texts: Routledge.

Crouch, C., & Mazur, E. (2001). Peer Instruction: Ten Years of Experience and Results. *American Journal of Physics, 69*(9), 970-977.

Cuseo, J. (2007). The Empirical Case against Large Class Size: Adverse Effects on the Teaching, Learning, and Retention of First Year Students. *Journal of Faculty Development*, *21*(1), 5-21.

De Jong, T. (2009). Cognitive Load Theory, Educational Research, and Instructional Design: Some Food for Thought. *Instructional Science*, *38*, 105-134.

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved Learning in a Large-Enrollment Physics Class. *Science, 332*(6031), 862-864.

Fagen, A., Crouch, C., & Mazur, E. (2002). Peer Instruction: Results from a Range of Classrooms. *The Physics Teacher, 40*(1), 206-209.

Fensham, P. Gunstone, R., & White, T. (1994). The Content of Science: A Constructivist Approach to Its Teaching and Learning: Routledge.

Gray, K., Adams, W., Wieman, C., & Perkins, K. (2008). Students Know What Physicists Believe, But They Don't Agree: A Study Using the CLASS Survey. *Physics Review Special Topics: Physics Education Research, 4*, 1.

Guimond, S., & Roussel, L. (2001) Bragging About one's School Grades: Gender Stereotyping and Student's Perception of Their Abilities in Science, Mathematics and Language. *Social Psychology in Education*, *4*, 275-293.

Hake, R. (1998). Interactive-engagement Versus Traditional Methods: A Six Thousand-student Survey of Mechanics Test Data for Introductory Physics Courses. *American Journal of Physics, 66*, 64-74.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force Concept Inventory. *The Physics Teacher, 30*, 141-153.

Hitt, G., Isakovic, A., Fawwaz, O., Bawa'aneh, M., El-Kork, N., Makkiyil, S., & Qatten, I. (2014). Secondary Implementation of Interactive Engagement Teaching Techniques: Choices and Challenges in a Gulf Arab Context. *Physics Review Special Topics: Physics Education Research, 10*.

Huffman, D., & Heller, P. (1995). What Does the Force Concept Inventory Actually Measure? *The Physics Teacher*, *33*(3), 138-143.

Iaria, G., Hubball, H. (2008). Assessing Student Engagement in Small and Large Classes. *Transformative Dialogues: Teaching and Learning Journal, 2*(1), 1-8.

Kalman, C. Milner-Bolotin, M., & Antimirova, T. (2010). Comparison of the Effectiveness of Collaborative Groups and Peer Instruction in a Large Introductory Physics Course for Science Majors. *Canadian Journal of Physics*, *88*(5), 325-332.

Kelly, A. (1985). The Construction of Masculine Science. *British Journal of Sociology of Education, 6*(2), 133-154.

Mazur, E. (1997). Peer Instruction: A User's Manual. Upper Saddle River, NJ: Prentice Hall.

McCarthy, J. & Anderson, L. (2000). Active Learning Techniques Versus Traditional Teaching Styles: Two Experiments from History and Political Science. *Innovations in Higher Education, 24*, 279-294.

McDermott, L. (1990). Teacher Preparation in Physics and Other Sciences: The need for Special Science Courses for Teachers. *American Journal of Physics, 58*, 734-742.

McDermott, L., & Redish, E. (1999). Resource Letter: PER-1: Physics Education Research. *American Journal of Physics, 67*, 755-773.

McDonald, M. (2001). Systematic Assessment of Learning Outcomes: Developing Multiple-choice Exams. Sudbury, MA: Jones and Bartlett Publishers.

Milner-Bolotin, M., Antimirova, T., Noack, A., & Petrov, A. (2011). Attitudes about Science and Conceptual Physics Learning in University Introductory Physics Courses. *Physics Review Special Topics: Physics Education Research*, *7*.

Murphy, H., & Roopchand, N. (2003). Intrinsic Motivation and Self-esteem in Traditional and Mature Students at a Post-1992 University in the North-east of England. *Educational Studies*, *29*, 243-259.

Noack, A., Antimirova, T., & Milner-Bolotin, M. (2009). Student Diversity and the Persistence of Gender Effects on Conceptual Physics Learning. *Canadian Journal of Physics, 87*(12), 1269-1274.

Perkins, K., Adams, W., Finkelstein, N., Pollock, S., & Wieman, C. (2004). Correlating Student Beliefs with Student Learning using the Colorado Learning Attitudes about Science Survey. *Proceedings from Physics Education Research Conference 2004*. Sacramento, CA.

Perkins, K., Gratny, M., Adams, W., Finkelstein, N., & Wieman, C. (2005). Towards Characterizing the Relationship between Students' Self-reported Interest in and their Surveyed Beliefs about Physics. *Proceedings from Physics Education Research Conference 2005*. Salt Lake City, UT.

Pollock, S. (2004). No Single Cause: Learning Gains, Student Attitudes, and the Impacts of Multiple Effective Reforms. *Proceedings from Physics Education Research Conference 2004*. Sacramento, CA.

Redish, J., Saul, M., & Steinberg, N. (1998). Student Expectations in Introductory Physics. *American Journal of Physics, 66*, 212-224.

Savinainen, A., & Scott, P. (2002). Using Force Concept Inventory to Monitor Student Learning and to Plan Teaching. *Physics Education, 37*(1), 53-58.

Simkins, S., & Maier, M. (eds.) (2010). *Just in Time Teaching: Across the Disciplines, Across the Academy.* Sterling, VA: Stylus Publishing.

Singh, C. (2005). Impact of Peer Interaction on Conceptual Test Performance. *American Journal of Physics*, *73(*5), 446-451.

Smith, M., Wood, W., Adams, W., Wieman, C., Knight, J., Guild, N., & Su, T. (2009). Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science, 323*(5910), 122-124.

Thornton, R., Kuhl, D., Cummings, K., & Marx, J. (2009). Comparing the Force and Motion Conceptual Evaluation and the Force Concept Inventory. *Physical Review Special Topics – Physics Education Research*, *5*(1).

Thornton, R., & Sokoloff, D. (1998). Assessing Student Learning of Newton's Laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics, 66*, 338-352.

Winter, D., Lemons, P., Bookman, J., & Hoese, W. (2001). Novice Instructors and Student Centered Instruction: Identifying and Addressing Obstacles to Learning in the College Science Laboratory. *Journal of Scholarship of Teaching and Learning, 2*, 15-42.

Higher Education Quality Council of Ontario

An agency of the Government of Ontario