Higher Education Quality Council of Ontario
An agency of the Government of Ontario

Teaching Scientific Inquiry: Inquiry-based training for biology graduate teaching assistants improves undergraduate learning outcomes

P. W. Hughes, Carleton 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:

Hughes, P. W. (2014). *Teaching Scientific Inquiry: Inquiry-based training for biology graduate teaching assistants improves undergraduate learning outcomes.* Toronto: Higher Education Quality Council of Ontario.

Acknowledgements

I would like to thank Michelle Ellefson for her help designing this intervention and analyzing the results. I would also like to thank Andrew Simons and Susan Aitken for their help and encouragement, as well as all graduate teaching assistant participants and undergraduate survey respondents, without whom this work would not have been possible.

Table of Contents

List of Tables

List of Figures

1 Executive Summary

1.1 Context

Higher education institutions are gradually increasing the teaching load borne by graduate teaching assistants (GTAs), a trend that is particularly severe in the STEM disciplines (science, technology, engineering and mathematics). As GTAs bear more responsibility for teaching STEM undergraduate students, student success will require investment in pedagogical training for GTAs, and improving GTA teaching ability using GTA training sessions will become more critical.

1.2 Research Question

The goal of this study was to examine whether training biology lab GTAs in inquiry-based methods improves teaching effectiveness to a greater degree than does the existing “best practice” model. Undergraduate biology courses, like those in other scientific disciplines, usually involve the use of structured-inquiry laboratory activities to teach undergraduates how to apply scientific inquiry skills. This pedagogical approach is rooted in inquiry learning (Brown, 2010; Wilke & Straits, 2005). The research question was: “*Does providing graduate teaching assistants with pedagogical training in inquiry-based methods improve their teaching effectiveness in undergraduate science laboratory activities (relative to ‘best practices’ training)?*” I hypothesized that offering an inquiry-based training regimen, consisting of a theoretical introduction to the process of inquiry-based learning and training in facilitating and assessing student-centered inquiry, would improve GTA laboratory teaching relative to standard GTA training methods, which mainly focus on offering “best practice” tips and strategies.

1.3 Research Methods

I used a quasi-randomized control trial to compare the two types of GTA training. Fifty-two GTAs were recruited and randomly assigned to one of two groups. Both groups completed a two-seminar training regimen, with the first session focusing on lab teaching skills and the second on marking and evaluation skills. The inquiry-based learning group (IBLG) received information in an “inquiry” context; the inquiry-learning process was explained to GTAs and their ability to teach inquiry was assessed. The control group (CG) received instruction in instructional “best practices” (the most common form of GTA training), which introduces GTAs to a variety of task-oriented tips gathered from training materials and educational research studies designed to make lab instruction more effective. I hypothesized that inquiry-based learning methods would improve measures of teaching effectiveness over and above the standard best practices training.

Teaching effectiveness was assessed using three measures: (1) the 32-item, 9-factor Student Evaluation of Educational Quality (SEEQ) inventory developed by Herbert Marsh (Marsh & Bailey, 1993; Marsh, 1982, 1987); (2) a 6-item Cognitive Learning Evaluation (CLE) instrument, developed for this study to assess scientific inquiry skills but derived from the Revised Bloom’s Taxonomy of Learning (Krathwohl, 2002); and (3) undergraduate students’ final grades, standardized by course. The SEEQ inventory and CLE were coded as online surveys and given to both the GTA participants as well as the students enrolled in introductory biology labs run by the GTAs. The surveys collected responses from all 52 GTAs and from 603 undergraduates. Undergraduate student grades were collected from students themselves and course grade descriptive statistics were collected from course instructors (lab coordinators and professors).

1.4 Findings

The main prediction of the hypothesis was supported in that all three measures recorded higher teaching effectiveness scores for GTAs enrolled in the inquiry-based learning training group. Particular SEEQ and CLE factors/items showed strong differences and undergraduate responses showed greater differences than GTA self-assessments. Students scored IBL-group GTAs higher in six of nine SEEQ factors (including learning value, instructor enthusiasm, organization, rapport, assignment feedback and overall instructional quality) and in four of six CLE items, including three of the four highest-level cognitive skills, which corresponded to the deeper learning objectives of the Revised Bloom's Taxonomy (Krathwohl, 2002). Undergraduate students with GTAs who received inquiry-based learning training had significantly higher mean course grades than students with GTAs who received the control group training. Overall, inquiry-based learning training for GTAs was associated with higher teaching effectiveness and improved undergraduate students' scientific inquiry skills.

2 Previous Research and Experimental Overview

2.1 Graduate Teaching Assistants in Ontario

In Ontario's universities, as at others in North America, graduate teaching assistants (GTAs) are part-time educators employed to provide teaching support to undergraduate course instructors. Such support takes a variety of forms: some GTAs provide one-on-one supervision of undergraduates, some mark tests and assignments, others provide supplementary lectures or tutorials for students, and some supervise students during lab activities. The teaching role of a GTA is typically very flexible and GTA teaching often accounts for a substantial proportion of the contact hours undergraduates have with subject specialists from their home department (Lowman & Mathie, 1993; Luft et al., 2004).

The desire to improve North American university science lab education presents some fundamental challenges. While GTA lab instructors are required to teach at a fairly high level for undergraduate lab instruction to be successful, GTAs (like many faculty members) are not trained teachers, and providing pedagogical training on the scale required can be time-consuming and expensive. As university enrolment has increased, so too has the use of GTA teaching labour grown in North American higher education institutions (Creech & Sweeder, 2012). GTAs have shouldered much of the additional teaching load, presumably because they represent a smaller per-unit labour cost than course instructors or faculty. This pattern – of accommodating increased class sizes by hiring more GTAs (and sessional faculty on part-time contracts) to help a fixed number of full-time faculty – has increased interest in providing pedagogical training to GTAs, since the quality of undergraduate learning has been found to correlate highly with perceived GTA teaching effectiveness (Bond-Robinson & Rodriques, 2006; Hardré, 2005).

How GTAs should be trained is a topic of much debate, but one consensus view seems to be that both the quantity and quality of GTA training needs to increase (Black & Bonwell, 1991; Buerkel-Rothfuss & Gray, 1990; Hardré, 2005; Lowman & Mathie, 1993; Meyers & Prieto, 2000; Nyquist, Abbott & Wulff, 1989; Park, 2004; L. Prieto, 2002; Rushin et al., 1997; Verleger & Velasquez, 2007; Volkmann & Zgagacz, 2004). Some studies have demonstrated fairly conclusively that if pedagogical training is provided, it measurably improves GTA teaching effectiveness (Black & Bonwell, 1991; Hardré & Burris, 2010; Marbach-Ad et al., 2012; Shannon, Twale & Moore, 1998). Additionally, although many university professors lack formal pedagogical training, many universities offer training programs designed to provide them with rudimentary teaching skills. Until recently, many GTAs were trained using the same programs, or using GTA-specific training programs modeled on this approach (e.g., Rushin et al., 1997).

In 2012, 19 of 23 universities in Ontario offered some form of GTA training (Miles & Polovina-Vukovic, 2012). These training programs took a variety of different forms, including full- or half-day orientation sessions, teaching workshops, in-person microteaching sessions, observational learning and webinars. Carleton University (the site of this experiment) offers five hours of paid pedagogical training per year for all full-time GTAs. Other factors vary among the types of GTA training offered by different universities, including whether training: (1) is mandatory or not; (2) is only for new GTAs, experienced GTAs, or some mixture of both; (3) is paid or not; (4) is course-, department- or faculty-specific; (5) primarily covers course content or teaching skills; and (6) uses evidence-based teaching methods or not.

2.2 GTAs in STEM Disciplines

Teaching undergraduates to link content knowledge to conceptual understanding is a common objective of many GTA teachers in STEM (science, technology, engineering and mathematics) disciplines. These pedagogical tasks are difficult even for seasoned instructors, whether or not they have pedagogical training. Specific challenges faced by STEM GTAs include the facts that: STEM content knowledge is complicated and non-intuitive; scientific learning involves belief change (Norton et al., 2005; Tompkins & Dimiduck, 2011); scientific inquiry skills are higher-order cognitive functions (Wilke & Straits, 2005); and scientific principles can be challenging to formalize and discuss (Lin et al., 2013; Linn et al., 2006). Effectively teaching scientific inquiry skills is difficult without some degree of pedagogical training. As GTAs shoulder a proportionately greater share of undergraduate teaching in Ontario universities than faculty members, providing pedagogical training to GTAs, and STEM GTAs in particular, will only become more important.

Laboratory (“lab”) activities, in introductory STEM courses are designed to teach undergraduates both discipline-specific content and scientific inquiry skills. Lab activities are also among the most difficult activities to teach (Ertepinar & Geban, 1996; Vale et al., 2012; Wallace et al., 2003). This is so precisely because labs teach more than mere content; unlike in lecture halls, where students often passively receive information, in labs undergraduate students engage in structured- or open-inquiry activities to learn to “think like a scientist” – that is, to engage in scientific inquiry and employ the scientific method. Training teachers in how to teach scientific inquiry well is therefore a crucial determinant of overall undergraduate success in science programs (Luft et al., 2004; Roehrig et al., 2003).

Like research universities throughout North America, most Ontario universities employ lab activities intensively in their introductory science courses. Because lab activities require a high degree of supervision and a correspondingly high student-to-GTA ratio, the success of this model depends largely on the availability and quality of GTA labour. Labs, as active learning activities, require higher instructor-to-student ratios than traditional lectures or other passive learning activities, such as online lectures and assignment-based learning. For instance, Carleton’s introductory biology labs usually have one lab demonstrator per ten students. The most cost-effective way for universities to achieve these high instructor-to-student ratios is through hiring masses of GTAs as lab demonstrators. Coupled with the long-term trend in increasing university attendance in Ontario (Finnie & Pavlic, 2013), a trend which shows no sign of abating, as well as the fact that GTA turnover is high (since graduate students finish their academic program within two to five years), the pool of GTAs requiring training in lab instruction techniques is continually growing.

The high number of GTAs required for lab courses, coupled with the relative difficulty of teaching structured-inquiry classes, combine to present GTA trainers with three different challenges: firstly, lab activities are among the most difficult to teach; secondly, demand for GTA labour is continuous, since older, experienced GTAs are constantly being replaced with newer, inexperienced GTAs; and, thirdly, GTAs’ training – if compulsory or paid at all – is usually limited to a few training hours (sometimes even a single orientation session).

2.3 Inquiry-Based Learning in STEM Disciplines: Undergraduate Labs

In postsecondary STEM courses, interest in learner-centered pedagogies, particularly inquiry-based learning (IBL), has increased since the year 2000 (Bao et al., 2009). This is because constructivist learning theories such as IBL emphasize the development of higher-order cognitive skills that have been identified as critical for the development of scientific thinking (i.e., theory explication and experimental protocol construction). As structured-inquiry activities, laboratory classes differ notably from other undergraduate teaching formats in that successful learning outcomes depend not only on teachers’ ability to relate scientific content (i.e., what the chemical properties of DNA are) but also on their ability to relate higher-order cognitive functions, such as applying the scientific method and interpreting data (i.e., what the results of a particular hypothesis would look like given the known chemical properties of DNA). The growing use of inquiry-based teaching methods in university-level courses has prompted calls for pedagogical training specific to inquiry-based instruction.

In the STEM disciplines, inquiry-based learning is used primarily to develop general scientific inquiry skills. Although there are studies supporting their use as pedagogical tools for this purpose, whether or not inquiry-oriented methods are generally more successful than instructor-centered methods remains an open academic question. Some critics maintain that inquiry methods are generally ineffective in teaching inquiry skills since the lack of guidance reifies existing errors (e.g., Kirschner, Sweller & Clark, 2010), while many other studies show improvement in higher-order cognitive skills, both in the K-12 science classroom (Hanauer et al., 2006) and in postsecondary science classrooms (Luckie, Maleszewski, Loznak & Krha, 2004; Price, 2012; Volkmann & Zgagacz, 2004). In response to criticisms of inquiry-based methods related to lack of guidance, many authors have developed undergraduate lab activities designed specifically to target the development of inquiry skills through the use of guided or process inquiry (Apedoe, Walker & Reeves, 2006; Goldey et al., 2012; Iler & Justice, 2012; Schoffstall & Gaddis, 2007; Uno, 2009; Wallace et al., 2003). Such methods are designed to provide "just enough" instruction to scaffold learning effectively. That is, inquiry is guided, so as not to let students aimlessly drift, but is not directed, in order to provide genuine student-centered learning.

Because they play an important role in facilitating learning in undergraduate labs, several authors have suggested that GTAs would benefit from inquiry-based training (Bohrer, Stegenga & Ferrier, 2007; French & Russell, 2002). Mainly, this was suggested because GTAs do not have the requisite skills to facilitate inquiry activities effectively, a problem which may be related to general dissatisfaction with GTA teaching in undergraduate labs (Roehrig et al., 2003).

2.4 Undergraduate and GTA Views of GTA Teaching Effectiveness

The degree to which undergraduates and GTAs agree on the constituents of effective GTA teaching is not well studied. One study found that when asked what important qualities "good" GTAs possessed, undergraduates and GTAs had similar responses (Herrington & Nakhleh, 2003). In another, the author found that undergraduates and course instructors (i.e., faculty) were relatively harsh critics of GTAs across academic disciplines (Rodriques & Bond-Robinson, 2006).

Although GTAs may have an acceptable understanding of source material, they may lack the ability to determine whether or not their teaching is effective. In a recent study (Kendall & Schussler, 2012), undergraduates identified several areas in which GTA teachers were less effective than experienced course professors; hesitancy, unresponsiveness and confusion were highlighted as three common attributes of GTA teachers. Developing GTA self-efficacy has been a major goal of most GTA training programs (Prieto & Altmaier, 1994), and increasing confidence and student rapport through group GTA training have been shown to be effective ways of improving GTA teaching scores, in the eyes of both undergraduates and the GTAs themselves (Prieto & Meyers, 1999; Tompkins & Dimiduck, 2011).

2.5 Experimental Overview

In order to test the hypothesis that offering GTAs pedagogical training grounded in inquiry-based learning (IBL) principles would improve GTA teaching effectiveness when leading inquiry-oriented activities (such as labs), I used a quasi-randomized control trial (RCT) with two training regimens offered to two different, quasi-randomly assigned groups of GTAs. I then measured teaching effectiveness (using teaching assessments scored by undergraduate students, course grades and GTA self-assessments) to determine the relative effectiveness of the IBL treatment. The idea underlying this approach is that GTAs might benefit from an explicit description of the constructivist learning theories upon which structured-inquiry activities are built. Of course, in theory any GTA training regimen might be improved simply by incorporating more material. To control for this, I compared GTA training regimens that: (1) took the same amount of training time; and (2) were as similar as possible in format to existing programs used by Ontario universities. An informal survey of the GTA training programs offered at Ontario universities indicated that there is no standard length of GTA training program in the province. The duration of GTA training programs on offer varied from brief, hour-long orientations to intensive, twenty-hour seminars (Dawson, Dimitrov, Meadows & Olsen, 2013). I therefore opted for the two-seminar format used at Carleton that offered five hours of training in total, since it was intermediate in length yet still permitted multiple training sessions to be used.

A quasi-randomized control trial design was chosen because randomized control trials are the most rigorous way of determining whether differences exist between two treatments. The experiment was "quasi-randomized" since GTAs were allocated randomly only after controlling for GTA academic program (so that both training regimens had the same proportion of PhD and MSc students).

The experiment was set up with one intervention group (i.e., IBL treatment) and a treatment control group (i.e., status quo treatment). The two treatments each received five hours of training separated into a pair of 2.5-hour seminars. For both groups, the first seminar focused on teaching skills and the second on assessment skills. In brief (although a detailed description of the two training groups follows), GTAs in the IBL training regimen received one seminar on how to implement inquiry-based teaching in biology labs and another seminar on evaluating learning outcomes from an IBL perspective. GTAs in the control training regimen received both a teaching seminar and an assessment seminar, but these were taught from a list of "best practices" compiled from GTA training material given to students at various North American universities. Both groups: (1) received only five hours of GTA training (which is typical of most GTA training programs); (2) covered the same subject material in training seminars; (3) were taught by the same instructor; and (4) were taught using the same teaching methods. The sole difference between the two treatments revolved around the theoretical grounding of the material; in the IBL group, GTAs were exposed to IBL pedagogy as a method of teaching scientific inquiry. GTAs in this group were taught how to facilitate guided- and open-inquiry activities through the use of "scaffolding" methods. In the control group, GTAs were exposed to what lab GTAs usually receive in GTA training sessions: practical, hands-on methods that are commonly used to orient new GTAs, generally related to teaching content to students quickly and effectively.

Unlike some randomized control trials, I did not include a negative control group (i.e., a group that did not receive either the IBL or status quo treatments), since including untrained GTAs would not have been ethically permissible, either to the GTAs or to the undergraduates they taught.

3. Research Questions and Predictions

Formally stated, the main research question I considered was:

> (1) *Does providing graduate teaching assistants with pedagogical training in inquiry-based methods improve their teaching effectiveness in undergraduate science laboratory activities (relative to "best practices" training)?*

For this question, I defined increased teaching effectiveness in terms of three metrics: (1) greater general instructional ability; (2) greater ability to teach scientific inquiry tasks (i.e., greater ability to teach higher-order cognitive skills); (3) and higher course grades. All comparisons are made between those GTAs who completed pedagogical training in inquiry-based methods and those GTAs who completed "best practices" training.

Each metric addresses a distinct area of emphasis for teaching undergraduate students. General instructional ability is important for a straightforward reason – general teaching skills are important for relating any type of content to students. General instructional ability is composed of both cognitive and affective skills, like organization and rapport. To measure scientific inquiry, I assessed cognitive skills from the Revised Bloom's Taxonomy of Learning (Krathwohl, 2002). Cognitive skills that represent students' depth of understanding of a principle are important because higher-order skills indicate a more developed awareness of subject material and are indicative of a greater ability to ask further questions and contextualize responses. For example, when students conduct experiments, lower-order cognitive skills govern their grasp of scientific knowledge – such as how a particular unknown chemical would behave if it was a hydrophobic compound – while higher-order cognitive skills govern their ability to ask questions using an experimental methodology – such as choosing an experiment to assess a particular feature (such as hydrophobicity) of an unknown chemical. Although GTAs are responsible for teaching undergraduates scientific content, this could happen equally well in a lecture class; the point of using the lab as a teaching method is to develop higher-order cognitive skills in undergraduate students. In addition, to the extent that undergraduate grades are the established feedback metric for undergraduate students, it is important to ask whether or not GTA training regimens affect undergraduate grades.

The main research question was therefore resolved into three research questions, each of which asks whether GTA pedagogical training affects teaching ability as measured using a different metric:

> (1a) *Does providing graduate teaching assistants with pedagogical training in inquiry-based methods improve their teaching effectiveness in undergraduate science laboratory activities (relative to "best practices" training) with respect to* ***general instructional ability****?*

> (1b) *Does providing graduate teaching assistants with pedagogical training in inquiry-based methods improve their teaching effectiveness in undergraduate science laboratory activities (relative to "best practices" training) with respect to* ***scientific inquiry?***

> (1c) *Does providing graduate teaching assistants with pedagogical training in inquiry-based methods improve their teaching effectiveness in undergraduate science laboratory activities (relative to "best practices" training) with respect to* ***undergraduate grades?***

Because GTA pedagogical training grounded in inquiry-based learning is designed to relate the fundamental purpose of lab teaching (i.e., developing scientific inquiry), I hypothesized that GTAs who would take part in the inquiry-based learning pedagogical training would score higher on all three measures of teaching effectiveness than GTAs who received standard (control) GTA training. Standard GTA pedagogical training is an important part of acculturation to the teaching model of an individual department, but it does not specifically address the teaching needs of lab GTAs.

The secondary focus of this study asks how well GTA self-assessments agree with the teaching assessments given by undergraduate students. Two of the three measures for this study are online surveys, and as these are administered to both GTAs and undergraduates, it is possible to make a direct comparison between the results for the two groups.

The secondary research question is therefore:

> (2) *To what degree do GTA self-assessments of GTA teaching effectiveness agree with undergraduate student assessments of GTA teaching, and does inquiry-based pedagogical training have an effect on the correlation between the two?*

Based on previous studies comparing instructors' and students' assessments of instructor teaching effectiveness, I hypothesized that GTAs' and undergraduates' views of GTA teaching effectiveness would be similar (Marsh & Roche, 1994, 1997), and that if inquiry-based pedagogical training improved GTA teaching effectiveness, this would be noted by both GTAs and undergraduates equally.

4. Method

4.1 Experimental Design

The experiment was conducted in the Department of Biology at Carleton University in the Fall 2012 semester. Carleton is a research-intensive university with approximately 27,000 students enrolled in more than 400 academic programs (as of the 2011-2012 academic year – see Carleton University, 2013). The Department of Biology caters to 765 students and offers 71 undergraduate courses in all areas of biology; 17 of these are large (>100 student) introductory courses at the first- and second-year level. Although no statistics on the size of these classes have been published, approximately 30.1% of first-year classes and 24.9% of second-year courses at Carleton contained more than 100 students in the 2011-2012 academic year; this is likely true for introductory biology courses as well.

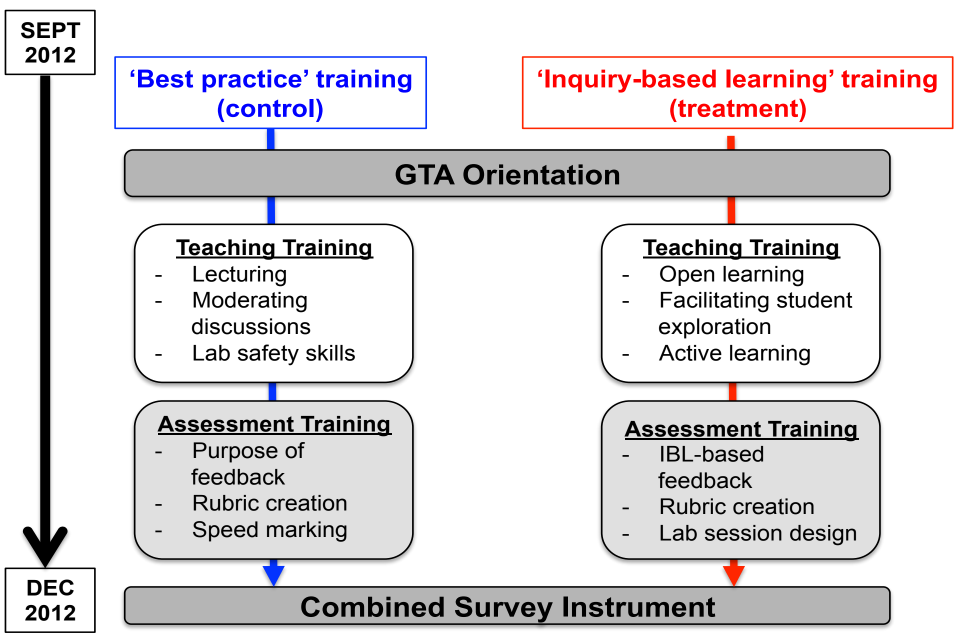
Carleton's undergraduate biology program is typical of an undergraduate science program at a North American university. Students enrolled as biology majors take a variety of introductory courses in their first and second years, usually with a large cohort of students. Most introductory courses have complementary lecture and lab components; such courses, being labour-intensive to run, are the primary source of employment for the department's roughly 100 GTAs. GTAs are typically graduate students pursuing research degrees in the Department of Biology, although rarely other students are hired as biology GTAs (e.g., chemistry graduate students or high-GPA fourth-year biology undergraduates). GTAs typically undertake five hours of pedagogical training per academic year and are responsible for 135 hours of work per semester in their assigned course.

In the Department of Biology at Carleton, most GTAs (~75%) work as lab demonstrators. Of these, approximately 80% are assigned to first- or second-year courses. GTAs report either to a lab coordinator (a full-time staff member responsible for organizing undergraduate labs) or to a lead GTA. Most GTAs are responsible for leading lab activities, teaching students lab procedures and marking student work.

For this experiment, I used a quasi-randomized control trial to compare IBL and "best practice" (control) GTA training regimens (see Figure 1). This design was chosen because it is the most robust type of comparative experiment, since randomization reduces the likelihood of systemic bias in subject allocation to treatment.

In this particular trial, allocation between the two treatment levels was randomized only after accounting for academic program (so that the same proportion of PhD and MSc students were in each level), hence "quasi-randomized" rather than completely randomized. Students were divided into two groups and allocated randomly into the two levels of the experimental variable (training regimen). Stratified random assignment according to academic program was performed because the low proportion of PhD students might have created an unbalanced design.

Figure 1: Diagram of Experimental Design



4.2 Recruitment and Research Ethics Approval

GTAs interested in obtaining pedagogical training were invited to participate in the experiment in September 2012. Prospective participants were only eligible for inclusion if they: (1) were employed by Carleton University as a GTA in an introductory biology course (i.e. a first- or second-year course without specific course prerequisites) with a lab component; (2) were a graduate student enrolled in a program of study in the Department of Biology at Carleton University; (3) consented to take five hours of GTA training in one of two experimental treatments, assigned at random; and (4) consented to self-assess their teaching effectiveness at the end of the training regimen. GTAs were allowed to count hours spent in the two training sessions of this experiment toward their five mandatory training hours. Fifty-four GTAs enrolled initially, but only 52 successfully completed the experiment. GTA participants included GTAs of both genders, all experience levels (including new GTAs), and from various subfields in the biosciences (molecular biology, evolutionary biology, ecology, etc.).

This study was presented to the Carleton University Research Ethics Board in September 2012 and was approved in October 2012. The informed consent form given to GTA survey respondents is included as Appendix A and the informed consent form given to undergraduate respondents is included as Appendix B.

4.3 Participants and Randomization

4.3.1 GTA Experimental Participants

The final distribution of GTA participants had 52 GTAs in the experiment, with 28 in the IBL treatment and 24 in the control treatment (Table 1). Sixty-three GTAs initially volunteered to participate in this experiment, but 9 GTAs did not have laboratory course teaching assignments and were thus ineligible to participate. Two GTAs, both in the control group, began the training regimen but did not complete it; they did not complete the online survey and were not included in the final results of the experiment. Proportions of GTAs in PhD/MSc academic programs were controlled before random allocation to treatment level. The proportion of male and female participants and the mean years of GTA experience were not controlled prior to randomization, but both levels ended up being similar (within approximately 10% of each other) in each training group. No qualified teachers were included in either training group; although it is common in some disciplines (particularly education) for GTAs to possess teaching qualifications that allow them to teach K-12 classes, this is uncommon in STEM disciplines.

Table 1: GTA Participant Demographic Data

		Treatment	
		Control	**IBL**
Number of GTAs		24	28
Academic Program	PhD	9 (37.5%)	11 (39.3%)
	MSc	15 (62.5%)	17 (60.7%)
Gender	Male	11 (45.8%)	10 (35.7%)
	Female	13 (54.2%)	18 (64.2%)
Mean Years as GTA		1.21	1.02
Int'l GTAs		4 (16.7%)	3 (10.7%)
Qualified Teachers		0 (0.0%)	0 (0.0%)

Of the 603 undergraduate respondents who volunteered to take the online survey, 167 students met the qualification criteria (i.e., they were a student in the laboratory section of one of the GTAs in this experiment – see Table 2). Of these, 87 respondents were in the lab section of a GTA in the control training group and 80 were in the lab section of a GTA in the IBL training group. A greater proportion of IBL GTA respondents was male and a greater proportion of IBL GTA respondents was enrolled as students taking a major in the Faculty of Science. There were no other statistically significant differences between the respondents for the two training groups.

4.3.2 Undergraduate Respondents

Table 2: Undergraduate Respondent Demographic Data

		Treatment	
		Control	**IBL**
Number		87	80
Major	Science	77.0%	87.8%
	Non-Science	23.0%	12.2%
Gender	Male	60.9%	79.3%
	Female	39.1%	20.7%
Number of Biology Courses Previously Taken		1.44	1.19

4.4 GTA Training Groups

In order to eliminate possible confounding effects of factors other than course content, I attempted to maintain a high degree of similarity between the pairs of seminars. They were held on different days of the same week, in the same seminar room, were blinded (for instance, GTA participants knew that they were attending session "2A" but were not made aware of the specific differences between their session and session "1A") and both were taught by me. The assessment seminar was held two weeks after the teaching seminar.

I attempted to minimize the possible confounding effect of teaching method. A variety of methods were used, including both instructor-centred techniques (such as lectures) as well as student-led techniques (such as peer-evaluated microteaching and small group brainstorming). As much as was possible, I used roughly the same mixture of teaching techniques in both treatments.

4.4.1 Inquiry-Based Learning GTA Training Group

GTAs in the inquiry-based learning (IBL) training group treatment were given exposure to the rudiments of constructivist learning theory and were trained to apply these principles while teaching biology labs. The main goals of the teaching seminar for the IBL GTA training group were: (1) to teach GTAs how inquiry-based practices teach students to reason independently and apply the scientific method; and (2) to teach GTAs how to facilitate structured inquiry and open-ended learning as lab activities.

The format of the teaching seminar for the IBL group was as follows:

Table 3: Teaching Seminar Activity Chart for the Inquiry-Based Learning Training Group

Item	Time Allocated	Activity	Desired Learning Outcome	Teaching Method
1	30 min	GTA General Orientation	- understanding the role of GTAs	Instructor lecture (PowerPoint)
2	15 min	Brainstorming Activity: Laboratory Learning Outcomes	- identification of determinants of lab learning success	Student-led small groups, answers shared with class and discussed
3	45 min	Introduction to Inquiry-Based Learning	- understanding of the Revised Bloom's Taxonomy of Learning - how scientific inquiry is taught using inquiry - understanding of facilitation as a teaching method	Instructor lecture (PowerPoint)
4	30 min	Facilitation Situation Activity – Sample Lab Scenarios	- understanding of how to apply facilitation techniques in teaching situations	Large-group activity
5	15 min	Troubleshooting Inquiry-Learning Activities	- how to handle unexpected problems in inquiry-based lab activities	Instructor lecture (PowerPoint)
6	15 min	Questions	- clear up unresolved questions	Student-led question and answer session

There were three main goals for the assessment training seminar for the IBL training group: (1) to teach GTAs the importance of informative and equitable assessment; (2) to explain how scientific inquiry is assessed (as opposed to how knowledge/content understanding is assessed); and (3) to practice designing and applying rubrics to inquiry tasks.

This training seminar began with a PowerPoint lecture, taught by the instructor, on the history and purpose of assessment. Emphasis was placed on the value of feedback in the learning process. This was followed by a brief brainstorming session that attempted to match appropriate methods of feedback for various laboratory-oriented learning tasks (e.g., how should feedback be given on cladograms?); this session was designed to encourage GTAs to think about the purpose of each learning activity and how errors should be addressed to best achieve the desired learning outcome.

An overview of the difficulty of assessing scientific inquiry tasks followed. The difference between knowledge recall and higher-order cognitive skills (connected to Bloom's Taxonomy) was reiterated to GTAs, along with the value of lab activities in relating inquiry skills to undergraduate students. GTAs were given examples of how to scaffold learning activities to encourage undergraduate students to ask and answer scientific questions, as well as how to assess the level of inquiry taking place.

The next two tasks were combined into a single exercise, where students evaluated, developed and applied assessment rubrics. Rubric development was introduced in an interactive question-and-answer session. Next, GTAs were given a lab assignment taken from an actual introductory biology lesson, as well as three sample lab reports from that assignment. GTAs collaborated to construct a 20-point assessment rubric (considering both content knowledge and scientific inquiry skills), then split into pairs to apply it to the sample reports. The resulting assessments were shared with the whole group and instructor feedback was provided to correct errors in rubric application. For this rubric development activity, emphasis was placed on rewarding inquiry skills and thinking about how feedback should be structured in order to highlight areas of improvement and growth in student work with respect to scientific inquiry. The last part of the assessment seminar was an open question-and-answer session that allowed students to resolve difficulties they had with assessment development and application.

The format of the assessment seminar for the IBL group was as follows:

Table 4: Assessment Seminar Activity Chart for Inquiry-Based Learning Group

Item	Time Allocated	Activity	Desired Learning Outcome	Teaching Method
1	30 min	History and Purpose of Assessment	- understanding the role of assessment in the learning process	Instructor lecture (PowerPoint)
2	15 min	Brainstorming Activity: Laboratory Assessment Methods	- identification of: (1) purpose of assessment in labs; (2) criteria determining successful assessment methods for labs	Student-led small groups, answers shared with class and discussed
3	30 min	How to Assess Scientific Inquiry	- understanding of the distinction between lower- and higher-order cognitive skills - understanding how to use assessment tolls to give feedback on cognitive skill development	Instructor lecture (PowerPoint)
4	15 min	Assessment Rubric Examples	- how rubrics are interpreted	Student-led question and answer
5	45 min	Small Group Rubric Design	- how rubrics are interpreted - understanding rubric inclusion criteria - understanding how feedback can assess different cognitive skills	Student-led small groups, answers shared with class and discussed
6	15 min	Questions	- clear up unresolved questions	Student-led question and answer

4.4.2 Control GTA Training Group

The control GTA training group was designed to teach “best practices” to GTAs to enable them to more efficiently and equitably teach content to undergraduates in their lab section. Scientific inquiry was not discussed and the emphasis was on content teaching rather than developing inquiry skills. “Best practices” were taken from existing training materials, both from university teaching and learning centres and from the educational research literature. The control training regimen included a “teaching” seminar on best teaching practices and an “assessment” seminar on fast and equitable grading as well as providing helpful assignment feedback.

There were two main goals for the control group teaching seminar: (1) to explain the role of GTAs as laboratory teachers to new GTAs; and (2) to identify strategies that facilitated effective lab teaching and lab management. The orientation portion of this seminar was the same as for the IBL training group assessment seminar, but the second half focused on best practices rather than explaining how to teach scientific inquiry. The general focus of this group was on teaching content knowledge rather than higher-order cognitive skills.

As for the IBL group version, this training seminar began with a general GTA orientation via PowerPoint lecture. Again, a brief brainstorming session followed, where the focus was on identifying important laboratory activity learning outcomes.

What followed next was a list of teaching best practices gathered from pre-existing GTA training materials, including journal articles, pamphlets, PowerPoint presentations and websites (mostly from teaching and learning centres at research universities). Although there was a relatively wide variety of best practices from which to choose, many involved giving GTAs advice on how to lecture and deal with student questions. Nearly all focused on honing a teaching message to convey content knowledge to undergraduate students effectively. A few others focused on providing useful demonstrations to undergraduates during lab activities.

This was followed by a mini-microteaching activity where GTAs “taught” unfamiliar techniques to one another using the tips they had just learned. GTAs formed pairs, alternating teaching and providing feedback to one another. The “teacher” GTA was to instruct the “student” as clearly as possible, and feedback was given afterward. This activity was repeated three times for each GTA in each pair. The activity was followed by a large group discussion of which best practices were most useful to GTAs.

The last part of this teaching seminar was an open question-and-answer session. It was designed to allow students to ask questions about best practices for teaching. The format of the teaching seminar for the control group was as follows:

Table 5: Teaching Seminar Activity Chart for Control Group

Item	Time Allocated	Activity	Desired Learning Outcome	Teaching Method
1	30 min	GTA General Orientation	- understanding the role of GTAs	Instructor lecture (PowerPoint)
2	15 min	Brainstorming Activity: Laboratory Learning Outcomes	- identification of determinants of lab learning success	Student-led small groups, answers shared with class and discussed
3	45 min	Lab Teaching Best Practices	- understanding of evidence-based methods for teaching effectively - understanding specific teaching skills, learning styles and learning outcomes	Instructor lecture (PowerPoint)
4	30 min	Mini-Microteaching Activity – Sample Lab Scenarios	- understanding of how to apply best practice teaching techniques in lab activity situations	Microteaching, large-group activity
5	15 min	Troubleshooting Inquiry-Learning Activities	- how to handle unexpected problems in lab activities	Instructor lecture (PowerPoint)
6	15 min	Questions	- clear up unresolved questions	Student-led question and answer

There were two main goals for the control group assessment seminar: (1) to teach GTAs the importance of informative and equitable assessment; and (2) to identify strategies that facilitated quick, efficient and fair grading of student work. The first portion of this seminar was the same as for the IBL training group assessment seminar, but the second half was different. The focus of this group was on assessing taught content knowledge rather than higher-order cognitive skills.

The format of the assessment seminar for the control group was as follows:

Table 6: Assessment Seminar Activity Chart for Control Group

Item	Time Allocated	Activity	Desired Learning Outcome	Teaching Method
1	30 min	History and Purpose of Assessment	- understanding of role of assessment in the learning process	Lecture (PowerPoint)
2	15 min	Brainstorming Activity: Laboratory Assessment Methods	- GTA identification of purpose of assessment in labs; criteria determining successful assessment methods for labs	Student-led small groups, answers shared with class and discussed
3	30 min	Grading Best Practices	- understanding of evidence-based methods for grading effectively - understanding specific grading requirements for in-lab and lab report activities	Instructor Lecture (PowerPoint)
4	45 min	Small Group Rubric Design	- how rubrics are interpreted - understanding rubric inclusion criteria - understanding how feedback is different in-lab versus on lab reports	Student-led small groups, answers shared with class and discussed
5	15 min	How to Grade Equitably and Quickly	- how to maintain grading consistency - how to grade lab feedback quickly	Instructor Lecture (PowerPoint)
6	15 min	Questions	- clear up unresolved questions	Student-led question and answer

This training seminar began with an instructor-centred PowerPoint lecture on the history and purpose of assessment, just as for the IBL training group assessment seminar. Again, a brief brainstorming session followed, where the focus was on matching appropriate methods of feedback for various laboratory-oriented learning tasks.

What followed next was a list of assessment best practices gathered from GTA training materials, including journal articles, pamphlets, PowerPoint presentations and websites (again taken from teaching and learning centres at research universities). Most tips involved teaching GTAs to grade papers fairly and quickly (e.g., instead of marking one paper then moving on to the next, mark Question 1 for all papers, then move to Question 2). In general, recommendations on providing feedback were less common but were not totally absent. Many best practices advocated using a shorthand system to make feedback faster and clearer.

A rubric development and application exercise followed and was similar to the one used in the IBL training group assessment seminar. However, instead of designing a rubric to address scientific inquiry, GTAs were asked just to develop rubrics that addressed content knowledge. They applied rubrics in pairs and discussed the results as a class just as in the IBL training group.

The last part of the assessment seminar was again an open question-and-answer session. It was designed to allow students to ask questions about best practices for assessment and to recognize when content was successfully learned.

4.5 Measures

In order to test GTA teaching effectiveness, I opted to use validated multiple-choice question inventories due to their reliability, consistency and ease of data collection. Demographic questions and lab section identifiers were combined with the student evaluation of educational quality (SEEQ) inventory, a custom six-question cognitive learning evaluation (CLE) instrument, and a grade self-report into a single questionnaire. In my view, the advantages of online survey data collection, including the potential for automatic data coding and wide dissemination, outweighed the main disadvantage, which was a low response rate.

I also chose to evaluate undergraduate student grades since grades are the most generally understood formal measurements of learning achievement used to evaluate undergraduate performance in a course. I adjusted these grades to standardize them for course mean grade. In the following sections I explain in greater detail the SEEQ inventory, CLE instrument and course grade measurements that I used in this experiment.

4.5.1 SEEQ Inventory: General Instructional Ability

The student evaluation of educational quality (SEEQ) inventory is a nine-factor, 32-item validated survey typically used to evaluate course instructors at North American universities. The original SEEQ inventory (Marsh, 1982) was written to resolve general teaching ability into multiple factors. Nine factors are measured using the SEEQ survey: (1) learning/academic value; (2) instructor enthusiasm; (3) organization; (4) group interaction; (5) individual rapport; (6) breadth of coverage; (7) examination and grading; (8) assignments; and (9) overall instructional ability (Table 7).

I used the first eight factors from Marsh's original SEEQ inventory to create two SEEQ instruments (one for GTAs and another for undergraduates). I changed factor 9 from its original (1982) version. Marsh's original version of the SEEQ inventory did not use "overall instructional ability" but rather "course difficulty" as the ninth SEEQ factor (Marsh, 1982; Marsh & Bailey, 1993). Since course instructors typically select course content and assessment materials, GTAs are not generally responsible for the "difficulty" of a course. As such, I opted to implement a version of SEEQ developed by the University of Saskatchewan, which includes "overall instructional ability" as factor 9. After these modifications, two SEEQ instruments were created: one was used by GTAs for self-assessment and the other by undergraduates to rate their lab GTA. These instruments contained identical assessments. They only differed in the language used (e.g., items on the GTA self-assessment read "rate your teaching ability", whereas the same items on the undergraduate assessment read "rate your GTA's teaching ability").

All SEEQ factors were scored on a five-point Likert scale ("strongly disagree" = 1; "disagree" = 2; "neutral"= 3; "agree"= 4; and "strongly agree"= 5). Each factor has one to four items, all positively worded. The mean score for these items represented the assessment score for that factor. All respondents were entitled to answer "Refuse/Not Applicable/Don't Know" for any item. That question was then given a null value and excluded from further analyses.

Table 7: SEEQ Factors: Measures of Overall Teaching Effectiveness

SEEQ Factor	Items	Description	Sample Item
(1) Learning/ Academic Value	4	Refers to the feeling of achievement and academic success that students obtain from participation in a course. High instructor scores indicate that the instructor is successfully imparting useful information to students, and helping them feel that what they have learned is worthwhile and challenging.	"Through my GTA, I have learned and understood the subject materials in this course."
(2) Instructor Enthusiasm	4	Refers to the instructor's ability to created attentiveness and interest in the educational material on behalf of the student. High instructor scores indicate that the instructor is creating engagement through dynamic presentation, and relating course material in a way that evokes interest.	"My GTA was energized and dynamic in conducting the course."
(3) Organization	4	Refers to the structure and transparency of the instructor's explanation of subject matter. High instructor scores indicate that the instructor is relating information clearly and precisely, in a way that is easy for students to understand.	"GTA explanations were clear."
(4) Group Interaction	4	Refers to the ability to foster academically useful social interactions within the classroom. High instructor scores indicate that the instructor is encouraging group work in a positive way, and is motivating students to share knowledge effectively.	"Students were invited to express their own ideas and/or question the GTA."
(5) Individual Rapport	4	Refers to the capacity to engage personally with individual learners and provide academically significant help and encouragement. High instructor scores indicate that the instructor is able to relate to students on a personal level and provide meaningful guidance.	"My GTA had a genuine interest in individual students."
(6) Breadth of Coverage	4	Refers to the ability to explain and compare alternative ideas, theories and techniques in a way that highlights essential features. High instructor scores indicate that the instructor is able to relate knowledge to students effectively through contrasting specific ideas.	"My GTA contrasted the implications of various theories."
(7) Examination and Grading	3	Refers to the ability to provide fair and useful evaluative feedback. High instructor scores indicate that the instructor equitably assesses student work and provides meaningful correction to students.	"My GTA's feedback on examinations/graded materials was valuable"
(8) Assignments	2	Refers to the ability to create or use assignments to relate material to students. High instructor scores indicate that the instructor is capable of designing or implementing assessments in such a way that new subject matter is taught or that errors are corrected.	"Readings/texts/references suggested by my GTA were valuable"
(9) Overall Instructional Ability	2	A general assessment of teaching effectiveness.	"Overall, my GTA was a good teacher"

4.5.2 CLE Instrument: Scientific Inquiry

The cognitive learning assessment (CLE) instrument is a short six-item questionnaire evaluating individual learning outcomes. I created six items based directly on the Revised Bloom's Taxonomy (Krathwohl, 2002) in order to assess the depth of student learning outcomes. I formatted the questions as in the SEEQ inventory: respondents were asked directly if the GTA provided adequate instruction according to the six cognitive skills in Bloom's cognitive domain and were shown concrete examples of each skill. The six skills of the cognitive domain are, in hierarchical order of increasing learning depth: (1) knowledge; (2) comprehension; (3) problem-solving; (4) conceptual-analytic; (5) planning; and (6) evaluation (Table 8).

As for the SEEQ, CLE items were scored on a five-point Likert scale (“strongly disagree” = 1; “disagree” = 2; “neutral”= 3; “agree”= 4; and “strongly agree”= 5). All respondents were entitled to answer “Refuse/Not Applicable/Don’t Know” to any item. That question was then given a null value and excluded from further analyses.

Table 8: CLE Items: Cognitive Learning Skills and Examples

CLE Item	Abstract Example	Concrete Example	Item
(1) Knowledge	Recalling information	Learning exact molecular weights from the periodic table	“My GTA helped me learn knowledge skills.”
(2) Comprehension	Comparing or contrasting two ideas	Learning to restate a word problem using equations	“My GTA helped me learn comprehension skills.”
(3) Application/ Problem-solving	Applying knowledge to find a solution to a specific question	Learning to select the appropriate statistical test for an analysis	“My GTA helped me learn problem-solving skills.”
(4) Conceptual Analysis	Determining causes and identifying relationships	Learning to troubleshoot a lab protocol	“My GTA helped me learn planning skills.”
(5) Planning and Synthesis	Creating a strategy by using ideas in a new way	Learning to design a new lab protocol using first principles	“My GTA helped me learn planning skills.”
(6) Evaluation	Using critical reasoning to make specific judgments about ideas	Learning to identify the most relevant theoretical approach to design a set of experiments	“My GTA helped me learn evaluation skills.”

4.5.3 Undergraduate Course Grades

Undergraduate students self-reported final course grades. Final grades reflected contributions from both lab and lecture components, although GTAs graded only lab assignments. Grading frameworks differed amongst courses, with the lab component accounting for 25% to 50% of the final grade. Final exams, which were entirely separate from the lab component, were the largest single other contributor to final grades, ranging from 30% to 75% of the final grade. Exams were typically multiple-choice, objectively graded questions (i.e., by Scantron). The objective nature of the final course grades should mitigate the likelihood that GTA grading standards differed between training groups (influenced by a Hawthorne effect[1]). Although it would have been preferable to obtain lab-only grades, this was not possible as students are not privy to them prior to receiving a final mark. Grades were self-reported since institutional data were not available.

Reported final grades fell on the 12-point GPA (grade point average) scale used by Carleton University. Using this scale, student work is graded by percentile score (0-100%), which is then linked to a corresponding letter and GPA grade (Table 9).

[1] The Hawthorne effect is a form of statistical bias during intervention-based experiments where participants’ beliefs about the efficacy of an intervention influences their behaviour and thereby influences the recorded effectiveness of the intervention.

Table 9: Grading and GPA Scale for Carleton University

Percentage	Letter Grade	GPA
90-100	A+	12
85-89.9	A	11
80-84.9	A-	10
77-79.9	B+	9
73-76.9	B	8
70-72.9	B-	7
67-69.9	C+	6
63-66.9	C	5
60-62.9	C-	4
57-59.9	D+	3
53-56.9	D	2
50-52.9	D-	1
0-49	F	0

Course means (mean GPA values) were collected from course professors and grades were standardized by subtracting course mean GPA scores (calculated by averaging undergraduate self-reported grades per course) from individual GTA GPA scores (the mean individual GPA score for the lab group of a given GTA). The resulting standardized grades ranged from 3.7 to -3.1 and represented how much better (i.e., positive values) or worse (i.e., negative values) a given GTA lab group mean grade was than the average grade for any group in that course (in terms of GPA). Standardized grades were used instead of raw grades to account for the variation between courses, professors and lab material.

4.6 Statistical Analyses

A variety of statistical methods were used, since the various instruments yielded different kinds of data. A significant portion of the raw data required post-collection processing or recoding (e.g., coding Likert responses on a scale of 1 to 5) prior to statistical analysis. To perform individual statistical analyses, I separated GTA self-assessments from undergraduate assessments for both the SEEQ and CLE instruments. Since both the SEEQ and CLE instruments contained multiple (possibly related) factors and items, I opted for a multivariate analysis of each to account for the possibility that the factors/items measured similar (or correlated) measures.

For the SEEQ instrument, raw survey data were resolved into the nine SEEQ factors by finding the mean Likert response values for all items (e.g., if a given SEEQ factor had three survey items, the average response across these items was found, and this value was the total score for that SEEQ factor for that respondent), then subjected to statistical analysis. Mean scores for all nine factors were analyzed using a factorial multivariate analysis of covariance (MANCOVA), where experimental treatment (IBL or control groups) and academic program (master’s or PhD) were included as independent variables, along with previous GTA experience (in months) as a covariate. Where experimental treatment was a significant predictor of the multivariate response, an analysis of covariance (ANCOVA) was run to examine the potential significance of training group on each SEEQ factor. Tukey HSD tests were used to resolve homogenous subsets. The MANCOVA was performed again without the covariate (becoming instead a multivariate analysis of variance termed a MANOVA) and none of the significance values (at the .05 significance level) changed, so I concluded that the significance level of multivariate effect did not depend on its inclusion. In this document, the MANCOVA results are reported for clarity.

For the CLE instrument, a similar analysis was performed. No processing of raw scores was necessary, since each CLE level was assessed by one item. Like the SEEQ data, CLE scores were analyzed using a factorial MANCOVA, where experimental treatment (IBL or control groups) and academic program (master's or PhD) were included as independent variables, along with previous GTA experience (in months) as a covariate. Where experimental treatment was a significant predictor of the multivariate response, a one-way ANCOVA was run to examine the potential significance of training group on each CLE item. Tukey HSD tests were again used to resolve homogenous subsets. Again, the MANCOVA was performed once more without the covariate and none of the significance values (at the .05 significance level) changed, so I concluded that the significance level of multivariate effect did not depend on its inclusion. Again, the MANCOVA results are reported in this document for clarity.

To analyze grade data, I used a two-factor ANCOVA to examine standardized grade scores, again with experimental treatment (control or IBL) and GTA academic program (master's or PhD) as independent variables and GTA experience (in years) as a covariate. Again, the analysis was re-run without the covariate (i.e., as an ANOVA) and the overall findings were consistent with the ANCOVA results. As such, only the ANCOVA results are reported here.

For all analyses where a significance test was repeated, Type I error was controlled for using a Bonferroni-corrected alpha value (e.g., the SEEQ inventory had nine factors, so I used $\alpha = .0056$). Effect sizes are reported using partial η^2 as my estimate of effect size. For individual SEEQ factors and CLE items, Pearson's *r* was used to illustrate the degree of linear dependence of that factor.

5 Results

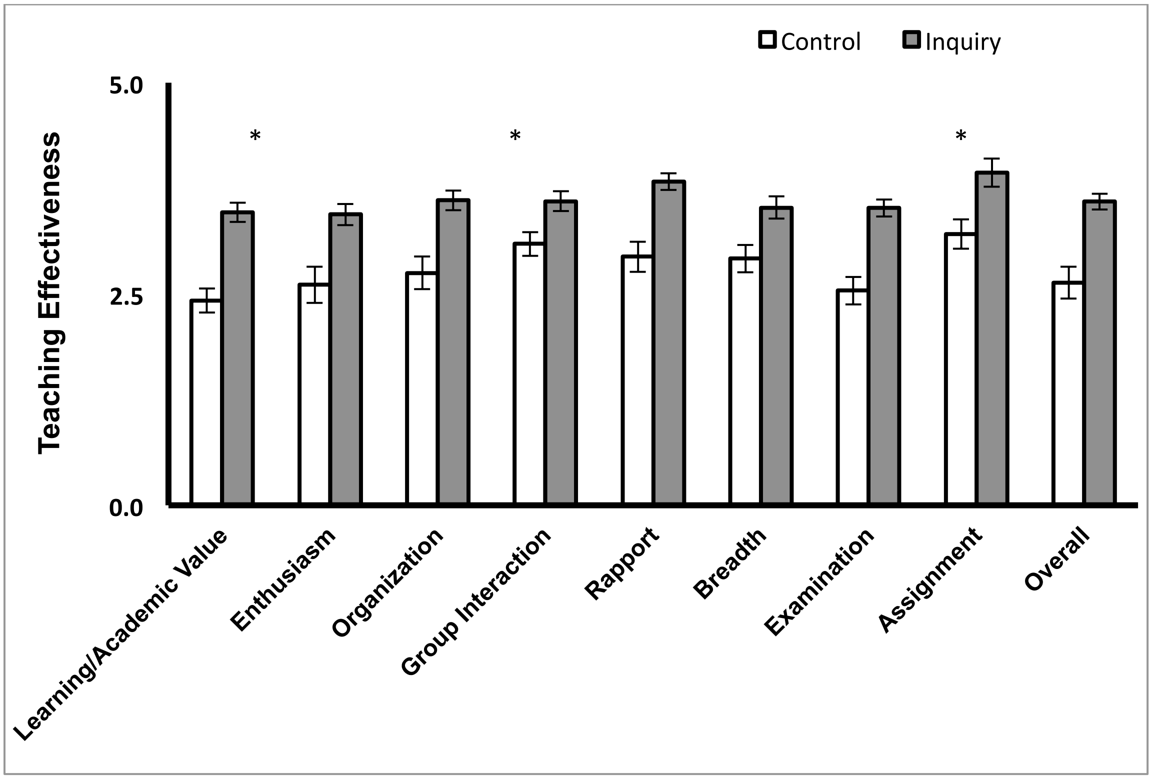
5.1 General Instructional Ability

5.1.1 GTA Self-Assessment

According to the MANCOVA analysis of SEEQ inventory data from GTA self-assessments, there was a significant multivariate main effect for training group (*Pillai's trace* = 0.38, $F(9,39) = 1.68$, $p = .017$, *partial* $\eta^2 = 0.38$, observed power = .89); GTAs in the IBL group scored significantly higher across SEEQ factors than GTAs in the control training group. There was no significant effect for GTA academic program (*Pillai's trace* = 0.15, $F(9,39) = 0.74$, $p = .67$, *partial* $\eta^2 = 0.15$, observed power = .31), GTA experience (*Pillai's trace* = 0.22, $F(9,39) = 1.22$, $p = .31$, *partial* $\eta^2 = 0.22$, observed power = .51) or the interaction between training group and GTA academic program (*Pillai's trace* = 0.22, $F(9,39) = 1.19$, $p = .33$, *partial* $\eta^2 = 0.22$, observed power = .50). When the covariate was dropped and the analysis was repeated as a MANOVA, no differences in the significance levels of any multivariate or univariate effects were found.

IBL group GTAs scored significantly higher than control group GTAs for three specific SEEQ factors: learning/academic value (Factor 1) $F(1,47) = 11.86$, $p < .001$, $r = .45$; group interaction (Factor 4) $F(1,47) = 11.86$, $p < .001$, $r = .48$; and overall instructional ability $F(1,47) = 15.34$, $p < .0005$, $r = .50$ (Figure 2). This analysis was performed using a Bonferroni-corrected alpha value of 0.0056.

Figure 2: Mean GTA Self-Assessed Teaching Effectiveness for SEEQ Factors (by GTA training group)



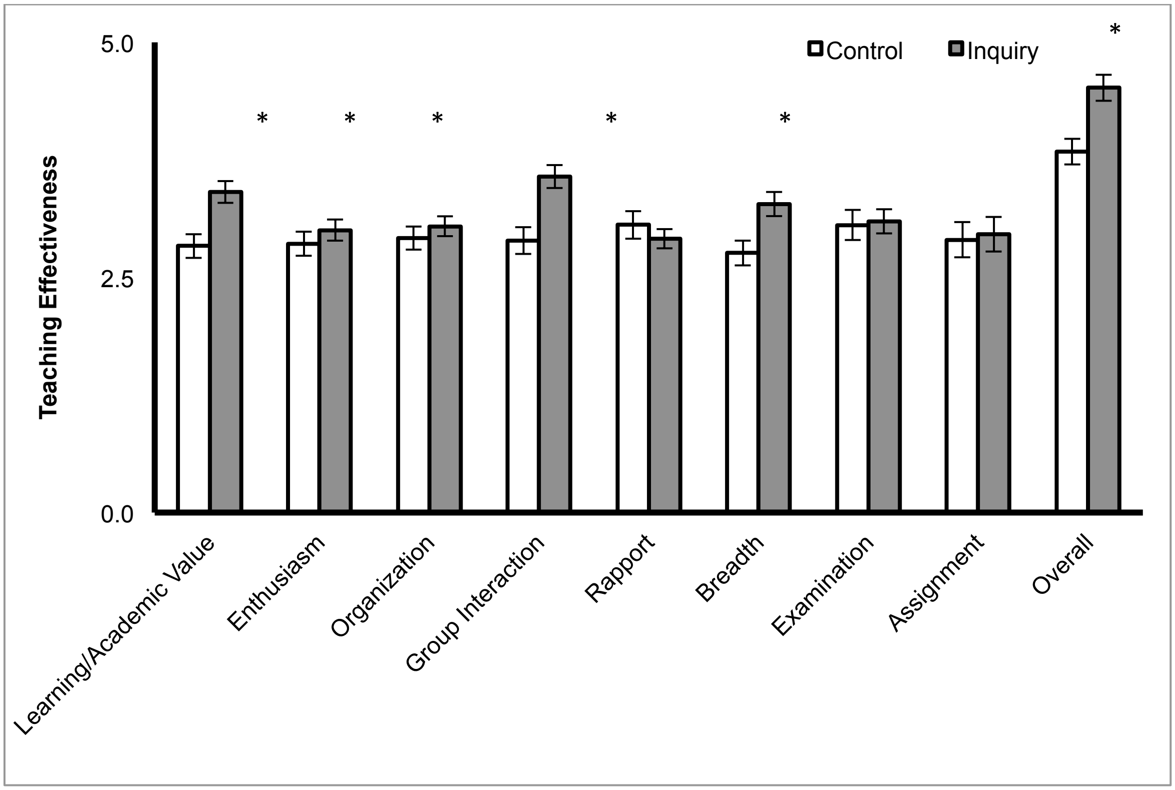
Note: Starred column pairs indicate statistically significant differences (Bonferroni-corrected alpha = 0.0056)

5.1.2 Undergraduate Assessment

According to the MANCOVA analysis of SEEQ inventory data from undergraduate assessments of GTAs, there was a significant multivariate main effect of training group (*Pillai's Trace* = .51, $F(9,34) = 3.93$, $p = .002$, *partial* $\eta^2 = 0.51$, observed power = .98). GTAs in the IBL training group scored significantly higher than GTAs in the control training group. There was no significant effect of GTA academic program (*Pillai's Trace* = .16, $F(9,34) = 0.70$, $p = .71$, *partial* $\eta^2 = 0.16$, observed power = .28) or GTA experience (*Pillai's Trace* = .18, $F(9,34) = 0.81$, $p = .61$, *partial* $\eta^2 = 0.18$). There was no significant interaction between training group and academic program (*Pillai's Trace* = .29, $F(9,34) = 1.57$, $p = .16$, *partial* $\eta^2 = 0.29$, observed power = .63). When the covariate was dropped and the analysis was repeated as a MANOVA, no differences in the significance levels of any multivariate or univariate effects were found.

IBL group GTAs scored significantly higher than control group GTAs on six specific SEEQ factors: learning/academic value (Factor 1), $F(1,42) = 26.47$, $p < .0005$, *partial* $\eta^2 = .39$; enthusiasm (Factor 2), $F(1,42) = 10.30$, $p < .003$, *partial* $\eta^2 = .20$; organization (Factor 3), $F(1,42) = 12.09$, $p < .001$, *partial* $\eta^2 = .22$; rapport (Factor 5), $F(1,42) = 14.47$, $p < .0005$, *partial* $\eta^2 = .26$; assignments (Factor 7), $F(1,42) = 25.79$, $p < .0005$, *partial* $\eta^2 = .38$; and overall instructional ability (Factor 9), $F(1,42) = 18.64$, $p < .0005$, *partial* $\eta^2 = .31$ (Figure 3). Univariate analyses were performed using a Bonferroni-corrected alpha value of 0.0056.

Figure 3: Mean Undergraduate Assessment of GTA Teaching Effectiveness for SEEQ Factors (by GTA training group)



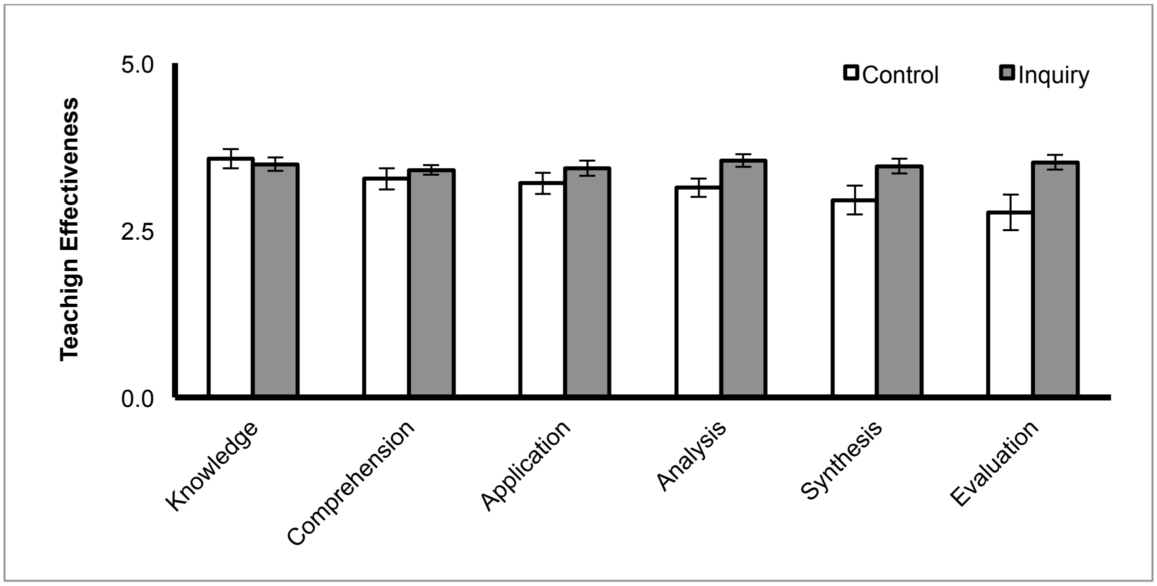
Note: Starred column pairs indicate statistically significant differences (Bonferroni-corrected alpha = 0.0056)

5.2 Scientific Inquiry

5.2.1 GTA Self-Assessment

According to the MANCOVA analysis of CLE instrument data from GTA self-assessments, there was no significant multivariate main effect for training group, (*Pillai's trace* = 0.19, *F*(6,42) = 1.68, *p* = .15, *partial* η^2 = 0.19, observed power = .57), meaning that there was no significant difference in self-assessed teaching effectiveness (by CLE item) between IBL and control training groups. In addition, there was no significant effect of GTA experience (*Pillai's trace* = 0.14, *F*(6,42) = 1.14, *p* = .36, *partial* η^2 = 0.14, observed power = .40), GTA academic program (*Pillai's trace* = 0.10, *F*(6,42) = 0.79, *p* = .58, *partial* η^2 = 0.10, observed power = .28) or the interaction between training group and program (*Pillai's trace* = 0.09, *F*(6,42) = 0.69, *p* = .66, *partial* η^2 = 0.09, observed power = .24). Given these results, univariate tests were not performed on training group. When the covariate was dropped and the analysis was repeated as a MANOVA, no differences in the significance level of any multivariate effects were found.

Figure 4: Mean GTA Self-Assessed Teaching Effectiveness for CLE Instrument (by GTA training group)



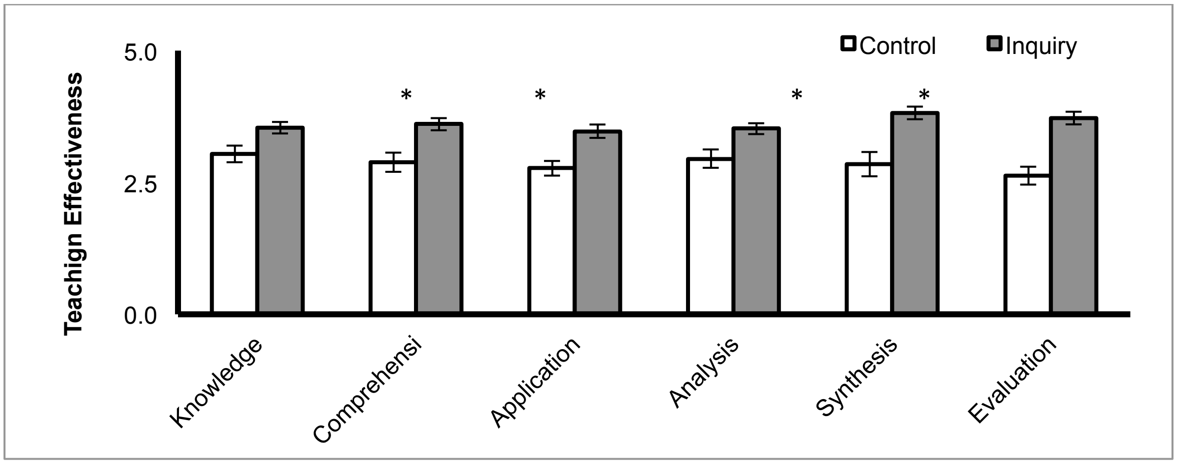
Note: Starred column pairs indicate statistically significant differences (Bonferroni-corrected alpha = 0.0083)

5.2.2 Undergraduate Assessment

According to the MANCOVA analysis of CLE instrument data from undergraduate assessments of GTAs, there was a significant multivariate main effect for training group on undergraduate CLE responses (*Pillai's trace* = 0.41, $F(6,40) = 4.71$, $p < .001$, *partial* $\eta^2 = 0.41$, observed power = .98). GTAs in the IBL training group scored significantly higher across CLE items than GTAs in the control group. There was no significant effect of GTA experience (*Pillai's trace* = 0.08, $F(6,40) = 0.60$, $p = .73$, *partial* $\eta^2 = 0.08$, observed power = .21). In addition, there was no significant effect for GTA academic program (*Pillai's trace* = 0.03, $F(6,40) = 0.17$, $p = .98$, *partial* $\eta^2 = 0.03$, observed power = .09) or a significant interaction between training group and program (*Pillai's trace* = 0.03, $F(6,40) = 0.22$, $p = .97$, *partial* $\eta^2 = 0.03$, observed power = .10). When the covariate was dropped and the analysis was repeated as a MANOVA, no differences in the significance levels of any multivariate or univariate effects were found.

IBL group GTAs scored significantly higher than control group GTAs on four specific CLE items. These included: comprehension skills (Item 2), $F(1,45) = 9.68$, $p < .003$, $r = .42$; problem-solving/application skills (Item 3), $F(1,45) = 9.76$, $p < .003$, $r = .42$; planning/synthesis skills (Item 5), $F(1,45) = 11.22$, $p < .001$, $r = .48$; and evaluation skills (Item 6), $F(1,45) = 26.12$, $p < .0005$, $r = .61$ (Figure 5). These analyses were performed using a Bonferroni-corrected alpha value of 0.0083.

Figure 5: Mean Undergraduate Assessment of GTA Teaching Effectiveness for CLE Instrument (by GTA training group)

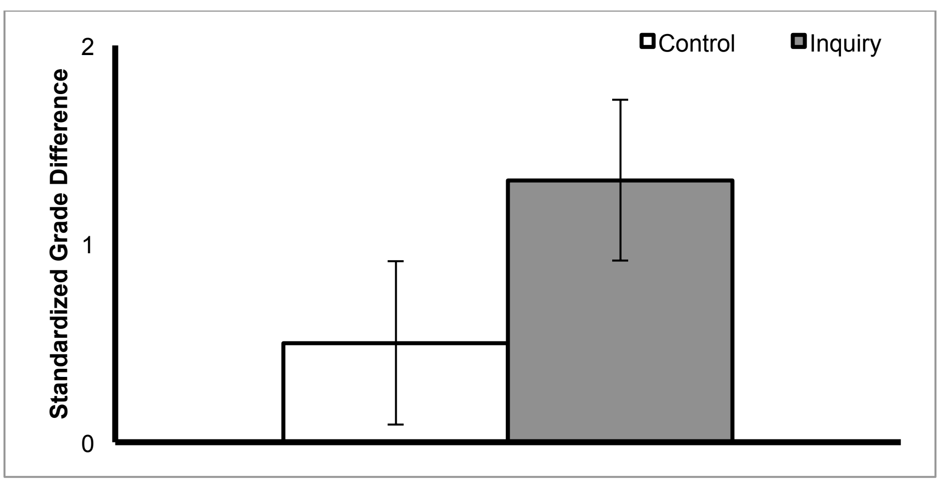


Note: Starred column pairs indicate statistically significant differences (Bonferroni-corrected alpha = 0.0083)

5.3 Undergraduate Grades

From an ANCOVA analysis of standardized grade difference data, there was a significant main effect of training group on undergraduate grades, $F(1,44) = 26.31$, $p < .022$, $r = .34$. Students in courses whose lab GTA was part of the IBL training group achieved significantly higher self-reported grades than students whose lab GTA was part of the control training group. There was no significant main effect for GTA teaching experience ($F(1,44) = .692$, $p = .69$, $r = .09$), GTA academic program ($F(1,44) = 1.66$, $p = .56$, $r = .19$) or an interaction between training group and GTA academic program ($F(1,44) = 0.53$, $p = .47$, $r = .11$, see Figure 6).

Figure 6: Undergraduate Grades by Training Group



6. Discussion

6.1 Does Grounding Lab GTA Training in Inquiry-Based Learning Improve GTAs' General Instructional Ability?

The results of this study support the hypothesis that inquiry-based learning training improves GTAs' general instructional ability. GTAs in the IBL training group consistently outperformed control GTAs across all SEEQ factors. IBL GTAs were rated more highly on three of nine factors by the GTAs themselves and on six of nine factors by undergraduates.

With respect to the SEEQ data, there was both agreement and disagreement between undergraduate assessments and GTA self-assessments. IBL GTAs' self-assessments showed a self-perception of improved teaching ability in three areas: SEEQ factors 1, 4 and 9. According to undergraduates, IBL GTAs scored more highly on SEEQ factors 1, 2, 3, 5, 7 and 9.

SEEQ factor 1 (learning/academic value) refers to the feeling of stimulation and interest in the learning process, as well as engagement with course material. SEEQ factor 2 (enthusiasm) refers to attention arousal and capture, as well as the instructor's ability to generate sustained student interest in the subject matter. SEEQ factor 3 (organization) refers to structure and context, and the instructor's ability to construct logical links. SEEQ factor 4 (group interaction) refers to the ability to maximize learning using social interaction and group activities. SEEQ factor 5 (individual rapport) refers to the ability to become personally involved with students one-on-one to offer direction and guide the learning process. SEEQ factor 7 (examination) refers to instructors' ability to provide useful feedback to students. SEEQ factor 9 refers to "overall" instructional quality. See Table 7 for a full description of SEEQ factors.

Both GTA and undergraduate responses highlight increased teaching effectiveness in areas that are of particular importance to structured-inquiry activities such as science labs. Structured inquiry is by its nature equally a self-directed and instructor-directed exercise. Students likely value the ability for GTAs to assist them in both areas: to facilitate the process of structured inquiry and to relate content effectively. The fact that both undergraduates and GTAs highlight general measures of instructional quality, like learning value and overall instructional ability, indicates that GTAs are perceived to be relating content effectively. The fact that students also cite enthusiasm, group interaction and rapport as significantly better in IBL-trained GTAs may indicate that IBL training, which encourages active feedback as the lab activity is performed, has enabled IBL GTAs to correctly diagnose student needs, both individually and in lab groups. In the IBL training, GTAs were instructed to, among other teaching strategies, encourage students to ask each other questions and solve problems in groups rather than ask for answers from lab coordinators or GTAs. Since lab activities are traditionally individualistic activities, the promotion of a social atmosphere within eight- to ten-student lab groups is possibly due to the GTA training environment.

Notably, undergraduates identified two areas of GTA improvement that GTAs did not recognize in their own teaching: SEEQ factors 2 (enthusiasm) and 5 (rapport). This is interesting because it suggests that GTA teaching ability may improve even when GTAs are not aware of the improvement. Although previous work has indicated that student assessments and instructor self-assessments show consistent ratings of instructor teaching effectiveness (Marsh & Roche, 1994, 1997; Marsh, 1982; Miller, 1988), this was not what was observed here.

6.2 Does Grounding Lab GTA Training in Inquiry-Based Learning Improve GTAs' Ability to Teach Scientific Inquiry?

The results of this study support the hypothesis that inquiry-based learning training improves GTAs' ability to teach higher-order cognitive skills (such as scientific inquiry) to undergraduates to a greater degree than traditional "best practices" training. There were several instances where GTAs in the IBL training group outperformed control training group GTAs across the range of cognitive skills identified by the CLE; there were no instances where control group GTAs outperformed IBL group GTAs. Interestingly, none of these differences were identified by GTA self-assessments. Although undergraduate assessments of GTA teaching ability showed a significantly higher score for IBL GTAs for CLE items 2 (comprehension), 3 (application/problem-solving), 5 (planning and synthesis) and 6 (evaluation), no significant differences were found between IBL GTA self-assessments and control group GTA self-assessments for any CLE item.

Undergraduates scored IBL GTAs much higher than control group GTAs on four CLE items: 2, 3, 5 and 6. The CLE items are arranged hierarchically after the six skills in the cognitive domain of Bloom's Taxonomy. Therefore lower numbers represent lower-order cognitive skills and higher numbers indicate higher-order cognitive skills; these latter items are the skills that represent scientific inquiry. For example, CLE item 2 represents comprehension skills; this is the ability to compare or translate two ideas, for instance taking a word problem and constructing a formal mathematical model. CLE item 3 refers to problem-solving skills, such as the ability to select an appropriate statistical test to address a particular hypothesis. CLE item 5 deals with planning skills, such as the design of an experimental protocol from first principles. CLE item 6 refers to evaluation skills, such as identifying the correct abstract or theoretical approach to design a set of experiments. Notably, the cognitive skills specified by CLE items 5 and 6 are two of the hardest skills to teach, and are correspondingly valued when taught effectively. Undergraduate students indicated that IBL training group GTAs outperformed control group GTAs, especially in three of the top four cognitive skills. This supports the hypothesis that IBL training for GTAs increases (at least perceived) teaching effectiveness with respect to higher-order cognitive skills. Because planning and evaluation skills are involved in experimental design and scientific analysis, these are sought-after outcomes of the educational process in the lab.

6.3 Does Grounding Lab GTA Training in Inquiry-Based Learning Affect Undergraduate Grades?

Based on the information collected, the data from this experiment support the hypothesis that inquiry-based learning training for GTAs is associated with improved undergraduate grades, relative to traditional GTA training (Figure 6). When overall self-reported course grades were standardized for course instructor mean grade, final grades of undergraduate students with IBL-group instructors outperformed undergraduate students of control group instructors by a grade differential of approximately 1.2 CGPA points (~5-8%). This is a modest gain, but this trend is statistically significant.

Some caveats with respect to this conclusion are necessary. First, undergraduates self-reported their own grades, and it is not known if they did so accurately. Second, undergraduates self-reported final course grades, not just grades for the lab portion of their course. Third, no experimental control for course instructor was used, although GTAs were roughly equally distributed among course instructors and course instructor class averages are controlled for statistically. In addition, since undergraduates were blinded as to the training group allocation of their GTA (assuming that they knew that GTAs were undergoing any kind of pedagogical training), there is no reason to believe that undergraduates in one group would be subject to any one of the caveats more than undergraduates of the other. Thus, this finding – statistically significant differences in final course grade – is an important piece of evidence that IBL training for GTAs is a genuine improvement over the control "status quo" training.

6.4 GTA and Undergraduate Views of GTA Teaching Ability

Although GTAs and undergraduates used the same survey instruments to assess the same object (GTA teaching effectiveness), their responses did not completely overlap. Statistically significant differences between the two training groups also varied between the SEEQ and CLE assessments of teaching ability. Generally, undergraduate responses identified a greater number of differences in teaching effectiveness between the IBL and control training groups, although both GTAs and undergraduate responses indicated that IBL training group GTAs were more effective lab teachers than control training group GTAs.

With respect to the SEEQ inventory, both undergraduates and GTA responses identified improvement in SEEQ factors 1 and 9, but only undergraduate respondents identified improvement in SEEQ factors 2 and 5. Factors 1 (learning/academic value) and 9 (overall instructional ability) are general indicators of teaching effectiveness, but factors 2 (enthusiasm) and 5 (rapport) are more specific. Because effective structured-inquiry learning relies on the presence of a learning facilitator who is capable of providing guidance to students, higher ratings for SEEQ factors 2 and 5 suggest an increased ability to act as this facilitator. That GTAs did not seem to notice this improvement is intriguing. These results suggest that improvements made in teaching effectiveness may not be accompanied by corresponding improvements in awareness of teaching effectiveness.

CLE instrument data seem to confirm the idea that undergraduate students rate IBL training group GTAs higher in skills associated with inquiry facilitation, but this is not recognized by the GTAs themselves. Once again, IBL training group GTAs are rated better than control training group GTAs for CLE items 2, 3, 5 and 6, but no self-assessed differences were found between training groups. That undergraduate students consistently rated IBL training group GTAs higher in teaching ability, particularly for the higher-order cognitive skills representative of scientific inquiry, suggests that GTAs from the IBL training group were genuinely better at teaching the structured-inquiry lab activity. Once again, however, GTAs themselves did not “see” any improvement.

In conclusion, GTA self-assessments may be less valid measurements of GTA teaching efficacy than are undergraduate assessments of GTAs. Certainly, GTAs are much less experienced as teachers than undergraduates are as learners; they may have insight into the educational assessment process that new teachers such as GTAs may not have. However, since GTAs typically have more postsecondary student experience than undergraduates do, GTAs may be making decisions about the value of their lab teaching based on their student experience. Where evaluations of GTA teaching involve GTA assessments, it may be prudent to obtain undergraduate assessment data as well (despite the additional effort and expense).

6.5 Should Inquiry-Based Training be Provided for Lab GTAs in the STEM Disciplines?

The results of this study support the hypothesis that inquiry-based learning training for GTAs improves their teaching effectiveness according to each and all of the metrics of instructional ability that I used in this experiment.

GTAs trained in inquiry-based teaching methods outperformed GTAs trained using control “best practice” methods, but more importantly IBL GTAs consistently scored higher as teachers of scientific inquiry. Teaching this skill is the main reason that lab activities are used as a component of undergraduate degree programs in STEM disciplines, and a training regimen that has been shown to improve GTA teaching ability in a structured-inquiry environment is a valuable tool for improving undergraduate learning outcomes.

7. Conclusions and Further Research

Although there are some limitations as to the generalizability of these findings, this evidence shows that teaching GTAs the rudiments of inquiry-based learning pedagogy is an important improvement that can take place anywhere lab activities are offered as part of a STEM undergraduate program (or indeed other programs using IBL as well). Clearly, more research into the causes of GTA teaching improvement may be necessary, as well as a larger, more comprehensive study. Nevertheless, this insight is important for GTA training programs and that the data collected in this study suggest that inquiry-based material should, where possible, be incorporated into the GTA training curriculum wherever STEM GTAs will be teaching using structured- or open-inquiry labs.

Further research into how inquiry-based methods improve GTA lab teaching is necessary. In this study, although GTAs given IBL pedagogical training were better teachers in the eyes of undergraduate students, they did not seem to self-assess their teaching abilities to be much different from those of GTAs not offered IBL training. The reason that this is the case should be explored in future studies. Further, although this study showed that students in lab groups led by IBL GTAs ended up with significantly higher course grades, two outcomes remain unclear: first, whether or not this grade improvement is a generalizable finding, instead of an artefact of this study design and grade measurement; and second, whether students trained in scientific inquiry by IBL training group GTAs score higher in scientific inquiry skills outside of this particular course (say, in an introductory physics course).

Finally, it is unknown whether the effects of improved GTA training last longer than the duration of the course involved; this is another area that could be fruitful for further study. It is possible that one five-hour intervention is not enough to permanently improve the teaching effectiveness of a GTA, although the duration of the improvement studied here is not known. Long-term recall of the pedagogical principles involved may be an important area of follow-up research.

In conclusion, these data suggest that offering GTAs training in relevant pedagogical methods is a sound and efficient strategy to improve undergraduate teaching in science labs, and that postsecondary institutions making use of lab GTAs should support such training schemes, where time and resources permit.

8. References

Apedoe, X., Walker, S., & Reeves, T. (2006). Integrating inquiry-based learning into undergraduate geology. *Journal of Geoscience Education*, *54*(3), 414-421. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.1751&rep=rep1&type=pdf

Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., Liu, Q., et al. (2009). Learning and Scientific Reasoning. *Science*, *323*, 586-587.

Black, B., & Bonwell, C. (1991). The training of teaching assistants in departments of history. *The History Teacher*, *24*(4), 435-444. Retrieved from http://www.jstor.org/stable/10.2307/494702

Bohrer, K., Stegenga, B. D., & Ferrier, A. (2007). Training and Mentoring TAs in Inquiry-Based Methods. *Tested Studies for Laboratory Teaching. Proceedings of the 28th Workshop/Conference of the Association for Biology Laboratory Education* (Vol. 28, pp. 335-346).

Bond-Robinson, J., & Rodriques, R. (2006). Catalyzing graduate teaching assistants' laboratory teaching through design research. *Journal of Chemical Education*, *83*(2), 313-323. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed083p313

Brown, S. (2010). A Process-Oriented Guided Inquiry Approach to Teaching Medicinal Chemistry. *American Journal of Pharmaceutical Education*, *74*(7), 1-6. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2972515/

Buerkel-Rothfuss, N., & Gray, P. (1990). Graduate teaching assistant training in speech communication and noncommunication departments: a national survey. *Communication Education*, *39*, 292-307. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/03634529009378811

Creech, L. R., & Sweeder, R. D. (2012). Analysis of student performance in large-enrollment life science courses. *CBE Life Sciences Education*, *11*(4), 386-391. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3516794&tool=pmcentrez&rendertype=abstract

Dawson, D. L., Dimitrov, N., Meadows, K. N., & Olsen, K. (2013). *Bridging the Gap: The Impact of the "Teaching in the Canadian Classroom" Program on the Teaching Effectiveness of International Teaching Assistants* (pp. 1-46). Toronto: Higher Education Quality Council of Ontario.

Ertepinar, H., & Geban, Ö. (1996). Effect of instruction supplied with the investigative-oriented laboratory approach on achievement in a science course. *Educational Research*, *38*(3), 333-341. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/0013188960380306

Finnie, R., & Pavlic, D. (2013). *Background Characteristics and Patterns of Access to Postsecondary Education in Ontario: Evidence from Longitudinal Tax Data* (p. 67). Toronto: Higher Education Quality Council of Ontario.

French, D., & Russell, C. (2002). Do Graduate Teaching Assistants Benefit from Teaching Inquiry-Based Laboratories? *BioScience*, *52*(11), 1036. Retrieved from http://www.jstor.org/stable/1314349

Goldey, E. S., Abercrombie, C. L., Ivy, T. M., Kusher, D. I., Moeller, J. F., Rayner, D. A., Smith, C. F., et al. (2012). Biological inquiry: a new course and assessment plan in response to the call to transform undergraduate biology. *CBE Life Sciences Education*, *11*(4), 353-363. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3516791&tool=pmcentrez&rendertype=abstract

Hanauer, D. I., Jacobs-Sera, D., Pedulla, M. L., Cresawn, S. G., Hendrix, R. W., & Hatfull, G. F. (2006). Teaching Scientific Inquiry. *Science*, *314*(5807), 1880-1881.

Hardré, P. L. (2005). Instructional Design as a Professional Development Tool-of-Choice for Graduate Teaching Assistants. *Innovative Higher Education*, *30*(3), 163-175. Retrieved from http://www.springerlink.com/index/10.1007/s10755-005-6301-8

Hardré, P. L., & Burris, A. O. (2010). What contributes to teaching assistant development: differential responses to key design features. *Instructional Science*, *40*(1), 93-118. Retrieved from http://www.springerlink.com/index/10.1007/s11251-010-9163-0

Herrington, D. G., & Nakhleh, M. B. (2003). What Defines Effective Chemistry Laboratory Instruction? Teaching Assistant and Student Perspectives. *Journal of Chemical Education*, *80*(10), 1197. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed080p1197

Iler, H., & Justice, D. (2012). Discovering 13C NMR, 1H NMR, and IR Spectroscopy in the General Chemistry Laboratory through a Sequence of Guided-Inquiry Exercises. *Journal of Chemical Education*, *89*, 1178-1182. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed2005664

Kendall, K. D., & Schussler, E. E. (2012). Does instructor type matter? Undergraduate student perception of graduate teaching assistants and professors. *CBE Life Sciences Education*, *11*(2), 187-199. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3366904&tool=pmcentrez&rendertype=abstract

Kirschner, P. A., Sweller, J., & Clark, R. E. (2010). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Based Teaching Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75-86.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, *41*(4), 212-218. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15430421tip4104_2

Lin, S.-Y., Henderson, C., Mamudi, W., Singh, C., & Yerushalmi, E. (2013). Teaching assistants' beliefs regarding example solutions in introductory physics. *Physics Education Research*, *9*(1), 1-23. Retrieved from http://link.aps.org/doi/10.1103/PhysRevSTPER.9.010120

Linn, M. C., Lee, H., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and Assessing Knowledge Integration in Science. *Science*, *313*, 1-2.

Lowman, J., & Mathie, V. A. (1993). What Should Graduate Teaching Assistants know about Teaching? *Teaching of Psychology*, *20*(2), 84-88.

Luckie, D. B., Maleszewski, J. J., Loznak, S. D., & Krha, M. (2004). Infusion of collaborative inquiry throughout a biology curriculum increases student learning: a four-year study of “Teams and Streams”. *Advances in Physiology Education*, *28*(1-4), 199-209. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15545349

Luft, J. A., Kurdziel, J. P., Roehrig, G. H., & Turner, J. (2004). Growing a garden without water: Graduate teaching assistants in introductory science laboratories at a doctoral/research university. *Journal of Research in Science Teaching*, *41*(3), 211-233. Retrieved from http://doi.wiley.com/10.1002/tea.20004

Marbach-Ad, G., Schaefer, K. L., Kumi, B. C., Friedman, L. A., Thompson, K. V., & Doyle, M. P. (2012). Development and Evaluation of a Prep Course for Chemistry Graduate Teaching Assistants at a Research University. *Journal of Chemical Education*, *89*(7), 865-872. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed200563b

Marsh, H., & Bailey, M. (1993). Multidimensional Students’ Evaluations of Teaching Effectiveness: A Profile Analysis. *Journal of Higher Education*, *64*(1), 1-18. Retrieved from http://www.jstor.org/stable/10.2307/2959975

Marsh, H. W., & Roche, L. A. (1994). *The use of students’ evaluations of university teaching to improve teaching effectiveness*. Canberra: Australian Government Publishing Service.

Marsh, H. W., & Roche, L. A. (1997). Making students’ evaluations of teaching effectiveness effective. *American Psychologist*, *52*, 1187-1197.

Marsh, H. W. (1982). SEEQ: a reliable, valid and useful instrument for collecting students’ evaluations of university teaching. *British Journal of Educational Psychology*, *52*(1), 77-95.

Marsh, H. W. (1987). Students’ evaluations of University teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research*, *11*(3), 253-388. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/0883035587900012

Meyers, S., & Prieto, L. (2000). Training in the Teaching of Psychology: what is done and examining the differences. *Teaching of Psychology*, *27*(4), 258-261. Retrieved from http://www.tandfonline.com/doi/full/10.1207/S15328023TOP2704_03

Miles, C. A., & Polovina-Vukovic, D. (2012). *The Role of New Faculty Orientations in Improving the Effectiveness of University Teaching - Part 1: University Sector* (p. 33). Toronto: Higher Education Quality Council of Ontario.

Miller, A. (1988). Student assessment of teaching in higher education. *Higher Education*, *15*, 3-15. Retrieved from http://www.springerlink.com/index/l205463033172464.pdf

Norton, L., Richardson, T. E., Hartley, J., Newstead, S., & Mayes, J. (2005). Teachers’ beliefs and intentions concerning teaching in higher education. *Higher Education*, *50*(4), 537-571. Retrieved from http://www.springerlink.com/index/10.1007/s10734-004-6363-z

Nyquist, J. D., Abbott, R. D., & Wulff, D. H. (1989). The challenge of TA training in the 1990s. *New Directions for Teaching and Learning*, *39*, 7-14.

Park, C. (2004). The graduate teaching assistant (GTA): lessons from North American experience. *Teaching in Higher Education*, *9*(3), 349-361. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/1356251042000216660

Price, R. M. (2012). How We Got Here: An Inquiry-Based Activity About Human Evolution. *Science*, *338*, 1554-1555.

Prieto, L. (2002). Teaching Assistant Development: Research and Impressions. *Essays from Excellence in Teaching*, *2002*(January 2002). Retrieved from http://www.teachpsych.com/ebooks/eit2002/eit02-01.pdf

Prieto, L. R., & Altmaier, E. M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Research in Higher Education*, *35*(4), 481-497. Retrieved from http://link.springer.com/10.1007/BF02496384

Prieto, L. R., & Meyers, S. A. (1999). Effects of Training and Supervision on the Self-Efficacy of Psychology Graduate Teaching Assistants. *Teaching of Psychology*, *26*(4), 264-266. Retrieved from http://top.sagepub.com/lookup/doi/10.1207/S15328023TOP260404

Rodriques, R. A. B., & Bond-Robinson, J. (2006). Comparing Faculty and Student Perspectives of Graduate Teaching Assistants' Teaching. *Journal of Chemical Education*, *83*(2), 305. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed083p305

Roehrig, G. H., Turner, J. A., Kurdziel, J. P., & Luft, J. A. (2003). Graduate Teaching Assistants and Inquiry-Based Instruction: Implications for Graduate Teaching Assistant Training. *Journal of Chemical Education*, *80*(10), 1206. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed080p1206

Rushin, J., Saix, J.-D., Lumsden, A., Streubel, D. ., Summers, G., & Bernson, C. (1997). Graduate Teaching Assistant Training: A Basis for Improvement of College Biology Teaching & Faculty Development? *The American Biology Teacher*, *59*(2), 86-90. Retrieved from http://www.jstor.org/stable/10.2307/4450255

Schoffstall, A. M., & Gaddis, B. A. (2007). Incorporating Guided-Inquiry Learning into the Organic Chemistry Laboratory. *Journal of Chemical Education*, *84*(5), 848. Retrieved from http://pubs.acs.org/doi/abs/10.1021/ed084p848

Shannon, D., Twale, D., & Moore, M. (1998). TA Teaching Effectiveness: the impact of training and teaching experience. *Journal of Higher Education*, *69*(4), 440-466. Retrieved from http://www.jstor.org/stable/10.2307/2649274

Tompkins, L. J., & Dimiduck, K. (2011). *Confident or Overconfident: Changes in the self-reported confidence levels of new teaching assistants pre- and post-training* (pp. 1-11). Ithaca, NY.

University Website, Carleton University. (2013). Facts and Reports - About Carleton. Retrieved April 20, 2013, from http://www1.carleton.ca/about/facts/

Uno, G. E. (2009). Botanical literacy: What and how should students learn about plants? *American journal of botany*, *96*(10), 1753-1759. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21622295

Vale, R. D., DeRisi, J., Phillips, R., Mullins, R. D., Waterman, C., & Mitchison, T. J. (2012). Graduate education. Interdisciplinary graduate training in teaching labs. *Science*, *338*(6114), 1542-1543. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/23258877

Verleger, M., & Velasquez, J. (2007). An engineering teaching assistant orientation program: Guidelines, reactions, and lessons learned from a one day intensive training program. *2007 37Th Annual Frontiers in Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports*, S4G–3–S4G–7. Ieee. Retrieved from http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4417912

Volkmann, M. J., & Zgagacz, M. (2004). Learning to teach physics through inquiry: The lived experience of a graduate teaching assistant. *Journal of Research in Science Teaching*, *41*(6), 584-602. Retrieved from http://doi.wiley.com/10.1002/tea.20017

Wallace, C. S., Tsoi, M. Y., Calkin, J., & Darley, M. (2003). Learning from inquiry-based laboratories in nonmajor biology: An interpretive study of the relationships among inquiry experience, epistemologies, and conceptual growth. *Journal of Research in Science Teaching*, *40*(10), 986-1024. Retrieved from http://doi.wiley.com/10.1002/tea.10127

Wilke, R. R., & Straits, W. J. (2005). Practical Advice for Teaching Inquiry-Based Science Process Skills in the Biological Sciences. *The American Biology Teacher*, *67*(9), 534-540. Retrieved from http://www.bioone.org/doi/abs/10.1662/0002-7685(2005)067[0534:PAFTIS]2.0.CO;2

Higher Education Quality Council of Ontario

An agency of the Government of Ontario