Developing Inquiry Skills with Undergraduate Students at McMaster 1999-2004: Ongoing Formative and Summative Evaluation for the Faculties of Science and Health Science

By

KRISTINA SYLVIA TRIM

BSW, MA

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

McMaster University

Your file Votre référence
ISBN: 978-0-494-28213-7
Our file Notre référence
ISBN: 978-0-494-28213-7

Evaluating Inquiry Skills with McMaster Science and Health Science Students
1999- 2004

Doctor of Philosophy (2006) McMaster University

Department of Clinical Epidemiology and Biostatistics Hamilton Ontario

TITLE:	Developing Inquiry Skills with Undergraduate Students at McMaster 1999-2004: Ongoing Formative and Summative Evaluation for the Faculties of Science and Health Science
AUTHOR:	KRISTINA SYLVIA TRIM, BSW, MA (McMaster University)
SUPERVISOR:	Dr. Geoffrey R. Norman
NUMBER OF PAGES:	iii, 176

Abstract

Context: In 1997, McMaster University recognized that students were primarily in content-driven, high enrollment lecture courses. One remedy to this problem was the McMaster Inquiry Project and its goal was to implement inquiry method courses so students had opportunities to be active learners throughout their undergraduate experience.

Objective: This dissertation reports on the program evaluation of Science and Health Science Inquiry courses from 1999 to 2004. Research questions were: did Inquiry courses foster attributes of academic self-confidence, critical thinking, and a preference to delve deeply into subject areas instead of test memorization; and, were these gains in evidence at graduation.

Design: Program evaluation over six years. Initially, a qualitative study was conducted to identify relevant domains. In subsequent years, quantitative, longitudinal data was collected from Inquiry students. Cross-sectional and longitudinal data was collected from control group students. Follow-up data was collected from first year Inquiry course students at graduation and compared to graduating students who had not taken a first year Inquiry course.

Setting: Undergraduate Science and Health Science programs at a research-intensive university with health professional programs in an urban, Canadian community.

Participants: Thirty-three students participated in the qualitative study. Approximately 80 to 100 students from each course (i.e., first year science, fourth year science and first year health science) participated in the project for all four years.

Instrument: The development and validation of two instruments are described: an objective measure of critical thinking scored by raters called the Mini-Inquiry Benchmarking Exercise and a self-assessed measure called the Science Inquiry Academic Self-Confidence Scale. The self-assessed measure, Approaches to Learning Scale (Lancaster), was also validated for these samples.

Outcome Measures: Critical thinking outcomes included the Appraisal of Arguments, Information Seeking, Self-Assessment And Overall Critical Thinking Ability scores. Self-confidence outcomes included the Inquiry Task Self-Confidence and the Projected Academic Goal Completion scores. The items used in the Lancaster scale corresponded to the Meaning, Reproducing and Organized Orientations sub-scores.

Results: Inquiry Task Self-Confidence, Information Search Strategy, Overall Critical Thinking Assessment, and Self-Assessment Skills were the four areas that all Inquiry students made progress and retained their gains at follow-up. Projected Academic Goal Self-Confidence and Appraisal of Arguments were important gains for first year inquiry students that were retained by them in their graduating year while fourth year Inquiry students did not make gains in these outcomes. None of the Inquiry students groups

changed their Meaning Orientation preference in the longitudinal or follow-up comparisons. The reproducing orientation appeared to be intensely disliked by 4th year and the follow-up Inquiry students while first year students did not change in their reproducing orientation. Conversely, the organized orientation was relevant to first year students but not for 4th year or follow-up students. Longitudinal results for the students not enrolled in a first year Inquiry course (control group) showed a decrease in their Appraisal of Arguments, Overall Critical Thinking Assessment, Self-Assessment Skills, and Organized Orientation scores. Non-inquiry students made gains on the concrete skills of Information Search Strategy and Inquiry Task Self-Confidence scores but these were half the magnitude of first year inquiry students. Most troubling was that non-inquiry students increased in their Reproducing Orientation preferences and had no changes on their Projected Academic Goal Self-Confidence.

Conclusions: Creating active learning opportunities where students feel that they have the appropriate level of autonomy and support to learn what it means to study at university is essential. Attending a research-intensive university can be a positive experience for undergraduates by exposing them to academic research skills. The McMaster Inquiry Project for science and health science students appears to help students develop successfully into active learners.

Acknowledgements

I would like to thank my Thesis Committee for their continuing support throughout the past six years: Geoffrey Norman, Del Harnish, Miroslav Lovric, and Lee Brooks.

I would like to thank the following people for their advice and support of the following people in completing this thesis:

Bruce Weaver, Assistant Professor, Lakehead University
Dale Roy, Director, Centre for Leadership in Learning, McMaster University
Mary Lou Schmuck, Data Analyst, Program for Educational Research Development, McMaster University
Christopher Knapper, Consultant, Centre for Leadership in Learning, McMaster University
Daphne Maurer, Professor, Psychology Department, McMaster University
Erika Kustra, Educational Consultant, Centre for Leadership in Learning, McMaster University
Jennifer McKinnell, Librarian, McMaster Health Sciences Library, McMaster University
Barbara Bloemhof, Assistant Professor, Brock University
Jennifer Landicho, Bachelor of Health Sciences Program, McMaster University
Stash Nastos, Bachelor of Health Sciences Program, McMaster University
Victoria Caminiti, Research Assistant, Brock University
Indu Mand-Bains, Janani Sivisithamparam, Catherine Pierre, Brian Heal, Amanda Jerome, Jonathan Wong, Elizabeth Faulkner, Krystyna Karwalajtys, Research Assistants, McMaster University

I would also like to thank all of the 26 instructors who have been involved in creating and revising the Inquiry courses elements since its inception.

I would like to the J.W. McConnell Foundation for providing funding for the project.

I would also like to sincerely thank the 1750 plus students and 60 plus peer tutors for being participating in the course and giving us feedback over the years. Without them, this dissertation would not have been possible.

Dedication

To my family—
for their love, support and patience.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Rationale for the Inquiry Project: In 1997, McMaster University recognized that first year science students have few opportunities to engage in small group, self-directed research skills development while spending their time in primarily content-driven, high enrollment lectures. In response, McMaster's academic administrators developed policies and programs to remedy these problems. Some of the initiatives included an active faculty recruiting strategy for the next five years. New faculty members were offered partial relief from their teaching responsibilities in their first year so that they may attend in their first year active learning strategies course design seminars offered by the Centre for Leadership and Learning (McMaster University Planning Committee, 2000). A third initiative was the McMaster Inquiry Project. The central goal of the McMaster Inquiry Project was to inculcate the concept of 'student as active learner' throughout the entire undergraduate experience: undergraduates should not learn primarily as passive recipients of information, but as active participants in their education to better develop critical thinking and communications skills (Centre for Leadership and Learning, 1999). This initiative encompassed the Humanities, Social Science and Science faculties over a three-year period and was intended to be offered in all years, culminating in all students completing a research project in their final year. The focus of this dissertation is the McMaster Inquiry project as it concerns the Faculty of Science.

The University's decision to focus on the Inquiry Project was heavily influenced by the Boyer Commission Report (McMaster University Planning Committee, 2000). While acknowledging that McMaster was a research intensive institution, the University Planning Committee also noted that, if stronger linkages were developed between research and teaching, students would have the opportunity to be presented with "cutting-edge content in the field [which would] also allow students to develop the generic, critical thinking, inquiry, research and communication skills that are inherent elements of the scholarly life and that form the foundation of a well-educated student" (p. 4). However the Planning Committee also noted that this linkage would take time, organizational commitment and budget reallocations to ensure that the goal of high quality undergraduate education was achieved (McMaster University Planning Committee, 2000).

A second organizational initiative for the Inquiry Project came from an external funding agency, the McConnell Foundation. Canadian universities were invited by the Foundation to compete for grants that would be strategic in helping the universities translate active learning initiatives into durable practices. The McConnell Foundation recognized that the "knowledge economy" required workers to move between learning and work environments and these transitions required more collaboration and problem-based learning strategies than in previous decades (J.W. McConnell Family Foundation, 2001). The Centre for Leadership in Learning at McMaster was awarded a three-year grant to assist in the "major challenge [of] helping professors to change their approach to teaching" (p. 2).

There were two key stakeholder groups who were interested in having the inquiry programs evaluated. The first group consisted of the University administration (i.e., Provost, Dean of Science, and University Planning Committee), and the Centre for Leadership and Learning who were responsible for reporting to the McConnell Foundation on the Project's progress. The University Administration stakeholders were interested in faculty development. The second group was the faculty members who became instructors in the course: their goals are centered on developing a course design that was workable for them and provides first year science students with the opportunity to engage in small-group, problem-based learning. It became clear in the process of this case analysis that the stakeholder groups' goals were separate yet interrelated and this overlap was been instrumental in the success of Inquiry.

Outline of the Dissertation: This dissertation describes the iterative development of the Science Inquiry Project from its inception 1999 to 2004. The second chapter is a systematic literature review of research studies available which examine the effects of inquiry interventions on high school, college and professional school students. The third chapter is a qualitative study that outlines the students' experiences of Inquiry during the winter 1999 semester.

The fourth through seventh chapters report on the program evaluation of Science and Health Science Inquiry courses for the academic years of 1999-2000 to 2003-2004. The fourth chapter describes the outcome measures that were developed based on the themes described in the quantitative study and also reports on the reliability and validity of these instruments used to measure these constructs. The fifth chapter describes the samples and methods used in the four years of the study. The sixth chapter reports what was found in the study while the seventh chapter discusses the implications of this study in the wider context of implementing inquiry interventions in higher education

References

Centre for Leadership in Learning. Proposals for the McMaster Inquiry Project. Working document last revised May 1999.

CHAPTER 2

Literature Review

According to Knapper and Cropley (2000), "nearly all universities claim to espouse life long learning goals, but their educational programmes, teaching methods and organizational structures often discourage lifelong learning" (p. xi). Candy, Crebert and O'Leary (1994) outline some major impediments to implementing active learning strategies that promote life long learning including: making extensive use of lectures; overloading the curriculum with too much detail; emphasising memorization of facts; using over-simplified and outdated models or theories; de-emphasising connections to skills or real-world applications; not giving appropriate and timely feedback; and not constructing assignments and learning activities that promote library usage (cited in Knapper and Cropley, 2000, p. 175).

Part of the reason there has been a slow adoption of active learning teaching practices were because of increasing faculty - student ratios and large class sizes which, faculty members often claim, makes it impossible to implement these practices. At McMaster University, the ratio of undergraduate students to full-time faculty persons has increased from 12 students to one professor in 1990 to 15.5 students to one professor in 1998 (McMaster University Planning Committee, 2000). Similarly, when examining the annual Macleans's rankings of Canadian universities, the number of first and second year classes with greater than 100 students per class was 30.02% in 1997 (1997, p. 50) but increased to 68% in 2003 (2003, p. 66): in both annual rating editions, McMaster was ranked 15th out of fifteen in the medical/doctoral category for class size.

The consequences of increasing enrollment with decreasing numbers of faculty available to students may be seen in a study published by a neighbouring university undergoing comparable social, economic and political pressures as McMaster. York University, which has large class sizes relative to other Canadian comprehensive universities (24% in 1997 and 64% in 2003 of greater than 100 students per class), conducted a survey to assess the 'value added' skills students gained during their university years. They surveyed of its students entering and graduating during the 1995-1996 academic year (n = 1306 entering, n = 2791 graduating with an approximate response rate of 51 to 58%) (Grayson, 1999, p. 119). The survey asked students to rate their abilities in domains such as communication (e.g., writing job application letters), analysis (e.g., identifying the main points in the lectures), personal (e.g., enjoying meeting new people, knowing themselves), organization (e.g., organizing priorities to prevent rushing at the last minute), comparative (e.g., comparing what is going on in Canada today to what was going on fifty years ago), basic numeracy (e.g., calculating a 15% discount on a $9.36 book), and computers (e.g., using a word processing program). Grayson linked these survey results to the students' secondary school marks and their graduating GPA. Value added to skills assessed in the survey were reported as standardized scores (z-scores: mean = 0, SD = 1.00) after controlling for confounding factors such as secondary school graduation grades, gender, home language and ethno-racial origin.

Grayson (1999) noted that there was a strong correlation between secondary school graduation marks and university cumulative GPA while the correlation between university cumulative GPA and perceived skills competence was weak. While students in all faculties tended to be below the mean for value-added skills at university entry, students in the Faculties of Business, Arts and Fine Arts tended to report improvement above the mean in all skill areas at graduation. However, Faculty of Science students, while they improved substantially in areas such as computer skills, organizing skills and basic numeracy, reported minimal gains and remained below the mean on analytic, communication, personal, and comparative skills (Grayson, 1999, p. 133). As Grayson noted, graduating science students' minimal gains for analytic, communication, personal, and comparative skills were unacceptable to his institution.

Since the late 1980's, there has been a call to implement active learning strategies into higher education teaching and learning, and a concomitant impetus for universities to graduate students with critical thinking, interpersonal and communication skills (i.e., Boyer Commission, 1998; National Research Council 1996; and the National Science Foundation, 1996). To achieve these objectives, these Committees advocate that students should have opportunities to work in small groups with peers to "have access to supportive, excellent undergraduate education in science, mathematics, engineering, and technology, and [to] learn these subjects by direct experience with the methods and processes of inquiry" (National Science Foundation, 1996).

Since this call to improve the quality of science education, numerous studies have examined the effects of these changes. Research in elementary and secondary schools has conclusively shown that small group, cooperative learning has a moderate positive effect on student learning outcomes (i.e., higher academic achievement) provided that the group was rewarded for the whole group performance (median Cohen's effect size d = +0.32 over 64 studies with control group comparisons) (Slavin, 1996, p. 46) and that there were group mechanisms whereby the group members established a common goal and the individuals were held accountable to their group for achieving that goal (median Cohen's effect size d = +0.32 over 52 studies with other teaching method group comparisons) (Slavin, 1996, p. 55). However, these meta-analyses apply to school age children who were mastering well-defined skills or knowledge (e.g., reading or mathematics skills, basic science or social science materials, etc.). There was less certainty about the efficacy of small group learning in higher education with self-directed tasks.

A meta-analysis by Springer et al., (1999), reviewed the effects of small-group learning curricula on science, mathematics, engineering, and technology undergraduates. They included studies that report on student groups (between two to ten students per group) working in or out of the classroom time in field conditions (i.e., not in experimental conditions). Based on 39 eligible studies, the authors reported higher academic achievement (mean Cohen's effect size of 0.51 representing 37 reports and 3,472 students), decreased attrition from the course or program (mean Cohen's effect size of 0.46 representing nine reports and 2,014 students), and more favourable academic attitudes (mean Cohen's effect size of 0.55 representing eleven reports and 1,293 students) when compared to control groups who did not experience collaborative or cooperative learning (pp. 29-30). However, by the authors' own admissions, with "relatively sparse descriptions of detailed teaching and learning practices in most studies" (p. 35) it was difficult to know the context with which these gains were most likely to occur. While it appeared that cooperative and collaborative learning has a positive effect on student learning outcomes, the question of *how* active learning strategies moderate students' cognitive and affective academic behaviours was less well established (Slavin, 1996, Springer et al., 1999).

While small group learning appears to increase students' achievement and motivation to learn (Slavin, 1996, Springer et al., 1999), there were fewer studies that report on the students' experiences using inquiry-based education strategies. There were numerous case studies reported in the literature by instructors that describe their attempts to replace laboratories and lectures with an inquiry curriculum. Consistently, these instructors reported that their students experienced an increase in self-monitoring skills, knowledge acquisition and retention, and improved classroom motivation (Arce & Bentancourt, 1997; Glasson and McKenzie, 1998; Ahern-Rindell 1999; Silivius and Stutzman, 1999; Tien et al., 1999; Henderson and Buising, 2001; Klionsky, 2001;

Kolkhorst et al., 2001; Lunsford, 2002; Westerlund and Stephenson, 2002; DiPasquale et al., 2003). However, these case reports, while informative, did not report comparison data (e.g., control groups, reference groups, pre and post test results, triangulated data, etc.) and were not systematic in their collection of empirical data.

Methods

In an effort to understand the impact of inquiry-based learning on student academic performance, I attempted to locate all of the studies that contained empirical findings. An electronic search of the following data bases (PubMed, ERIC, PsychInfo, Social Science Index, Sociological Abstracts and Social Work Abstracts) using the search term "inquiry" in the title was used to identify the appropriate subject headings for each data base. Once the appropriate subject heading was identified, the same databases were re-searched to obtain additional references. Dissertations were excluded since obtaining these works has proven costly and time-consuming.

A total of 411 unique reports concerning inquiry were identified; the majority of which were case reports discussing the practicalities of adopting inquiry methods in the classroom. Since 411 reports were deemed to be too unwieldy to review, the titles and abstracts were checked to determine if they contained any qualitative or quantitative reports on inquiry experiences for student samples. Based on these criteria, the list of eligible reports was reduced to seventy-nine studies. The subsequent reports were retrieved and reviewed by hand to determine if they were original studies. After this review, 32 reports were found not to be pertinent to this review: the majority of these reports were concerned with teacher behaviour and attitudes that had an impact on student adoption of inquiry or were not studies (e.g., reviews of studies). The remaining 45 studies did contain empirical data studying the impact of inquiry on students, with 21 reports using samples from primary to grade eight students and 26 reports using samples from grade nine and above. Upon careful review, it was decided that the subject populations most relevant to this systemic review were from high school, post-secondary or professional education levels since these samples were deemed to be most comparable to the students in McMaster Science Inquiry Project.

When reviewing these 26 papers, nine met the criterion but were not used for the following reasons: one did not report the cited comparison data; two focussed on student learning styles and not on the students' experiences of inquiry; and four had methodological problems (i.e., the outcome measure was the students' gains in learning about evolution but the control group had not been exposed to the evolution curriculum; students in the inquiry learning group were also exclusively above average learners while the control group were heterogeneous learners; the measures used to examine student inquiry performance were impossible to interpret). Of the 19 usable reports, six were qualitative reports and thirteen were quantitative reports. While this strategy may omit some reports, I believed that these papers were representative of the available empirical literature on Inquiry. The summaries of the data are provided below.

Results

Results of the Quantitative Studies: A series of meta-analyses reported to the National Association for Research in Science Teaching Conference in 1983 and conducted by the University of Colorado, examined the impact of the then 'New Science Curricula' on student learning outcomes. In the meta-analysis by Shymansky, Kyle and Alport (1983) examining the impact of the 'New Science Curricula' when compared to traditional teaching strategies on student learning outcomes such as achievement and process skills for American high school students. The average effect size was 0.39 (SD 0.44, 0.09 – 2.32) over 137 studies. The meta-analysis from the same research initiative Wise and Okey (1983) examined the impact of teaching various active learning strategies compared to traditional strategies (number of studies included = 160), and found that the effect size of inquiry/discovery teaching techniques on general achievement and problem solving was 0.41 (SD 0.87) over 39 studies while the effect size for critical thinking, creativity and affective measures was 0.15 (SD 0.29) over 20 studies.

McRobbie and Fraser (1993) examined the effects of classroom environment on student inquiry skills for senior high school chemistry students in Brisbane, Australia. During the chemistry lectures, 591 students in 76 classes (approximately seven students per classroom) in schools throughout Brisbane were asked questions regarding their classroom environment and inquiry skills (i.e., outcome measures defined as 'ability to make conclusions and generalizations from their work' and 'ability to design experiments'). In the same study, 596 students in 78 classes (approximately seven students per classroom) were asked questions about their learning environment and their attitudes toward laboratory learning (i.e., outcome measures defined as 'attitude to laboratory learning', 'nature of chemistry knowledge', 'cooperative learning and adoption of laboratory attitudes'). The data was analyzed in two separate analyses, one using the individual student scores, and the second using classroom mean scores. Both analyses controlled for students' general scholastic ability. The proportion of variance in chemistry knowledge and attitudes, explained by each inquiry factor, the variance component values ranged between 0.17 to 0.45 for individual differences and 0.41 to 0.60 the classroom differences. This study demonstrated "that students' perceptions of classroom psychosocial environments account for appreciable amounts of variance in student outcomes beyond that attributable to student characteristics such as general ability" (p. 83).

In the study by Gallagher (1994), path analyses were used to examine whether the effects of middle school science inquiry exposure had an impact on science persistence using data from the Longitudinal Study of American Youth cohort (n = 1166). The total model explained 46% of the variance, with grade seven science achievement having the largest direct impact as a factor contributing to grade 11 science completion (0.40), and inquiry-based instruction had the second largest direct impact (0.33). Other factors such as teacher enthusiasm (0.10), teaching stressing facts and principles in grade seven (0.06) and grade eight (0.06) also had direct impact on science persistence while parental education (0.25 indirect and 0.12 direct effect) and gender (0.06 indirect and 0.10 direct effect) also contributed to the model.

Chang and Mao (1999) studied the impact of an inquiry curriculum on junior high school students learning about the rotation of the earth on its axis in their earth science course. Six hundred and twelve Taiwanese grade nine students (mean age 15 years with an approximately equal split of male and female students) were studied over 16 earth science classes (approximately 38 students per class) over five schools with similar socio-economic and ethnic backgrounds. Using a random allocation procedure, eight classes (n = 319 students) were assigned to the inquiry curriculum where small groups of students collaborated on working on an equinox scenario to generate a data driven solution and present their findings to the class. The other eight classes (n = 293 students) were assigned to a lecture-based curriculum where students were provided with clear and detailed instructions and explanations of the equinox. However, both groups were exposed to the same assignments and materials during the four weeks of the intervention.

Students were measured on their knowledge of earth sciences (i.e., knowledge, comprehension and attainment) and attitudes towards teaching styles (i.e., class involvement, confidence on the subject matter, and learning interest in earth science) prior to the start of the intervention and again at the end. To control for the nesting of students within classrooms, classroom means were used as the unit of analyses. Using pre-test scores as a covariate, Chang and Mao (1999) found that inquiry students scored significantly higher on their knowledge of the material but their comprehension and application of the knowledge was not significantly different than the lecture groups. Also, inquiry students' attitudes were significantly improved from the students in the lecture format for their involvement and confidence in the material but not for their interest in learning the subject.

Von Secker and Lissitz (1999) examined the effects of science instructional practices (i.e., teacher-centred instruction, emphasis on critical thinking and emphasis on laboratory inquiry) on 2,018 grade ten students' science achievement scores over 163 schools using data from the High School Effectiveness Study that was part of the second wave of the National Education Longitudinal Study. Achievement scores were derived from the students' responses to higher order thinking and basic knowledge questions in biology, earth science, physics and chemistry subjects developed by the Educational Testing Service. Teaching style was determined by the teachers' descriptions of teaching and only those schools having more than four students were used. Using hierarchical linear modelling to control for the clustering of students within schools, teacher centred instruction was found to have a negative impact (effect size of –0.47) on student achievement whereas emphasising laboratory inquiry has a positive impact on student achievement (effect size of 0.39). When examining the indirect effects of socio-economic status, gender, and minority status, instruction which emphasized inquiry based laboratories, "was invariably associated with higher achievement overall and with more equitable achievement among students with different demographic profiles" (p. 1121).

Zoller (1999) reported on his efforts to introduce inquiry style strategies into a second year organic chemistry class in a Canadian university (n ~ 150 to 180). He compared the exam results for this group to the examination scores of undergraduate Israeli biology majors who were in small group inquiry classes, had completed their military service, and were focussed on becoming high school science teachers (n ~ 120). Zoller's inquiry intervention for the organic chemistry asked students to work in small groups during the lectures to consider what could be the possible reactions of given reagents and solvents as opposed to working through each step by the instructor. While the students in the organic chemistry class had mixed reactions (i.e., some were frustrated because he did not tell them specifically what they needed to know for the exam while others appreciated the opportunity to work through the questions themselves), the biology majors expressed little difficulty with this style of learning. When comparing student performance for both groups, class averages on the inquiry type examinations were the same or slightly better for both the Canadian and Israeli students when compared to their class averages on department established examination scores.

Magnussen, Ishida and Itano (2000), measured the impact of an inquiry-based curriculum on student nurses' ability to think critically. Over fifty percent of the curriculum used small group, inquiry style learning that focused on holistic, interdisciplinary, self-directed learning. They found that student nursing scores on the Watson Glasser Critical Thinking Appraisal test decreased at graduation when compared to their scores at entry (mean score of 56.35 at entry and 55.35 at graduation). However, comparisons were made using an independent sample t-test; it was hard to interpret this result and, more fundamentally, the Watson Glaser Critical Thinking Testdoes not measure the attributes that the curriculum espoused to teach the students.

Hofstein, Levy Nahum and Shore (2001) studied the impact of inquiry style laboratories compared to "cookbook" style laboratories on high school students. Hofstein et al. were interested in the degree to which the students in both groups reported convergence between their preferred learning style and their actual laboratory experience. The initial stage of the project involved developing about fifty inquiry style laboratories where students would work to "peer tutor in small investigative groups" (p. 200): the students were to work at their own pace and determine which resources were relevant to define hypotheses, perform relevant experiments, document the data collected and analyze and document the data to form cogent conclusions. Instructors were then introduced to the new inquiry laboratories by working in small instructor groups to pre-test and revise the laboratories: also, the instructors were prepared for the type of open-ended and unintended results that inquiry style laboratories could produce. These laboratories were subsequently incorporated into the grades 10 through 12 Israeli chemistry curriculums during the years 1997 to 2000. The "cookbook" style laboratories were from the current curriculum and were described as very task oriented with students not having the time to deviate from the prescribed order when completing the experiments.

When evaluating the congruence of the students' preferred learning environment to the one they actually experienced, students were asked to complete a previously developed (i.e., McRobbie et al., 1993), 140 item survey adapted for this study (Science Laboratory Environment Inventory) which asked the students about their learning experience (70 items) and what they would like to have in a laboratory environment (the same 70 items) (Hofstein et al., 2001). Of the 312 students, 129 students were in the inquiry laboratories and 183 were in the traditional laboratories. The inquiry students showed greater congruence between what was preferred and experienced than the traditional laboratory students. In particular, the inquiry students experienced a smaller gap in their involvement, open-endedness and teacher involvement of the laboratories: however, there was the same degree of congruence for the other elements such as the degree of knowledge integration, organization, and materials available in the environment.

Russell and French (2001) studied how pre-college level students experienced "cookbook" compared to inquiry-based laboratories and related that to their performance in and attitudes towards the course. Using a mixed methods approach, the authors studied 145 students over three different semesters and recorded the amount of time at specific on task and off task behaviours of the students during the first 10 minutes of the labs (this coding scheme had been developed from a previous semester's observations), an attitude and biology content survey administered during the same semester of observation. To validate their coding scheme to their current year of study, the authors' audiotaped 16 students who had participated in the observations.

Russell and French (2001) found that while women in the traditional labs were less likely to discuss the findings or attempt the lab procedures than men, there were no differences between males and females in the inquiry-based labs. Since all group members discussed how to proceed with testing their hypotheses for the lab in the inquiry-based style, women tended to report that they felt that all members were participating equally and the roles were more defined. In the traditional labs, men were more likely to get started with the lab and the women reported feeling intimated to participate. In contrast, inquiry lab students reported more of a benefit from group discussions prior to beginning the labs since they could receive corrective feedback from their peers in a supportive environment. In the inquiry based labs with the students developing and testing their own hypotheses, the students reported feeling that they were trying to understand the scientific concept more instead of trying to get the correct answer. Students from both the traditional and inquiry laboratories improved although the inquiry group students experienced greater improvement.

Schneider, Krajcik, Marx and Soloway (2002) examined the difference between their students who were in project based science inquiry curriculum and national standardized science test scores. Grade ten (n = 85) and eleven (n = 57) students at a small, alternative school (which did not have an accelerated curriculum or an enrollment limited to 'exceptional' students) in white, middle-class suburban Michigan were enrolled in a project-based science curriculum which encouraged self-direction, collaboration, and conversation to construct ideas in "what is science" and find solutions to real science problems. The National Assessment of Educational Progress (NAEP) tests for grade twelve students were used to test earth, physical and life sciences knowledge though multiple choice, scenario-based and demonstration problems. These results were then compared to the national means for white students who were not eligible for school lunch programs. Although the NAEP was intended to assess grade twelve student performance, the grade 10 and 11 students in this study outscored the national sample on 44% of the items. More specifically, the inquiry students had medium to large effect size differences (0.30 to 1.58) on 33 % of the items that required conceptual understanding and scientific investigation abilities when compared to national groups of similar socio-economic status and ethnicity.

Myers and Burgess (2003) designed an intervention to improve their physiology students' ability to design experiments and interpret data. Students in their Organismal Biology course could elect to enrol in the lecture only course (3 credits) or the lecture and lab course (5 credits). Course groups were compared for their previous cell biology course grades and on a pre-test evaluation tool and were found to be similar in both these measures. The curriculum in the lab component emphasised designing and interpreting the results of their experiments. For the pre-tests and post-test evaluation tool, students were asked to generate data, graph their data, interpret the relationships in their data, develop recommendations for the patient, and suggest improvements/alternatives to the experiment. An independent scorer evaluated all of the responses but was blind to the condition and the time point. The results showed that there were no significant differences between the groups on their ability to create a graph, interpret the data and make recommendations to the patient; there were significant differences on the students' ability in the lab section to make improvements to the study design.

In a study that was most similar to the McMaster Inquiry Project, O'Sullivan and Copper, (2003) reported on a comparison of two educational interventions, active learning vs. traditional lecturing, in a required first year general chemistry course for the United States Naval Academy. In this study report, there were approximately 1200 students instructed by about 20 instructors yearly and this experiment was repeated over four semesters. Midshipmen (undergraduate students) were assigned to two conditions: the active learning group which used problem-solving worksheets, creative testing strategies, hands-on learning activities, student presentations, games and group work to stimulate active learning (n of approximately 200 per semester); while the traditional condition continued with a primarily lecture based curricula with some active learning techniques (n of approximately 1000 per semester). Over the course of four semesters, the teaching team rotated leadership and faculty from the traditional group switched to the active learning teaching team to learn about the new teaching techniques.

There were no significant pre-intervention differences in SAT scores between students in the two teachings styles for all semesters. The outcome measures (i.e., mid-course and final exams and the final course grade) showed significant differences in the first semester but not for subsequent semesters. However, there were two difficulties with this study. First, the outcome measures used were not designed to tap into the domains originally specified to keep; namely, having a more positive attitude toward the course materials and encouraging students to "formulate their own ideas, draw conclusions from experimental evidence and participate in other similar activities" (O'Sullivan and Cooper, 2003, p. 1). Second, there were no restrictions on the type of instruction provided by the traditional learning group. Hence, with instructors rotating from the active learning group into the traditional lecture group in the second to fourth semesters, the difference between teaching styles may have been diluted because the instructors may have used more active learning techniques in their traditional classes. However, the significant findings from this study were that, while active learning instructors felt that the students covered less material in their courses, the active learning students scored as well as or better than the traditionally taught students. Also, the instructors felt that team teaching provided them with opportunities to learn more about teaching and experience more support for their innovative efforts.

Table 2.1: Summary of Small Group, Active Learning Compared to Traditional Teaching Methods Meta-analyses Reported in this Review

	Slavin, 1996	Springer et al., 1999	Shymansky, et al., 1983	Wise and Oakley, 1983	Weighted Average Score by attribute
Higher Student Academic Achievement (effect size)	ES = 0.32 n = 64	ES = 0.51 n = 37	ES = 0.39 n = 137	ES = 0.41 n = 39	ES = 0.39 n = 277
Independent Student Goal Setting for the Group (effect size)	ES = 0.32 n = 52		ES = 0.39 n = 137		ES = 0.39 n = 189
Decreased Student Attrition from school (effect size)		ES = 0.46 n = 9			ES = 0.46 n = 9
Improved Student Academic Attitudes (effect size)		ES = 0.55 n = 11		ES = 0.15 n = 20	ES = 0.29 n = 31

NB: The n reported is the number of studies this effect size represents. Weighted effect size is based on the effect sizes reported in this table divided by the total number studies. These results represent a crude estimated of the effect sizes over these meta-analyses.

Table 2.2: Summary of Major Findings of Inquiry Based Methods Compared to Traditional Teaching Methods from the Eleven Reports with Single Studies in this Review

Author	McRobbie and Fraser, 1993	Gallagher 1994	Chang and Mao, 1999	Von Secker and Lissitz, 1999	O'Sullivan and Cooper, 2003	Zoller, 1999
Sample	High school 76 - 78 classes n = 1187 students	Grade 11 n = 1166 students	Grade 9 16 classes n = 612	Grade 10 163 schools n = 2018	Students in first year chemistry n = 1200 students per year over four years	Undergraduate chemistry students 2 classes n ~ 150
Research Design	One group	One group using grade 7 performance to predict grade 11 individual outcomes	Randomly assigned classes to compared two groups' pre and post test scores	Compared three teaching styles as identified from teachers' descriptions of teaching	Compared traditionally taught classes to inquiry classes	Compared class results to university department-wide tests
Knowledge Acquisition during the Intervention			++		= / +	= / +
Improvement on Scientific Process Skills	++		=	++		
Student Improvement on Science Attitudes	++	+	=			
Improvement in Student Affective Attributes			++			

Legend: ++ good improvement from inquiry, + some improvement from inquiry, = about the same between two conditions, - decrease in skills in inquiry

Table 2.2 (con't): Summary of Major Findings of Inquiry Based Methods Compared to Traditional Teaching Methods from the Eleven Reports with Single Studies in this Review

Author	Magnussen et al., 2000	Hofstein et al, 2001	Russell and French, 2001	Schneider et al., 2002	Myers and Burgess, 2003
Sample	Nursing Students	High School n = 312	Undergraduate science students N= 145 Over three classes	Grade 10 and 11 N = 142	Undergraduate students
Research Design	Compared WGCAT scores at enrolment and at graduation	Compared two groups: traditional and inquiry conditions	Compared differences between males and females in traditional and Inquiry laboratories	Compared inquiry student results to National Testing Results	Compared two groups on pre and post test measures
Knowledge Acquisition during the Intervention		=	=	++	+
Improvement on Scientific Process Skills	=		++	++	=
Student Improvement on Science Attitudes		+	++		=
Improvement in Student Affective Attributes			++		

Legend: ++ good improvement from inquiry, + some improvement from inquiry, = about the same between two conditions, - decrease in skills in inquiry

In short, these quantitative reports demonstrated that there was a low to moderate but consistent student improvement in science achievement, attitude, and process skills when comparing inquiry-based interventions with traditional lecture-based interventions. Also, these authors, when designing their interventions, make the fundamental assumption that inquiry skills were learned by interacting with others either in dyads or in groups. Also, although Inquiry learning covers less of the syllabus materials, there was no difference in student knowledge scores.

Results of the Qualitative Studies: All of the six qualitative papers examined appeared to describe the ways in which students made meaning of the process of inquiry. Roth and Roychoudhury (1993) looked at how middle and high school students (n = 60 grade eight, n = 41 grade 11, and n = 29 grade 12 students) developed more precise language and experimental laboratory skills through the process of working with others in an open-ended laboratory experience where the students developed their own hypotheses. Ritchie and Rigano (1996) examined how two high school biology students, with the assistance of a PhD candidate as a mentor, become accustomed to designing and interpreting their own experiments in a university laboratory setting. Hammer (1997), through a series of personal reflections as the instructor to a group of high school physics students (n = 22 students), examined the process of knowing when to intervene as students were constructing their own explanations of scientific phenomena. Harrison, Grayson and Treagust (1999), examined how five grade 11 boys made conceptual changes to how they thought about notions of ‘heat’, ‘temperature’, and ‘thermal interaction' over 20 sessions of inquiry-based classroom discussions and laboratories. Kaartinen and Kumpulainen (2002) examined how teacher training students made sense of dissolving problems in "kitchen chemistry" (e.g., the differences in dissolving sugar, salt and flour, etc. in water) when working in groups (n = 18). Hmelo-Silver et al. (2002) examined how expert physicians/clinical trialists (one group with n = 4) made decisions when designing a phase II clinical trial using a computer simulation program when compared to four groups (n = 24) of fourth year medical students. All of these papers used dialogue analyses of the participants' interaction: these were documented in a variety of media including student or instructor reflective journals; audio or videotapes of the interactions; and key informant interviewing.

These reports documented that participants of all ages developed the higher order scientific process skills of identifying pertinent variables; generating more explicit and focused hypotheses; and analyzing and interpreting data with increasing specificity. Although these studies were working with individuals at different developmental stages and varying levels of aptitude and expertise, the gains in scientific process skills appeared to be context specific and underwent a similar process within a co-operative learning environment. Roth and Roychoudhury (1993) summarized the themes of their qualitative study as six assertions:

1. The identification of pertinent variables in a particular context increases with familiarity of the student in a specific physical and conceptual context (p. 133).

2. Student interpretations of experimental results evolved from simplistic and insufficiently supported statements to identification of complex relationships among the variables using multiple representations of their experimental data (p. 135).
3. The pursuit of problems of genuine interest motivated students to generate new hypotheses based on the interpretation of previous results. These hypotheses led to new focus questions for subsequent experiments (p. 138).
4. Students learn to define concepts, events, and actions in order to be able to design experiments and communicate results effectively (p. 140).
5. Given the freedom to choose research topics and to design their own experiments, students became very adept in planning and carrying out their experiments (p. 141).
6. Learning in an open inquiry lab over a period of 14 months, students developed highly competent integrated science process skills (p. 142).

The previous themes were echoed in the identified themes of Ritchie and Rigano (1996) but were phrased slightly differently: that "laboratory skill development preceded conceptual understanding" of the scientific materials (p. 805); that "blind alleys provided opportunities to practice skill, but increased frustrations and revived memories of fudging in school cookbook labs" (p. 808); and, "students developed into independent researcher during the project (p. 810). Hmelo-Silver et al. (2002) noted that novices and experts got to the same end point in designing their trial but took very different routes to get there. While experts were more fluent in their knowledge of how analogous drugs behaved and in the requirements of a phase II trial, novices had to learn what a phase II trial was and were more rule bound. However, novices were more persistent and interested in the task than the experts. Hmelo-Silver et al. (2002), Harrison et al., (1999), and Kaartinen and Kimulainen (2002), noted that collaborative learning experiences when students worked with authentic materials allow students opportunities to negotiate and clarify scientific meaning to come to an agreed upon understanding.

Hammer started his article by articulating the teaching dilemma of "how do you handle the intrinsic tension between students having to learn the requisite material yet also feel empowered to follow their own learning path? How do you handle the students coming to the wrong conclusions when they have been designing their own experiments and observing their experiences faithfully?" (Hammer, 1997, p. 513). In Hammer's (1997) view, "a curriculum succeeds not by guiding the flow of learning and instruction but by helping to establish an arena of activity rich with opportunities for student and teacher discovery" (p. 514). In this view, the role of the teacher was to scout ahead on the path of inquiry to determine if the students were making appropriate connections to learning objectives of the course. It was the teacher's responsibility to balance the sometimes contradictory but important tension between the students' inquiring to engage in the material and gaining a sense of what were the appropriate knowledge goals. However, a more important goal was for both the teacher and student to recognize, through their mutual process of discovery and the resultant fits and starts, messiness, and contradictory findings, that currently accepted knowledge must be held as uncertain or needs to be critiqued as part of the process of good scientific understanding.

Summary and Conclusions

Collaborative, inquiry-based learning, helped students become more proficient in developing a more realistic and flexible understanding of science concepts. Students, from high school to professional school, were shown to have gains in scientific processing skills while maintaining their content knowledge scores in an inquiry curriculum when compared to a traditional lecture curriculum. When placed in an authentic science context, students also appeared to make gains in their analytic hypothesis testing skills and developed a greater tolerance for ambiguity when developing their experimental designs.

References

Ahern-Rindell, A. (1999). Applying inquiry-based and cooperative group learning strategies to promote critical thinking. Journal of College Student Teaching, December, pp. 203-207.

Arce, J. & Betancourt, R. (1997). Student-Designed Experiments in Scientific Lab Instruction. Journal of College Science Teaching. 27(2): pp. 114-118.

Boyer Commission on Education Undergraduates in the Research University. (1998). Reinventing undergraduate education: A blueprint for Americas research universities. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.

Chang, C., & Mao, S. (1999). Comparison Taiwan Science Students' Outcomes with inquiry-group versus traditional instruction. The Journal of Educational Research, 92(6), pp. 340-346.

DiPasquale, D.M., Mason, C.L., & Kolkhorst, F.W. (2003). Exercise in Inquiry: Critical Thinking in an Inquiry-Based Exercise Physiology Laboratory Course. Journal of College Science Teaching. 32(6): pp. 388-393.

Gallagher, S.A. (1994). Middle school classroom predictors of science persistence. Journal of Research in Science Teaching, 31(7), pp. 721-734.

Glasson, G.E., & McKenzie, W.L. (1998). Investigative Learning in Undergraduate Freshman Biology Laboratories. Journal of College Science Teaching. 27(3): pp. 189-193.

Grayson J. (1999). Using surveys to measure value added in skills in four faculties. Canadian Journal of Higher Education, 29(1), 111-142.

Hammer, D. (1997). Discovery Learning and Discovery Teaching. Cognition and Instruction, 15(4): pp. 485-529.

Harrison, A. G., Grayson, D. J., & Treagust D. F. (1999). Investigating a grade 11 student's evolving conceptions of heat and temperature. Journal of Research in Science Teaching, 36(1): pp. 55-87

Henderson, L., & Buising, C. (2001). A Research-Based Molecular Biology Laboratory. Journal of College Science Teaching, 30(5): pp. 322-327.

Hofstein, A., Levy Nahum, & Shore, R. (2001). Assessment of the learning environment of Inquiry-type laboratories in high school chemistry. Learning Environments Research, 4, 196-207.

Hmelo-Silver C.E., Nagarajan, A., & Day, R. (2002) "It's harder than we thought it would be": A comparative case study of expert-novice experimentation strategies. Sicence Education 86(2), pp. 219-243.

Kaartinen, S., & Kumpalainen, K. (2002). Collaborative inquiry and the constructions of explanations in the learning of science. Learning and Instruction, 12, pp. 189-212.

Klionsky, D.J. (2001). Constructing Knowledge in the Lecture Hall: A Quiz-Based, Group-Learning Approach to Introductory Biology. Journal of College Science Teaching, 31(4): pp. 246-251.

Knapper C., & Cropley A. (2000). Lifelong learning in higher education (3rd ed.). Sterling, VA: Kogan Press.

Kolkhorst, K., Mason C., DiPasquale, D., Patterson, P., Buono, M. (2001). An inquiry-based learning model for an exercise physiology laboratory course. Advances in Physiology Education, 24(2), p[. 45-50.

Lunsford, E. (2002). Inquiry in the Community College Biology Lab: A Research Report and a Model for Making it Happen. Journal of College Science Teaching, 32(4): pp. 232-235.

Magnussen, L., Ishida, D., & Itano, J. (2000). The impact of the use of inquiry-based learning as a teaching methodology on the development of critical thinking. Journal of Nursing Education, 39(8), 360-364.

McMaster University Planning Committee (2000). McMaster University Academic Plan 2000 - 2005. [online] http://www.mcmaster.ca/newsevents/acadplan.htm Date of Access: March 8, 2004.

McRobbie, C.J., & Fraser, B.J. (1993). Associations between student outcomes and psychosocial science environment. Journal of Educational Research, 87(2), pp. 78-85.

Myers, M.J., & Burgess, A.B. (2003). Inquiry-Based Laboratory Course Improves Students' Ability to design experiments and interpret data. Advances in Physiology Education, 27(1), pp. 26-33.

National Research Council (1996). From analysis to action: Undergraduate education in science, mathematics, engineering, and technology. Washington, DC: National Academy Press.

National Science Foundation (1996). Shaping the Future: New expectations for undergraduate education in science, mathematics, engineering, and technology. Washington, DC: Report by the Advisory Committee to the National Science Foundation Directorate for Education and Human Resources. Date Accessed: March 8, 2004. URL http://www.ehr.nsf.gov/ehr/due/documents/review/96139/start.htm

O'Sullivan, D., & Copper, C.L. (2003). Evaluating active learning: A new initiative for general chemistry curriculum. Journal College Science Teaching, 32(7), pp. 448-452.

Ritchie, S.M., & Rigano, D.L. (1996) Laboratory Apprenticeship through a Student Research Project. Journal of Research in Science Teaching, 33(7), pp. 799-815.

Roth, W.M., & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. Journal of Researhc in Science Teaching, 30(2), pp. 127-152.

Russell, C.P., & French, D.P. (2001). Factors Affecting Participation in Traditional and Inquiry-Based Laboratories: A Description-Based on Observations and Interviews. Journal of College Science Teaching, 31(4): pp. 225-229.

Schneider, R.M., Krajcik, J., Marx, R.W., & Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of achievement. Journal of Research in Science Teaching, 39(5), pp. 410-422.

Silvius, J.E., & Stutzman, B.C. (1998/1999). A Botany Laboratory Inquiry Experience: Investigating the Effect of Soil Resistance on Bean Seedling Emergence. Journal of College Science Teaching, 28(3): pp. 203-207.

Slavin, RE. (1996). Research on cooperative learning and achievement; What we know, what we need to know. Contemporary Educational Psychology, 21(1), 43-69.

Shymansky, J.A., Kyle, W.C., & Alport, J.M. (1983). The effects of new science curricula on student performance. Journal of Research in Science Teaching, 20(5), 387-404.

Springer, L., Stanne, M.E., & Donovan, SS. (1999). Effects of small group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. Review of Educational Research, 69(1), 21-51.

Tien, L.T., Rickey, D., & Stacy, A.M. (1999). The MORE Thinking Frame: Guiding Students' Thinking in the Laboratory. Journal of College Science Teaching, 28(5): pp. 318-324.

Von Secker, D.E., & Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. Journal of Research in Science Teaching, 36(10), pp. 1110-1126.

Westerlund, J.F., & Stephenson, A.L. (2002). Phasing in Future Teachers: A Multidisciplinary Teaching Technique. Journal of College Science Teaching, 32(3): pp. 171-175.

Wise, K.C., & Okey, J.R. (1983). A meta-analysis of the effects of various science teaching strategies on achievement. Journal of Research in Science Teaching, 20(5), 419-435.

Zoller, U., (1999). Scaling-up of higher-order cognitive skills-oriented college chemistry teaching: an action-oriented research. Journal of Research in Science Teaching, 36(5), pp. 583-596.

[No Author] (November 17, 2003) Universities 2003: Measuring excellence an insider's guide. Maclean's, pp. 26-80.

[No Author] (November 24, 1997) Universities 97: The seventh annual ranking. Maclean's, pp. 28-82.

CHAPTER 3

The Initial Qualitative Study of The McMaster Science Inquiry Project: Winter Semester 1999

Initially, the Science Inquiry course was developed from a loose collaboration between the Centre for Leadership in Learning (Dale Roy), the instructors who taught the course in its first year (Del Harnish, Miroslav Lovric) and the Dean of Science's office (Peter Sutherland). Additional support came from research data that provided the instructors with the students' perspective (Kristina Trim). The overarching goal of the McMaster Science Inquiry Project had been established in the initial proposal to the McConnell Foundation as "to inculcate the concept of 'student as active learner' throughout the entire undergraduate experience. Our undergraduates should not learn primarily as passive recipients of information, but as active participants in their education in order to better develop critical thinking and communication skills" (Centre for Leadership in Learning, 1999).

In the previous academic year, the Inquiry course had been offered to students as a non-credit elective. However, the instructors were dissatisfied with the course since they felt the students did not see its value and put in less effort compared to other credit courses that demanded a fair amount of their attention. Prior to the winter 1999 semester, there was some uncertainty about how to best design the course. The instructors were asking themselves: what would be the most appropriate vehicle for first year students to acquire research skills; what was an appropriate skill set for this level of student; and, what was the role of the instructor in working with this group of students in this instructional format. The instructors realized that learning through inquiry was 'messy': they were limited in how much course structure they could implement with the students in the beginning of the semester; they had to be responsive to the students' needs as they arose; and they learned "on the job" how best to meet those needs. Most importantly, students could not be given a week by week course outline at the beginning of the semester since it signalled to the students that the instructor was setting the learning objectives and the students were not truly self-directed. As a result of these concerns, the instructors felt that taking a flexible approach to course design would allow them to respond to the students' needs as they were discovered.

At the beginning of January 1999, I approached the Science Inquiry instructors to conduct a formative evaluation of the course: instructors in three of the five sections consented to have their course evaluated (albeit one section's instructors were wary of the process). The course under study for this report is the Science Inquiry offered to first year, Natural Science program students in the winter term of 1999. The course format was in an experimental phase but several new elements were introduced at this time. These elements included having the professors work as facilitators with inquiry-based, small groups; the incorporation of computer-based discussions via a computer-mediated conferencing system called LearnLink (a customization of the First Class Software); and having all sections use the common Inquiry topic of global warming. The topic of global warming was chosen since none of the professors were experts in the field and they would be in a learning process with their students. Students would not rely on their professors to "have" the answers, but rather help the students learn research skills. Since the instructors were attempting to be as flexible as possible to experiment with the strategies to meet the students learning needs, I elected to conduct a qualitative study to best describe the students' experiences of the course.

Methods

Convenience Sampling: Information from a total of 33 students in two sections of the course was evaluated using the data collection techniques outlined below. The information from the third section was not evaluated for two reasons: the instructors were ambivalent about having their students recruited for focus groups or permitting me to read the LearnLink message postings; and the method of instruction in the course used many strategies that were not consistent with the open-ended small group inquiry format that was being used in the remaining sections. However, the data from the two course sections are probably the most appropriate to study since these instructors had an informal collaboration about implementing the goals of the course. As a result, this similarity ensured a homogeneous sample and facilitated comparisons between the students' and researcher's experiences. However, the number of students in these sections represents 55% of the total enrolment of Science Inquiry for 1999.

Data Collection: Data was collected in the following manner:

1. Participant observation: I enrolled in the course as a student and was subject to the same course requirements, including grading, as the students were. However, it was important to strike an appropriate balance between being enrolled as a student who was registered in an undergraduate science program and being clearly a 'mature' student with previous university degrees. I was able to fashion a role whereby I acted as a peer tutor, which meant that I could describe my

experiences and offer my thoughts as to what we may wish to do next or model research behaviours (e.g., how to conduct a literature review, offer to post meeting minutes on LearnLink after the first group meeting), but I would "sit on my hands" and deliberately not take a leadership role in my small group. In the larger class group, I would only tentatively offer my opinions when I felt that the class had tried to make sense of the material but were 'stuck' in knowing what other possible alternatives they could pursue. Queries about my status in the course by my class mates were met with the honest but brief explanation that I was a student enrolled in the course who was getting a credit but that I was also conducting research on how students experienced the course since it was in an experimental stage.

2. Focus Group Interviews: At the end of the semester, I posted a message on LearnLink to invite students to participate in focus group interviews. Six students (6/15 or 40%) in one section and nine students (9/18 or 50%) in another section agreed to participate in group interviews conducted during the students' regular class time. Although the instructors encouraged the students to participate and, in one section, introduced me to students who did not know me, they were not present for the interview. The interviews were audio taped with the students' written informed consent and each lasted for about one hour. Although there was an interview guide students were asked open-ended questions to compare their initial expectations of the course with their current understanding. The researcher interviewed in a manner consistent with qualitative methodologies, which included using a conversational style as much as possible, asking open-ended

questions and attempting to let the respondents talk about their experiences in their own words. Also, the researcher confirmed emerging themes by offering content from previous interviews and her own reflections. The addition of previous students' comments and the researcher's tentative notions and insights helped generate some validity of the patterns and themes that were developing.

3. Key Informant Interviewing: Participating instructors were interviewed informally on an "as-needed" basis. The goals of these interviews was to gain some insight into the instructors' pedagogical goals for the project and also to inform them of some of the general issues and concerns that some students may have expressed during class time or during the focus group interviews.
4. Content Analyses of computer-mediated discussion group text: All text messages available on the course messaging system, LearnLink, were reviewed for their relevant themes.

Data Analysis: Qualitative methodologies look at 'what is' in a situation and focus on how people 'make sense' of situations (Glaser and Strauss, 1967; Loftland, 1971; Morgan, 1997; Strauss and Corbin, 1998; Morse and Richards, 2002). Data analysis compares subjects or cases to determine how they fit together and how they differ (Glaser and Strauss, 1967; Loftland, 1971; Morgan, 1997; Strauss and Corbin, 1998; Morse and Richards, 2002). The collection of data, coding, data analysis and theme generation happened simultaneously in this study (Glaser and Strauss, 1967; Loftland, 1971; Morgan, 1997; Strauss and Corbin, 1998; Morse and Richards, 2002). The strong suit of this approach is an "orderly presentation rich in descriptive detail" (Loftland, 1971, p. 59) that allows the researcher to identify themes based on an intimate understanding of the material. Focus group interviews were listened to two or three times by the researcher (KT) and then transcribed verbatim. Notes from interviews and from reviewing the LearnLink messages were also generated and reviewed throughout the data collection process to generate tentative themes and notions. These tentative findings were introduced in subsequent respondent interviews to refine the developing themes.

Multiple data collection methods were used to triangulate data sources thereby improving the reliability and validity of the findings. In addition, data collection and analysis were conducted simultaneously with findings being reviewed with involved faculty members and students so that emerging themes could be validated directly by participants. This continuous process of data collection, data analysis and theory generation allowed the researchers to develop a coherent explanation of events.

Results

Theme One: Inquiry Provided Students with Academic Social Support Systems

Analysis of the data illustrated that first year university students often struggle with their adjustment from secondary school to university. In fact, some students felt that Inquiry was an opportunity to reconstruct social support systems, available to them in high school, which had facilitated their previous academic success. These supports included getting to know their classmates and professors in small group settings and in computer-based discussions. Many students also cited wanting to learn **how** to research as a reason for taking Inquiry: they saw the opportunity to work in inquiry-based, small groups as an opportunity to develop these skills. For example, some students talked about the advantages of group work in the following passages:

"Yes, this course definitely stimulated my interest in group work. I find it much better than regular 50 minute lectures. I discovered that you can learn much more when you have 'more heads than one'. All last semester it was just me and my textbook, now I see that working in groups is not only more fun, but is also very helpful."

"We had several group activities. These were also effective because they taught me the importance of working in groups and the importance of self-discipline and trust. It is hard to trust all of your group members and it isn't possible to cover up for someone who isn't doing their work. I learned a lot about cooperation and team work."

As for getting to know their professors, students talked about the following:

"(F)irst term, I didn't go to any of my profs, I didn't know any of them. But the second term, I don't know, cause I saw the relationship with (the professor). I just thought professors were totally mean and wouldn't help me but, ya know (laughs). But after (this professor), I think it's all right, I would approach my profs now."

Other students stated the following from the focus group interview:

"Focus Group Student One: He's really organized with what he's doing, like he knows (the subject) and doesn't overwhelm us.

Focus Group Student Two: Yeah, and he understands you have other courses, that they might be another priority and he understand that, and he doesn't pile it on, and make something due on the same day that you have a test or anything like that. He'll kind of rearrange the due dates to meet your schedule.

Focus Group Student Two: He's flexible

Focus Group Student One: Yeah, that's it.

Focus Group Student Two: Very flexible."

The ability to interact with fellow students and professors appeared to facilitate learning because, by the students' admissions, they felt more confident in their ability to learn. In addition, interacting directly with professors who are accomplished researchers, but unfamiliar with the substantive area of global warming, may have given the student role models to augment their internalized notions of reasonable and competent research behaviours. Unlike the knowledge transmission model of their lectures and text books, where the content is usually delivered stripped of its research context, these professors modelled the uncertainty, the rough guesses and debate that the students experienced in their own small group research process while also showing the students their heuristics and "rules of thumb" for making headway in learning a new topic.

Many students stated that they enjoyed using computer-based discussion groups on LearnLink. Students also developed computer literacy skills since they were required to submit personal and group evaluations electronically. Some advantages to LearnLink cited by the following three student quotes include the following:

"I found that LearnLink was useful because it maintained a consistent connection between everyone in the course. We got updates and were able to share our ideas. It is like having a continuous discussion for four months straight. It (wa)s fun."

"I got most of my assistance from other people in class and from Learn Link"

"Getting more acquainted with Learn Link was a good source to help each other with finding resources."

The most frequent comment about LearnLink was the students enjoyed using it because they could be social with their peers and their professors. One group of students exchanged quotes with each other on topics unrelated to their group project. Other students cited that being able to read messages from other students, which addressed issues with which they were personally struggling, made them feel more confident in tackling their own problems. It may be that having a forum to exchange ideas with other students facilitated individual learning indirectly, since the students had an opportunity to normalize their fears and expectations of themselves and others.

Theme Two: Students Preferred Active Learning Strategies to Passive Learning Strategies

As stated in the previous section, students enjoyed engaging with fellow students on the computer-mediated discussion group LearnLink because they were able to engage with their peers outside of class time. Students also made it apparent that they preferred and learned more when they could actively engage with the materials. For example, one section had their professor personally lead tours of the library with the objective that each student would search out one journal article for their topic during that session. The students who participated in these sessions later claimed that they would be able to find a journal article independently during the focus group interviews. In contrast, in the second section, the entire class attended a presentation by library staff that had no hands-on experience for the students. The students in the second section overwhelmingly stated that they found the session to be frustrating since they could not use the library's search capacity for themselves. Fewer students felt they could find a journal article on their own than in the first section. Many of the second section students spontaneously stated during the focus group interview that the library tour would have been more effective if it had been a hands-on session with peer support to answer questions when they were stuck, as they had experienced during their LearnLink tutorial.

Theme Three: Ensuring the Students had Opportunities to be Self-directed and Active Learners Needed a Delicate Balancing of Instructor Authority and Student Autonomy

In one section, the researcher behaved as a peer tutor to the group. Students claimed to the instructor that having a 'mature' student available for questions was helpful to them. In my experience, students would often express ideas and opinions about the course that they would be reluctant to share with the professor for fear of appearing to criticize or challenge the instructor's role. However, in my informal meetings with the professor, I would often pass on these concerns and ideas which gave the professor an opportunity to unobtrusively deal with these issues either during class time or on LearnLink. Similarly, the professor would have issues with the students (e.g., the students were not submitting the LearnLink self-evaluations, or not booking meeting times with the instructors) and, as a peer tutor, I could find out what impediments were stopping the students from following through and then deal with those issues. It seemed that the role of the peer tutor, although at times delicate, could be used to reduce some of the students' fears of the "authoritarian" professor who inhibits some of the their self-directed behaviours and attitudes while also increasing some of the students' assertive behaviours.

In terms of the small groups, the researcher's experiences are best described by my final self-evaluation submitted at the end of the term.

> "We've produced an instrument that, with some fine tuning of the variables and scales, would produce meaningful data. I would like to point out that Sharon and George initially suggested the (general population) survey (on the Drive Clean Program) with Debbie enthusiastically agreeing. I was the reluctant on thinking that, perhaps, maybe we could do a poster session (about the Drive Clean Program). My experience of building the instrument has been interesting. My role has been to provide them with the "real world" research behaviours, expectations, resources, etc., and then get out of their way so they can decide what they want to do. So I think we've been successful in generating a viable product which represents a group effort. [...] I feel I've been successful in maintaining a balance between encouraging the group to extend their expectations of what they are capable of, helping them stay focused when slightly off track, and staying out of their "process" of being self-directed about their knowledge.

As this quotation reflects, the required skills in being able to balance the students' genuine need to know something that would be out of their realm to understand, with the hardy belief that the students are better off discovering new knowledge and skills for themselves.

Theme Four: Students Re-discovered How to be Academically Successful in this Course

The initial impressions of the professors when reviewing the focus group transcripts were that the students "seemed young" and were struggling with developing skills that they felt the students would have developed prior to entering university (e.g., time management, basic library skills, and computer usage). The students also recognized that they needed basic skills to be successful and felt that Inquiry offered them the opportunity to develop them. The following quotation from one of the focus groups reflects this opinion.

> Focus Group Student: "Well, I picked it basically because I thought that it would be of value, like throughout whatever I planned to do. I would always have the background knowledge…the knowledge that we get in this course would be able to stay in the background so we could work on it.
>
> Interviewer: Sort of like a foundation…
>
> Focus Group Student: Yeah, yeah [….] like its not a specific field that I will never have any use for, rather the opposite
>
> Interviewer: So you thought this would be a course that would be really useful in terms of developing maybe more skills instead of just knowledge?
>
> Focus Group Student: Yeah, yeah"

Students were also struggling with how to organize their time and develop self-motivation within the university context. Emerging themes reflect the tension between wanting the predictability of the secondary school structure but also appreciating the "freedom" of pursuing their own learning objectives available in the Inquiry course.

> "But a lot of our courses don't offer that freedom. It's like due that day, (and) if it's not in that day then a lot of professors don't let you do it the next day. Like our stats course, if it's not in that day, than you've lost it."

> "Like they said, going to be like not lecturing, no teacher actually going to tell you what to do, it's like... your own based learning, you going to do what you want to do so, I've never actually had that experience at all."

Many students' recommendations for future Science Inquiry courses revolved around having more lectures, due dates and more extensive assignments that would make the course more "organized". This type of tension is typical of struggles that health science students describe in learning to adjust to problem-based learning where they are often frustrated that professors are not giving specific directions on how to complete their work and that there is no one "right" answer (Solomon and Finch, 1998).

Despite their recommendations for more traditional course expectations in future Inquiry courses, students felt overwhelmed by the Natural Sciences program's first term work load (i.e., Biology, Chemistry, Calculus, Physics and Psychology), and the effects of large class sizes and lecture formats.

> "I think I rediscovered working with other people cause for like the entire first semester, basically all you did was work with your textbook and the notes and that's it. So I sort of I found that not writing about stuff and not talking to other people and working with them over just a semester can really make a difference (laughs). Like, I completely forgot."

> "I'm seriously feeling just like a 'mark'. Like if (students) don't get the highest mark, they're like sick, totally. There's no learning experience, you just study to get 'the mark', that's it. It's your only incentive sometimes. But that's not the right thing obviously, to get marks..."

> "...Cause they expect you to be so passionate about a subject, so that at the end of four years you decide, ok, I want to go into medicine, or teach children. But if you never have an opportunity to interact with anyone like that, I mean never have that experience (to learn), then all you think about is marks, marks, marks, you don't even get a chance to think, do you love this , or not? Its just memorizing."

The effects of the pressures of a traditional learning environment, combined with many students wanting to get high marks to apply to professional schools inhibited many students from developing friendships with other students who share similar academic interests, increased feelings of competition and loneliness, and decreased the relevance of a university learning experience by focusing on the 'static' materials of text books and notes.

> "...and in the first semester, didn't you find that everyone just wanted to compete with each other so much like...if you wanted help, people wondered if you wanted the answers or you wanted <u>help</u>. And you're thinking, I just want help, chill out (laughs). But no one wants to chill out, so I just totally forgot about asking people. If I didn't know someone, I wasn't ready to like ask for help. So

this course taught me like that again, like if you need help, its' not a problem, like no one's going to kill you, hopefully."

"Probably that's a reservation that's with us now as a result of first semester. Where you didn't really want to like, know people that you didn't know and everyone minded their own business. You're a little bit less up to meeting people. Whereas in high school, everyone always knew everyone, that type of thing, no one was afraid to talk to anyone else."

The experience of small group cooperative learning, using LearnLink as a means to communicate with their class mates outside of class time and to have discussions with a professor, seemed to make university life more accessible to the students, although it did not completely remove their fears of engagement with others. Also, having an opportunity to focus on learning about how research is conducted, as opposed to rote memorization from the text book, seemed to encourage the students in developing support relationships that fostered academic curiosity and growth.

Because of their experiences in Inquiry, students moved from being less competitive for marks between students and developed an internalized notion of themselves as university students.

"Yeah, I don't see the end result of this course as a mark being on my transcript"

Focus Group Student One: "About the grading, I was not so worried as I was confused. I mean I'm not so worried about the kind of mark I will have but I think that, that it's good in a way, because it more reflects the way things will be when we get out of here. Because once we're out there and we're doing some research and whatever, there won't be anyone telling us what to do for this many percent, so…

Focus Group Student Two: yeah

Focus Group Student One (con't): so in a way, it teaches a lot of …

Focus Group Student Two: discipline

Focus Group Student One (con't): yeah, discipline, and also, for someone to know what they are supposed to do to get the result they want…

Interviewer: so you have to be self-monitoring in terms of how successful you're being

Focus Group Student One: but I think it's all your own responsibility"

Although students expressed some anxiety about how the final Inquiry mark was going to be assigned at the start of the term, students did not mention marks at all during class time by the end of the semester: it appeared that the students were more focused on completing their small group project for their meeting with the instructor than they were on "getting a mark". Defining their sense of self-worth and self-discipline during the first year of university studies is developmentally appropriate for persons in their late teens and early twenties. These students have often moved away from home in the last six months and are establishing identities outside their families of origin. In developing their identities as university students, Inquiry offers them an opportunity to learn about what is appropriate research in a less competitive environment with professors acting as real world role models.

Theme Five: Students Feel That They Learned How to Learn in Inquiry

Students defined their knowledge gains from Inquiry as developing university level skills in focusing research questions, finding and critically evaluating research, increasing their motivation and discipline, as well as being persuasive in presenting knowledge.

> "Well, the good thing about it was you could learn how to work individually as well as in a group. I mean that's what I find. You're basically like pacing yourself, telling yourself what you actually have to do and stuff and have the will power to do it (laughs) dedication or whatever".

> "But still, I thought global warming was a good topic although not a very interesting one. It forces you to make your topic, to make your topic interesting to you. And that's important because often research (isn't) fun and interesting, so you have to make it work for you."

> "Cause global warming is such a broad topic right? So (dividing the topic up between members of the small groups) allowed you to basically pinpoint and narrow down . . . and learn a lot more."

> "Cause that's a real test of knowledge, right? If you can…it's easy to research but if you can explain, sit down and talk to someone about all the knowledge that you've obtained then that says something, you can really reinforce what you've learned."

Discussion and Conclusion

The data analyzed in this report illustrated that the students were struggling to develop an identity within that university environment. This struggle was in keeping with the developmental milestones of young adults. They were anxious for role models who could show them how to be successful in using research skills and in developing cooperative relationships with their peers. Large class sizes, with few opportunities for discussions with professors or peers, appeared to alienate students and focused their attentions on getting high marks instead of learning. Courses like Science Inquiry may provide a model to help students bridge the gap from moving from high school to university successfully.

Research Outcomes that Affected Program Implementation: From the research data collected in the first year and the instructors' observations, there were three major findings. Firstly, it was apparent that Inquiry sensitized students to the diverse learning experiences available in a university compared to their secondary school environments. However, being aware of the breadth and depth of university level research was not sufficient for students to feel confident in approaching these tasks. They were more likely to approach new learning experiences and persist in skills mastery when bolstered by an academic social support system. Steps were taken to ensure that the 1999-2000 Inquiry course provided this social support system through increased instructor-student contact (e.g., individual student-teacher interviews), the introduction of peer tutor-student interactions, and continued student-student small group interactions. Secondly, the decision was made to extend the course to 36 hours of class time over two semesters (i.e., one hour per week first term, two hours per week second term) instead of the traditional one semester course. The instructors found that students would be more open to and would derive greater benefit from learning Inquiry skills in September when they were just beginning at university: students starting Inquiry courses in January were resistant to learning about the library, for example, since they had already "been to the library" although almost all students did not know that journals existed and could not name more than one electronic abstract database. Third, the instructors also realized that students should participate in more than one group problem and have at least one opportunity to choose a problem based on their interests. With at least two attempts, both the instructor and students could learn from any difficulties in the first problem and then make a second attempt with their newly learned insights.

In the second year of the course, the instructors worked as a teaching team by developing and using the same course elements and exercises in all sections. They agreed that the instructors, acting as facilitators, would benefit from ongoing discussion and collaboration with other first year Science Inquiry instructors to improve the quality of the course. The instructors from various disciplines [i.e., mathematics, biology, psychology, and health sciences] and a representative from the Centre for Leadership and Learning (CLL) met weekly to share their classroom experiences of what worked, problem-solve any difficulties, discuss how their ideas about teaching and learning were changing, and share any tools they found effective. The advantage of this collaboration was that the instructors were pilot testing and refining a series of teaching tools and sessions they found effective in helping students work through their skill set (e.g., how to get students into the library, mini-inquiry benchmarking exercise, group work skills, how to encourage peer tutor – student interactions, critical thinking skills, how to assess the validity and reliability of a website, the LearnLink tutorial, how to make presentations, etc.)

However, to encourage more faculty to teach Inquiry, they decided to have a three-year rotation of instructors. This rotation ensured that experienced instructors supported new inquiry instructors with third year instructors taking a leadership role. The team realized that having a tightly structured course, based on their developed tools and sessions, would impede new instructors' full participation in the teaching team. New instructors brought their own expertise and previous teaching and learning strategies which added additional depth to the collaboration. The decision to have a three-year instructor rotation meant that the Inquiry teaching model needed flexible boundaries. This flexibility meant that experienced instructors give new instructors room to be self-directed in their teaching while new instructors were amenable to the support available from the other instructors, peer tutors, CLL and research initiatives. Science Inquiry team meetings evolved into developing and sharing a 'tool kit' of flexible tactics that instructors could use to help the students explore the skill set in response to the students' needs. The teaching teams in the third offering of Inquiry became this type of partnership between the instructors, a Centre for Leadership and Learning representative, research support and their respective peer tutors. These teaching team meetings have continued in subsequent inquiry courses to the mutual benefit of all.

References

Centre for Leadership in Learning. Proposals for the McMaster Inquiry Project. Working document last revised May 1999.

Glaser B., & Strauss A. The discovery of grounded theory. 1967: Chicago; Aldine Publishing.

Loftland J. (1971). Analyzing social settings. Belmont, CA; Wadsworth Publishing.

Morgan, D.L. (1997). Focus Groups as Qualitative Research (2^{nd} ed.). Thousand Oaks, CA: Sage.

Morse, J.M., & Richards, L. (2002). Readme first for a user's guide to qualitative methods. Thousand Oaks, CA: Sage.

Schatzman L., & Strauss A. (1971). Field research. New Jersey, NY; Prentice-Hall.

Solomon P. & Finch E. A quantitative study identifying stressors associated with adapting to problem-based learning. Teaching and Learning in Medicine, 1998, 10(2), pp. 59-64.

Stauss A., & Corbin, J. (1998). Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory (2^{nd} ed). Thousand Oaks, CA: Sage.

CHAPTER 4

Development and Validation of Domains of the McMaster Science Inquiry Project: Academic Year 1999 to 2000

The qualitative study from the previous year provided the following insights into how students experienced the previous years' Inquiry course. The formative evaluations indicated that students felt that they gained research, critical thinking, self-assessment and interpersonal skills from taking this course. They appreciated that the course was 'hands-on' and felt that they would continue to benefit from this course in their academic and employment careers. Inquiry sensitized the students to the diverse learning experiences available in university when compared to their secondary school environments; however, being aware of the breadth and depth of university level research is not sufficient for students to feel confident in approaching these tasks. They were more likely to approach new learning experiences and persist in skills mastery when bolstered by an academic social support system. Also, students at the end of the course still struggled against having 'a right answer'. While they are eager to explore their own interests in a topic area, they still viewed their efforts as "inadequate" but were beginning to see that "knowledge" was constantly changing and influenced by its context.

However, to assess whether the course is responsible for these self-reported benefits, and not other factors (i.e., meeting normal developmental milestones of higher education), required more rigorous evaluation methods. It was anticipated that science inquiry courses would continue as a McMaster initiative for first year students and would expand with Inquiry courses offered to upper year students starting with year four students in 2000-2001 as well as in the inaugural year of the Bachelor of Health Science program. Therefore, we initiated a three year study of the following students: all first year Science and Health Sciences Inquiry students; all students taking a fourth year, large enrollment, Molecular Biology course taught with a combination of inquiry and lecture learning; and students who were not enrolled in Inquiry but were enrolled in a first year science program. The purpose of collecting this data was to assess the gains in inquiry skills (i.e., critical thinking, academic self-efficacy and learning preferences) between the variety of inquiry courses compared to a control group who had not taken a first year Inquiry course. We also resurveyed the students who had taken a first year Inquiry course in subsequent years to ascertain the degree to which these skills persisted over time.

There were no specific, standardized tools that measured "inquiry skill development" at the inception of this study (1999-2000) nor did I expect to find any such miracle instruments. However, translating themes generated by the qualitative study into a specific methodology to measure these outcomes needed to be balanced against the instructors' right to determine the best way of teaching the course to the students. Given the fact that the instructors were committed to using a previously developed teaching tool as a research tool, I felt that it would be appropriate to standardize this teaching tool (i.e., the mini-inquiry exercise) to study the domains identified in the qualitative study such as critical thinking, library search strategies, and self-assessment skills. Other domains such as improving academic self-efficacy and preference to understand the content areas as opposed to memorizing them were developed separately and discussed in later sections of this chapter.

Development of the Mini-Inquiry Benchmarking Exercise

In the second year (1999-2000), the instructors initiated a mini-inquiry benchmarking exercise to give the students a 'snapshot' of their inquiry skills development at different times in the year. The students were given a newspaper clipping, asked to complete individual and group exercises, and then evaluate their performance. Both the students and instructors found the mini-inquiry exercise highly informative and useful as a means for starting discussion about: student evaluation; giving the students the opportunity to gauge their own performance in the course; and, the elements of the Inquiry skill set. The instructors also hoped that the mini-inquiry benchmarking exercises could be used as a research tool.

In the summer of 2000, I attempted to code the mini-inquiry tool used by the instructors in the previous year to see if we could find trends in the students' learning. Analysing the mini-inquiry data was too difficult because of rater bias issues (i.e., different articles were used by each instructor in the September administration but the same clipping was administered to all students at the end of the course so the rater could identify the post-test). However, the instructors were willing to change the administration to generate reliable data. Also, the Centre for Leadership in Learning offered financial support for this program evaluation. Because of this level of support, I developed the following tools which were used in the three years of the summative program evaluation.

Findings from the attempted rating of a sub-sample of the 1999-2000 academic year's mini-inquiry benchmarking exercise was informative for developing the exercise as both a research and teaching tool. With two additional raters (Dr. Erika Kustra, an Educational

Consultant from the Centre for Leadership in Learning and Jerrome Marrin, a second year university student), we attempted to "make sense" inductively of how the students performed on the mini-inquiry exercise and to pre-test some possible coding sheets. Some of the insights were as follows:

1. There were marked differences in student performance based on the type of articles given to the student. Articles on divorce predictors given in post-test administration generated more superficial responses than in the pre-test administration. By contrast, the article on water contaminants, which had many specific details, generated more specificity from the students. The article on RU486 (the abortion pill) tended to have students formulate a moral or emotional response that indicated they had already decided what the answer should be. We concluded that it made sense to choose articles that cited some specific data or details, were personally relevant to the students but not likely to generate a polarized response.
2. It was also difficult to track the students' thinking process when they were asked to start with one question. Students needed an opportunity to "brainstorm" all of their ideas after reading the article so as not to limit where they may have taken their ideas if they had an opportunity for reflection. For instance, the student may have started with a question that, upon further reflection, they wanted to abandon for a different question. The way the exercise was structured only allowed the student to reflect on this change in the self-reflection on their performance based on their first choice of question.
3. Some of the instructions to the students were contradictory to the goals of Inquiry. For

example, some of the students generated a good first question (i.e., a complex inquiry question) but when asked to revise it to an 'easy to answer' question as specified by the written instructions, reverted to a more superficial and less complex question.

4. Students were quite good at citing sources of information they would use to answer their question but were not asked how they would use them. We felt that asking them why these sources would be useful would improve the evidence of their research ability and critical thinking gains.
5. The instructors wanted both group work and individual contribution components to the mini-inquiry benchmarking exercises. The three raters pre-tested a revised version of the instrument that included the group work component. In this version, the students worked in groups after they had individually completed the first part of the mini-inquiry (i.e., read the article, summarized the article, generated a question and a rationale for the question, and devised a search strategy). We took longer than a typical class time to complete the individual component of the exercise because we found that we needed time to reflect and think through our responses. We felt that it was too ambitious to complete both the individual and group administrations: the compromise was that the students would switch their paper with a peer and have a brief conversation and feedback session prior to completing the self-evaluation session.

From this pre-test, we agreed that in order to trigger the students' thinking, the newspaper article should cite some specific details or data, have some personal relevance to students' age group, should not provoke a solely moral or emotional response and should be reasonably short with an accessible scientific vocabulary. In addition, trigger articles should have a subject common to all students in each cohort (i.e., Science, Health Sciences, and fourth year Biology inquiry students). Also, we determined that the mini-inquiry exercise as a whole should be a close approximation to the desired academic behaviours of the students by the end of the course *on an individual level* and that the students should see a different trigger article in the September and April administration.

The standardized version of the mini-inquiry benchmarking exercise was agreed upon by the instructors and raters as a tool that would work as a teaching tool as well as an assessment tool that would yield reliable and valid data on how the students were experiencing Inquiry. The finalized tool assessed the students' inquiry skills and attributes using the following manoeuvre: each student was asked to complete a mini-inquiry benchmarking exercise during class time in the first and last month of the academic year (i.e., September and March). The students were given a stimulus to read (i.e., one of six to eight standardized newspaper clipping at each administration) and asked to:

A) brainstorm all the ideas, questions, and previous knowledge;

B) select one question that they wanted to research further;

C) describe, using as much detail as possible, how they would research their selected question;

D) swap their paper with a class peer for feedback; and

E) self-evaluate their performance on the mini-inquiry exercise.

Mini-inquiry stimului were rotated between class sections to minimize the likelihood that the students saw the same stimulus twice. Prior to its administration, students were informed of the purpose of the mini-inquiry exercise and that it would not be marked like a test paper or exam. Students had 45 minutes to complete the mini-inquiry exercise and they were given verbal prompts to help keep them on track to complete the exercise. (See Appendix 4.1 for a sample mini-inquiry benchmarking exercise.)

Newspaper clippings were standardized by searching the literature for approximately 40 articles of a scientific nature that had similar difficult word to word count ratio (20% to 48%), readability score (11 to 15 years education) using the Fleishman-Kincaid score, and similar content to all cohorts under study (e.g., biology, science inquiry, and health sciences). These articles were sent to seven to ten reviewers who were Inquiry instructors, previous mini-inquiry exercise raters, or undergraduate students. Article reviewers were asked to create two categories of articles: those articles they felt were very good to excellent for the mini-inquiry exercise and those they would not want to use the mini-inquiry exercise. The six to eight articles that had the most positive endorsements by from the article reviewers were used in the following year's mini-inquiry exercise. In the second and subsequent years, there was a worry that there was a ceiling effect of using newspaper clippings since the articles may not have had enough specific details for the fourth year students. The above described standardization procedure was used in subsequent mini-inquiry benchmarking administrations with medical abstracts (i.e., Medline) instead of newspaper clippings.

Developing the Coding Procedure to Rate Critical Thinking: The Evaluation of Information, Self-Assessment, and Library Skills Development

A 1990 Consensus Panel, lead by Peter Facione (Facione and Facione, 1996) developed a hierarchical definition of critical thinking as the "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference". The Consensus Panel also emphasized the need to foster dispositions essential to good critical thinking skills and attributes such self-assessment, library skills, information evaluations, strong academic self-efficacy beliefs and a preference for a meaning orientation to learning. In an article attempting to translate the actual definition of critical thinking into teaching and assessment strategies, Facione and Facione (1996) attempted to develop a guide for translating the conceptual definitions into "authentic" measures of critical thinking (p.129). These authentic measures are described below in Table 4.1.

Table 4.1: Teaching Domains that were Abstracted from the Mini-Inquiry Activities
Critical Thinking (Facione and Facione, 1991; Trim, 2000)
Interpretation -- *Conveying or Searching for meaning in the words, phrases and concepts* -- (e.g., Categorization, decoding and clarifying meaning)
Analysis -- *Conveying or Searching for comprehension of the overall meaning* -- (e.g., examining ideas, identifying and analysing arguments)
Evaluation – *Wondering/Questioning whether the text makes sense, the findings are valid, reliable* – (e.g. assessing claims and arguments)
Inference – *Drawing conclusions from the student's understanding of the text, making connections to other knowledge, etc.* – (e.g., querying evidence, conjecturing alternatives, drawing conclusions)

Table 4.1: Teaching Domains that were Abstracted from the Mini-Inquiry Activities
Search Strategy How specific is/are they? How likely would it/they yield good data? How practical is/are they?
Self-Assessment Willingness to Self-Assess? (Authenticity) Open-minded to personal strengths and weaknesses? Analytical? Self-Confidence? Inquisitiveness to multiple points of view?

For our study, an expert panel of Inquiry researchers, instructors, and educational developers reviewed these domains for face and content validity. The domains were then refined through extensive training procedures with the raters. Three raters were recruited from upper year inquiry students who had familiarity with the student cohorts and were aware of the pedagogic goals of inquiry. As part of the training procedure, the raters were asked to complete a mini-inquiry exercise identical to one used in the classroom and to read the article by Facione and Facione (1996). Subsequently, raters were asked to individually rate ten mini-inquiry exercises taken from that year's administration. The training mini-inquiry benchmarking exercises were taken from all class sections and were selected to represent a potential range of responses. Once the raters had coded the initial ten exercises, they were brought together to resolve discrepancies in their ratings and to receive guidance from the principal investigator (e.g., to minimize end aversion, discussion of theoretical constructs, moderate disputes between raters, etc.). Once the raters had resolved their first set of rating sheets, the raters were given a second set of 10 mini-inquiry exercises to rate individually and then resolve with the other raters. At this point, the raters were given their initial rating package. Subsequently, team meetings were held weekly with the research team to discuss any problems and to review missing data from the previously submitted rating sheets. As much as possible, raters were asked to return from the previous year to ensure a consistency of rating across the repeated administrations of the study. However, each year the training procedure was repeated to ensure that the new raters had the comparable level of understanding as the returning raters.

Rating Procedure: Completed mini-inquiry benchmarking forms for both September and March in all three groups: had all identifying information removed, were assigned an identification number that blinded the rater to which group and time point the mini-inquiry exercise was completed, were randomly allocated to groups of 100, and were coded by three raters on five point scales (1 = poor to 5 = excellent) for each item using the coding sheet developed. All three raters completed coding sheets on all mini-inquiry benchmarking exercises. (Please see Appendix 4.1 for an example of the coding sheet.)

Sections A and B [A) brainstorm all the ideas, questions, and previous knowledge; and B) select one question that they wanted to research further] were coded for evidence of critical thinking. Section C [C) describe, using as much detail as possible, how they would research their selected question] was coded for evidence of library search strategies. Sections D and E [D) swap their paper with a class peer for feedback; and E) self-evaluate their performance on the mini-inquiry exercise] were coded for evidence of self-evaluation skills. Raters were also asked to give their overall assessment on a five point likert scale for each mini-inquiry benchmarking.

Results from the Mini-Inquiry Benchmarking Reliability and Validity Study

Study Participants: Data from the first year of the study (2000-2001) was used to determine the reliability and validity of the mini-inquiry rating procedure. Data from samples of students from three different inquiry interventions were used to test whether this instrument was sensitive to measuring student differences in Inquiry skills from the beginning to the end of the course. Students in the **science** inquiry group enrolled in a 36 hour inquiry course over two semesters: science inquiry students took this course as an elective and had to be in first year Faculty of Science program. Students in the **health science** inquiry group were enrolled in a mandatory first year inquiry course with 72 hours of instruction. Students in the **fourth year science** inquiry group were enrolled in an elective fourth year molecular biology course with 36 hours of instruction. In terms of class organization, all courses involved small-group inquiry learning, computer discussion groups, and in-class discussion. However, the first year courses were offered in small sections of 20 to 30 students per instructor while the fourth year course had one instructor for the entire class. Also, the first year course had activities specifically designed to teach students how to evaluate information, use the library, work with peers, and self-assess their strengths and weaknesses. The fourth year course assumed that the students had learned these skills previously.

Data Analyses: This rating data was analyzed using SPSS 11 and GENOVA statistical software systems. Two techniques were used to establish the reliability of the sub-scales of the mini-inquiry exercise; internal consistency (Cronbach's Alpha) and inter-rater reliability using generalizability modelling [Student x (Rater : Time)] that accounted for the nesting of rater and time point (since all raters coded for both time points) (Crick and Brennan, 1983; Crossley et al., 2002; Streiner and Norman, 2004). Principal components factor analyses were conducted on each separate raters' scores to establish the validity of the sub-scales.

To validate this scale, ANCOVA analyses were conducted on each sub-scale to estimate the adjusted mean group differences on pre-test to post-test scores for each Inquiry course. The balanced model took into account the following factors as main effects: time (beginning or end of the course); course (first year Science, first year Health Science or fourth year Molecular Biology); article (whether the student saw the same or different article); rater (three raters); and the individual differences. The interaction effects were also entered into the model taking into account the nesting of the individual student within each course. The benefit of this method of analysis is that the mean values of interest can be calculated which take into account (adjust for) the effects of covariates. In this report, the F values were reported to give the reader an indication of the magnitude of a factor's impact on the model.

Results of Mini-Inquiry Benchmarking Study

Results: In Table 4.2, first year Science students (n = 81) had a mean age of 19.6 years, 24 % were males, and 29% spoke a language other than English at home. First year Health

Science students who had complete data (n = 97) had a mean age of 19.1 years, 18 % were males, and 9% spoke a language other than English at home. For the fourth year Science students who had complete data (n = 98) had a mean age of 22.1 years, 41 % were males, and 22% spoke a language other than English at home.

Table 4.2: Mini-Inquiry Benchmarking Study Participants by Comparison Group Year, Age, Gender, and Language Most Spoken at Home

	Age (Mean, SD)	Gender (% Males to % Females)	Language spoken in Parental Home Not English (%)
Science Inquiry n = 81	19.6 (2.1)	24% to 76 %	29%
Health Sciences n = 97	19.1 (0.42)	18% to 82 %	9%
4^{th} Year Sciences n = 98	22.1 (1.4)	41% to 59%	22%

Internal consistency (Cronbach's Alpha) for the three sub-scales ranged between 0.68 to 0.98 using data from each rater independently. Inter-rater reliability co-efficient was 0.57 for the appraisal of arguments score, 0.63 for the search strategy score, and 0.59 for self-assessment skills score. The non-rotated principal component analyses were conducted on each rater using the total of all 13 items and the total variance explained for each rater ranged from 64% to 77%. Principal component analyses on each rater on the 12 items using varimax rotation determined that three theoretically determined sub-scales (e.g., appraisal of arguments, search strategy, self assessment) had consistent factor loadings for all three raters. Based on these data, the sub-scales were deemed to have an acceptable reliability for this study.

Table 4.3: Mini-Inquiry Exercise Reliability, Internal Consistency and Percentage of Variance Explained by Rater

	Evaluation of Information (4 items)	Search Strategy (3 items)	Self-Assessment (5 items)
Inter-rater Reliability (n = 267)	0.57	0.61	0.59
Internal Consistency (n = 272)	0.68 – 0.89	0.74 – 0.98	0.80 – 0.98

Table 4.4: Mini-Inquiry Study Participants by Critical Thinking Sub-Score Comparisons by Program Type

	Adjusted Means	Appraisal of Arguments	Information Seeking	Self-Assessment	Overall Critical Thinking Assessment
Science	Pre	58.3	58.0	75.1	51.2
	Post	71.8	75.0	79.3	66.2
	Mean Difference	13.5	17.0	4.2	15
Health Science	Pre	69.5	67.1	74.6	67.8
	Post	80.1	85.9	81.2	80.1
	Mean Difference	10.6	18.8	6.6	12.3
Fourth Year	Pre	73.6	71.9	54.7	70.9
	Post	77.1	84.8	69.4	76.6
	Mean Difference	3.5	12.9	14.7	5.7

* All sub-scores totaling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores.

For all sub-scores, the main effect of time and course had large F values that were statistically significant: this result adds to the accumulating evidence that the mini-inquiry exercise does measure student change from the beginning to the end of the course and was able to determine differences between the type of course. The main effect of article and article x time was not significant which indicates that articles used in the mini-benchmarking exercise were sufficiently standardized and did not impact on student performance.

Comparisons of the adjusted mean values show that fourth year students did better than both groups of first year students. Similarly, health science students, who have more hours of instruction, tended to do better than first year science inquiry students. Not surprisingly, self-assessment scores were lowest for the fourth year students. Typically, self-assessment and metacognitive skills are only expressed when students have been explicitly told that these insights are welcomed in the learning environment. Self-assessment skills are usually deemed to be socially inappropriate in the classroom and are seen as making one vulnerable to ones' peers. In the fourth year class, where there was no classroom discussion of the value of these skills, the students would be less likely to write them in the mini-inquiry exercise whereas the students in the Science and Health Science programs had classroom discussions about their use in classroom learning.

The Development of the Science Inquiry Academic Self-Confidence and Approaches to Learning Scales

Both the Science Inquiry Academic Self-Confidence Scale and the Approaches to Learning Scale were validated in the same study: consequently, their results will be reported together.

Science Inquiry Academic Self-Confidence Scale (SIASCS)

Bandura's work on self-efficacy has been studied and reported extensively since its advent in the 1970's. He posited that self-efficacy beliefs mediate the cognitive, metacognitive, and motivational mechanisms for self-regulation and behavioural change (Bandura et al., 1996). Self-efficacy is a belief that one can successfully accomplish tasks, and through these beliefs, be influenced in one's choice of behavioural settings, activities, peer relationships and willingness to persist with skills mastery. Since the 1980's, several researchers have developed scales to measure primary through tertiary school students' self-efficacy beliefs in the areas of general academic achievement (Lent et al., 1984; Bandura et al. 1996; Rouxel, 1999), mathematics (Betz and Hackett; 1981; Betz and Hackett, 1989; Nauta et al., 1998), vocational skills (Betz et al, 1981;1986;1996; Lent et al, 1987; Parson and Betz, 1998; Betz and Schifano, 2000), and writing skills (Zimmerman et al.,1992). The findings of these authors strongly suggest that global rating scales are less likely to produce results sensitive to students actual change in beliefs because self-efficacy beliefs appear to be context specific (Lent et al.,1987 ; Betz et al.1996; 2000; Nauta et al., 1998). A review of the literature found that there were no self-efficacy scales developed for the science inquiry or active learning strategies.

These scales were developed using a methodology described by Betz et al. (2000) based on data from the 1999-2000 Science Inquiry Project data. The first scale, the Inquiry Task Self-Confidence scale measures 25 items that ask the students to rate their confidence to complete tasks such as evaluating websites, providing group members with feedback and finding information in campus libraries. The second scale, the Projected Academic Goal Completion Self-Confidence Scores confidence scale, measures 10 items which asks the students how confident on tasks such as asking a professor to write a reference for them or call the school of graduates studies for information. Previous measures of self-efficacy have asked respondents to answer yes or no as to whether they feel they can complete a task, and if yes, rate the strength of that belief on a ten-point scale. However, subsequent revisions by Nancy Betz and colleagues stated that researchers have tended to use either the dichotomous responses or the total strength scores, but not both (Parson and Betz, 1998). Betz (1996) now suggests using a five-point scale ranging from one (no confidence at all) to five (complete confidence) which is the method used in these inventories. (See Appendix 4.1 for the list of items in the SIASCS.)

Approaches to Studying Preferences Scale

There are several standardized instruments which measure the construct of student learning styles such as the Approaches to Studying Inventory (The Lancaster) and the Learning Style Inventory (The Kolb). Ramsden (1983) developed the Lancaster scale to examine how higher education students experienced their courses and their study methods. Ramsden noted that there is an important connection on how a student feels about a course, to how much they apply themselves to the materials to how effectively they learn (Ramsden, 1983). Ramsden conceded that learning approaches were not so much transferable skills as they were preferences to think deeply about the subject matter and to retain the material in memory. In his experience as an educator and educational developer, he felt that perceived good teaching and the existence of intrinsic interest in the subject were related to deep and organized approaches to learning. This scale was used for this study because "perhaps the success of 'problem-based learning' and its many variants was the substitution of the traditional passive learner role for a broader range of learning mechanisms initiated by the learners themselves" (Curry, 2002).

In the Lecturer's Handbook for the Approaches to Studying, the scale was described as being reviewed by an expert panel for content validity and having a student sample complete both the Approaches to Studying Inventory and a similar scale developed by John Biggs to establish construct validity. For the purposes of this study, we examined changes in student preferences for three sub-scales of the Lancaster Inventory (Ramsden, 1983): a deep or meaning learning orientation where the student was purported to retain a conceptual understanding of the subject material; surface or memorization orientation where the student was purported to "forget" the information as soon as it is learned; and an organized preference where student was purported to have little difficulty with procrastination and have good study habits. Inquiry learning, with the emphasis on self-directed and group learning, should give students an opportunity to increase their preference for meaning and organization orientations and decrease their preference for a reproducing orientation. (See Appendix 4.1 for the list of items in the Approaches to Studying Scales.)

Study Participants: To establish the validity of the scales, two studies were run at McMaster University in 2000. Subjects for the first study were recruited for validity and reliability data analysis from a first year calculus course (n = 647/1371 students or 47% of the enrolled students) and from a fourth year, biology course (n = 108/130 students or 83% of enrolled students) for a total number of 755 subjects. These subjects were administered an in-class paper survey including the SIASCS and the Approaches to Learning Inventory as part of a formative evaluation of the first year Calculus course at the end of the winter semester in 2000. Tables 4.5 and 4.6 show how the inclusion of these subjects allowed for a broad spectrum of university programs oriented toward math and science curriculum at all levels of undergraduate years. Also, there was a deliberate attempt to sample limited enrollment programs which, at this university, have a strong emphasis on active learning strategies through small group, inquiry-based learning pedagogies. However, for the purposes of comparing the results of the sub-scores, the data has been dichotomized into the following: upper year (year of entry equal to 1998 or earlier) and first and second year students (year of entry 1999 or later); inquiry-based courses (health science, arts and science and honours molecular biology) and non-inquiry students (years 1-3 science, kinesiology and nursing, humanities and social science, business and computer science).

Table 4.5:

Participants by Year of Entry and Gender

		Female	Male	Total
Year entry to university	≤1996	6	5	11
	1997	55	37	92
	1998	6	7	13
	1999	27	12	39
	2000	403	188	591
Total		497	249	746*

*nine cases with missing data for gender

Table 4.6: Participants by Year of Entry to University and Program

		Years 1-3 Science	Kinesiology and Nursing	Humanities and Social Science	Business and Computer Science	Health Science	Arts and Science	Honours Year Science	Total
Year entry to university	≤1996	2			1			8	11
	1997							93	93
	1998	3	3	2				5	13
	1999	16	7	5	5		4	2	39
	2000	355	48	48	43	51	48		593
Total		376	58	55	49	51	52	108	749*

*six cases with missing data for program

Subjects for the second study, the test-retest reliability generalizability study, were recruited from a first year science inquiry course. Twenty-nine subjects were asked to complete the SIASCS and Approaches to Studying on two separate occasions (i.e., prior to Christmas break and in the last week of January). Subjects for this study were homogenous in that they all were first year science students (i.e., 2000 year admission) and 79% (23/29) were female.

Data Analyses: All data were analyzed using SPSS 11. Two techniques were used to establish the reliability of the inventories and their factors; generalizibility analysis based on the test-retest data (Crossley et al., 2002; Streiner and Norman, 2004), and internal consistency (Cronbach's Alpha) based on the calculus and molecular biology students. Two techniques were used to establish the validity of the scales: a principal components analysis; and independent samples t-tests using variables hypothesized to show differences in academic self-efficacy beliefs and approaches to studying beliefs between the groups (i.e., program and year of admission). Criterion validity was also determined by correlating the all sub-scores from both the Approaches to Learning and SIASCS with student course grade (available for Calculus students only), type of program, and year of study.

Results of the Reliability and Validity of the SIASCS and Approaches to Studying Preference Scale

SIASCS Results: The first study (n = 755) confirmed that the inter-item reliability coefficients (Cronbach's Alpha) were 0.93 for the Inquiry Task Self-Confidence scale and 0.89 for the Projected Academic Goal Completion Self-Confidence Scores scale. The data from the second study confirmed that the generalizability co-efficient on the test-retest study (n = 27) for this sub-sample of students was 0.72 for the Inquiry Task Self-Confidence scale and the Projected Academic Goal Completion Self-Confidence scale was 0.77 which were acceptable for this study (Crossley et al., 2002).

To validate the scales, principal component analysis (non-rotated solutions) was conducted on the total scale (n = 755) and the total variance explained was 64%. Principal component analyses (varimax rotation solution) also determined that there were two separate sub-scales (e.g., inquiry task and projected academic goal completion). Independent samples t-test also showed that comparisons between sub-groups in the sample were statistically significant where upper year students had more confidence compared to first and second year students [Inquiry Task Mean Difference = 7.6 points, t value = 6.69 ($p < 0.0001$); Projected Academic Inquiry Goal Completion Mean Difference = 9.17 points, t value = 5.60 ($p < 0.0001$)] and inquiry students had more confidence than those students not taking an inquiry course [Inquiry Task Mean Difference = 7.05 points, t value = 8.19 ($p < 0.0001$); Projected Academic Inquiry Goal Completion Mean Difference = 9.84 points, t value = -7.95 ($p < 0.0001$)].

Figure 4.1: Comparison of Inquiry Task Self-Confidence and Projected Academic Goal Completion Self-Confidence Scores between First/Second Year Students and Upper Year Students

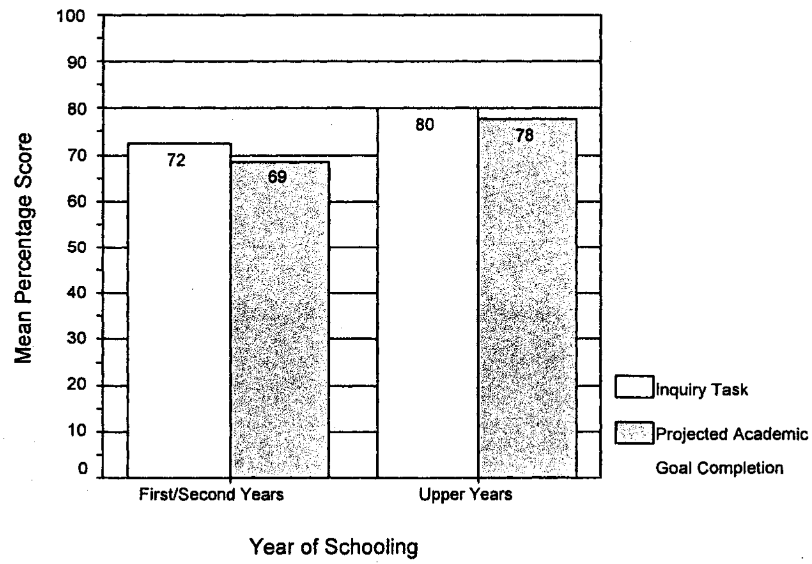
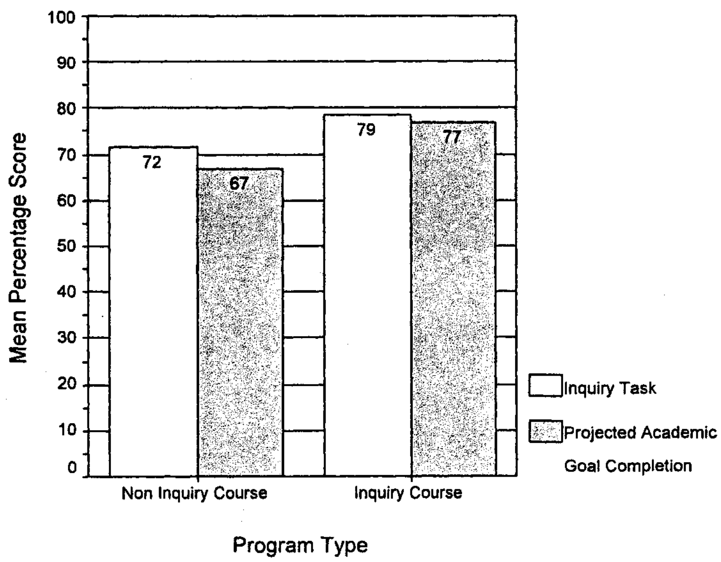


Figure 4.2: Comparison of Inquiry Task Self-Confidence and Projected Academic Goal Completion Self-Confidence Scores between Inquiry and Non-Inquiry Program Types



Approaches to Studying Preferences Results

The first study (n = 755) confirmed that the inter-item reliability coefficient (Cronbach's Alpha) was 0.93 for the Meaning Orientation scale, 0.89 for the Reproducing Orientation scale and 0.84 for the Organized Orientation Scale. From the second study (n = 27), the test-retest generaliziablity co-efficient was 0.74 for the Meaning Orientation Preference scale, 0.66 for the Reproducing Orientation scale and 0.65 for the Organized Orientation Scale which was acceptable for this study (Crossley et al., 2002).

Principal component analysis was conducted on the total scale and the total variance explained ranged was 57%. Of the thirty-four items from the original Lancaster Scale used in both studies, principal component analyses (varimax rotation) also determined that there were three separate sub-scales (i.e., seventeen items for meaning, thirteen items for reproducing and four items for organized orientation) and the items on each scale corresponded with the original Lancaster meaning, reproducing and organized orientation scales. Independent samples t-test also showed that comparisons between sub-groups in the sample were statistically significant where **upper year students** had a greater preference for a meaning [Mean Difference = 4.53 points, t value = 5.52 (p <0.0001)] and organized orientations [Mean Difference = 8.82 points, t value = 4.96 (p <0.0001)] and less preference for reproducing [Mean Difference = 6.99 points, t value = 7.55 (p <0.0001)] that those students in first or second year. Similarly, independent samples t-test also showed that comparisons between sub-groups in the sample were statistically significant where **inquiry students** had a greater preference for a meaning orientation[Mean Difference = 3.48 points, t value = 5.52 (p <0.0001)] and organized orientations [Mean

Difference = 6.77 points, t value = 4.97 (p <0.0001)] and less preference for reproducing orientation[Mean Difference = 6.98 points, t value = 10.11 (p <0.0001)] than those **students in not taking inquiry**.

Figure 4.3: Comparison of Meaning, Reproducing and Organized Orientation Scores between First/Second Year Students and Upper Year

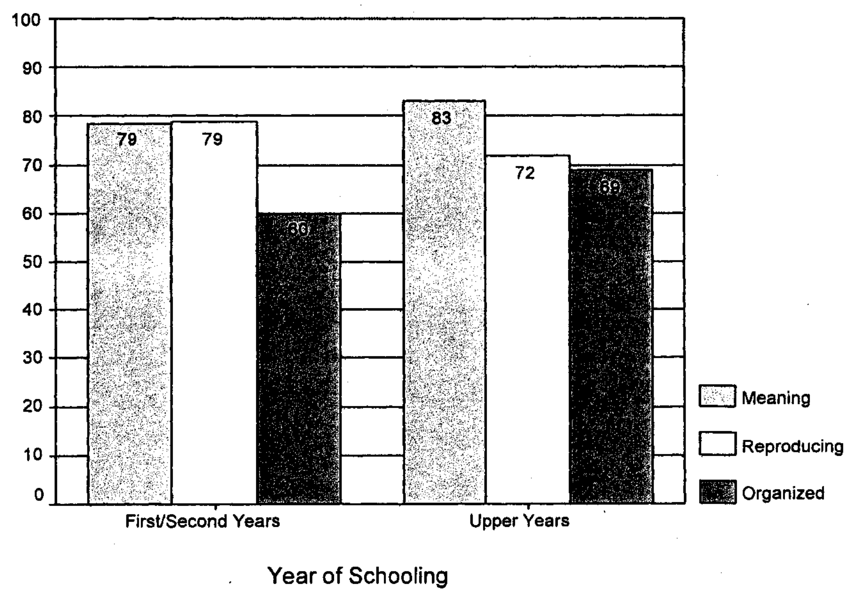
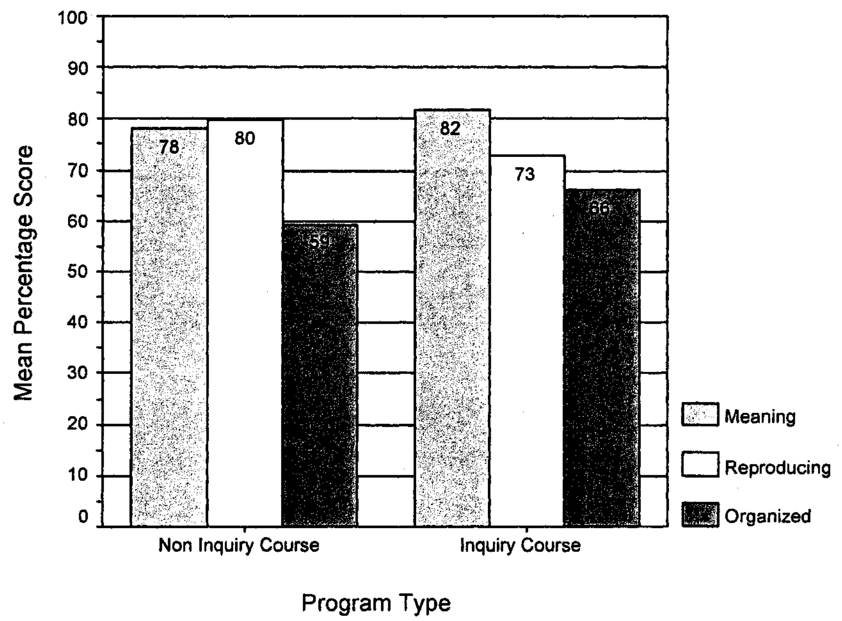


Figure 4.4: Comparison of Meaning, Reproducing, and Organized Orientation Scores between Inquiry and Non-Inquiry Program Types



Criterion Validity Study Results

The Pearson's correlation coefficients for all scores on both instruments (SIASCS and Approaches to Studying Preference Scales) were reported for the Calculus students only (n = 549) and also correlated to the final course grade in Table 4.8 below. When examining these scores, there are modest correlations between the scores indicated that these instruments are measuring related but not identical constructs. Also, expected inverse correlations for such constructs of meaning and reproducing orientations, were also seen to be negatively correlated. In addition, a one way ANOVA comparing high scoring Calculus students (i.e., A+ to A- grades) to average students (B+ to B- grades) to struggling students (C+ or lower) showed that there was a significant difference between these students on all scores. These differences show that the scales are able to distinguish between students in the expected direction and indicate that the Approaches to Studying and SIASCS are valid scales to use in higher education science and health science students.

Table 4.7: Pearsons' Correlation Coefficients on the SIASCS and Approaches to Studying Preference Scales and Final Calculus Course Grade

	Meaning Orientation	Reproducing Orientation	Organized Orientation	Inquiry Task Self-Confidence	Projected Academic Goal Completion	Final Calculus Course Grade
Meaning Orientation	1	-0.085*	0.291**	0.445**	0.429**	0.170 **
Reproducing Orientation		1	-0.223**	-0.110**	-0.169**	-0.138**
Organized Orientation			1	0.286**	0.300**	0.205**
Inquiry Task Self-Confidence				1	0.575**	0.153**
Projected Academic Goal Completion					1	0.361**

* $p < 0.01$ ** $p < 0.001$

Table 4.8: Mean Differences (SD) and ANOVA Results for the SIASCS and Approaches to Studying Preference Scales when Compared to Student Performance on Final Calculus Course Grade

	Inquiry Task Self-Confidence Mean (SD)	Projected Academic Goal Completion Mean (SD)	Meaning Orientation Mean (SD)	Reproducing Orientation Mean (SD)	Organized Orientation Mean (SD)
A+ to A- Grade (n = 201)	74.3 (10.6)	74.5 (14.6)	80.0 (8.1)	78.3 (8.1)	63.7 (16.8)
B+ to B- Grade (n = 129)	71.6 (9.6)	67.1 (12.2)	77.1 (7.2)	79.0 (9.1)	58.6 (15.2)
C+ or lower (n = 240)	69.2 (11.7)	61.4 (15.2)	76.1 (8.6)	80.3 (8.0)	56.5 (15.6)
ANOVA F value	12.1	45.6	13.8	3.115	11.7
p value	0.0001	0.0001	0.0001	0.045	0.0001

References

Bandura, A., Barbaranelli, C., Caprara, G.V., & Pastorelli C. (1996). Multifaceted impact of self-efficacy beliefs on academic functioning. Child Development; 67(3), pp. 1206-1222.

Betz, N., & Hackett, G. & Betz, N. (1989). An exploration of the mathematics self-efficacy/mathematics performance correspondence. Journal for Research in Mathematics Education; 20(3), pp. 261-273.

Betz, N., & Hackett, G. (1981). The relationship of career-related self-efficacy expectations to perceived career options in college women and men. Journal of Counselling Psychology; 28(5), pp. 399-419.

Betz, N., & Hackett, G. (1986). Applications of self-efficacy theory to understanding career choice behaviour. Journal of Social and Clinical Psychology; 4(3), pp. 279-289.

Betz, N., & Schifano, R. (2000). Evaluation of an intervention to increase realistic self-efficacy and interests in college women. Journal of Vocational Behavior; 56, pp. 35-52.

Betz, N.E., Harmon, L.W., Borgen, F.H. (1996). The Relationships of Self-Efficacy for the Holland Themes to Gender, Occupational Group Membership, and Vocational Interests. Journal of Counseling Psychology; 43(1), pp. 90-98.

Boyer Commission on Education Undergraduates in the Research University, Reinventing Undergraduate Education: A Blueprint for America's Research Universities.

Crick, J.E., & Brennan, R.L. (1983). Manual for GENOVA: A GENeralized analysis Of VAriance System. Iowa City, Iowa: The American College Testing Program.

Crossley J., Davies, H., Humphris, G., & Jolly, B. (2002). Generalisability: a key to unlock professional assessment. Medical Education; 36, pp. 972-978.

Curry, C. (2002). Individual Differences in Cognitive Style, Learning Style and Instructional Preference in Medical Education. In G.Norman, P. Cees, & Neuble (Eds.), International Handbook of Research in Medical Education (pp. 249-276). Kluwer Academic Publishers.

Facione, N. C., & Facione, P. A. (1996). Externalizing the critical thinking in knowledge development and clinical judgment. Nursing Outlook, 44, pp. 129-36.

Hackett, G. (1983). The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. Journal of Vocational Behavior, 23, pp. 329-345.

Lent, R., Brown, S., & Larkin, K. (1984). Relation of self-efficacy expectations to academic achievement and persistence. Journal of Counseling Psychology; 31(3), pp. 356-362.

Lent, R., Brown, S., & Larkin, K. (1987). Comparison of three theoretically derived variables in predicting career and academic behaviour: Self-efficacy, interest congruence, and consequence thinking. Journal of Counseling Psychology; 34(3), pp. 293-298.

Nauta, M. M., Epperson, D. L., & Kahn, J. H. (1998). A Multiple-Groups Analysis of Predictors of Higher Level Career Aspirations Among Women in Mathematics, Science, and Engineeirng Majors. Journal of Counseling Psychology; 45(4), pp. 483-496.

Parsons, E., and Betz, N. E. (Winter 1998). Test-Retest Reliability and Validity Studies of the Skills Confidence Inventory. Journal of Career Assessment; 6(1), pp. 1-12.

Ramsden, P. (1983). The Lancaster Approaches to Studying and Course Perceptions Questionnaire: Lecturers' Handbook. The Educational Methods Unit, Oxford Polytechnic.

Rouxel, G. (1999). Path analyses of the relations between self-efficacy, anxiety, and academic performance. European Journal of Psychology of Education; 14(3), pp. 403-421.

SPSS, Inc. (1999). SPSS 11 for Windows (Ver. 11.0.5).

Streiner, D., & Norman, G. (2004). Health measurement scales: A practical guide to their develoment and use (3rd ed.). Oxford: Oxford University Press.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal Setting. American Educational Research Journal; 29(3), pp. 663-676.

Appendix 4.1: Measurement Scales

Example of a Mini-Inquiry Exercise

Please circle the name of your instructor Student Name: ___________________

Instructor A Instructor B Instructor C

Mini-Inquiry Exercise Part I -- 2002-2003

What do we mean by bench marking? Bench marking gives you, as a student in this course, a sense of what your research and critical thinking skills are and how they will change over time. You will complete this exercise at the beginning and end of this course and to compare them so you can see your own progress. These bench marking exercises will help gather some evidence to see how you have done in this course. **But they will not be "marked" like a test or a term paper.**

Please read the following, then work on the questions below.

Am J Med 1998 Sep 28;105(3A):66S-73S

Parallels between post-polio fatigue and chronic fatigue syndrome: a common pathophysiology?

Bruno RL, Creange SJ, Frick NM.

Kids' Fatigue Management Program and The Post-Polio Institute, Englewood Hospital and Medical Center, New Jersey 07631, USA.

Fatigue is the most commonly reported and most debilitating of post-polio sequelae affecting the >1.8 million North American polio survivors. Post-polio fatigue is characterized by subjective reports of difficulty with attention, cognition, and maintaining wakefulness. These symptoms resemble those reported in nearly 2 dozen outbreaks of post-viral fatigue syndromes (PVFS) that have recurred during this century and that are related clinically, historically, anatomically, or physiologically to poliovirus infections. This article reviews recent studies that relate the symptoms of post-polio fatigue and chronic fatigue syndrome (CFS) to clinically significant deficits on neuropsychologic tests of attention, histopathologic and neuroradiologic evidence of brain lesions, impaired activation of the hypothalamic-pituitary-adrenal axis, increased prolactin secretion, and electroencephalogram (EEG) slow-wave activity. A possible common pathophysiology for post-polio fatigue and CFS, based on the Brain Fatigue Generator Model of PVFS, and a possible pharmacotherapy for PVFS based on replacement of depleted brain dopamine, will be described.

PMID: 9790485 [PubMed - indexed for MEDLINE]

1). Write down all the questions, ideas, previous knowledge, etc. that you can think of after reading the above. Please feel free to use the back of the page.

(2) Develop one question that maybe suitable for further research. Indicate in 1-2 sentences why you think it is appropriate.

(3) Identify the steps you would take to research this question. Identify specific sources of information that you possibly would use to find information about this question (e.g., not "www" but specific websites).

(4) Space for comments from peer discussion: (peer comments need to be initialled by their author)

Please circle the name of your instructor

Instructor A Instructor B Instructor C

Student Name: ____________________

Mini-Inquiry Exercise Part II -- 2002-2003

For each activity you participated in today, please describe your strengths and areas needing improvement:

Generating questions, ideas, knowledge, etc.:

Developing a question for further research:

Knowing where to look for an answer:

Did your discussion with a classmate make you think differently about what you had written?

Yes No

If yes, please tell us how in 1-2 sentences.

Example of a Mini-Inquiry Exercise Coding Sheet

Rater Initials	Date:		Id #:			
Topic of Article:	CJD	Post Polio		HPV	Autism	
Appraisal of Arguments/Brainstorming/Question Generation	Excellent		Average		Poor	
Interpretation -- *Conveying or Searching for meaning in the words, phrases and concepts* -- (e.g., Categorization, decoding and clarifying meaning)	5	4	3	2	1	NA
Analysis -- *Conveying or Searching for comprehension of the overall meaning* -- (e.g., examining ideas, identifying and analysing arguments)	5	4	3	2	1	NA
Evaluation – *Wondering/Questioning whether the text makes sense, the findings are valid, reliable* – (e.g, assessing claims and arguments)	5	4	3	2	1	NA
Inference – *Drawing conclusions from the student's understanding of the text, making connections to other knowledge, etc.* – (e.g., querying evidence, conjecturing alternatives, drawing conclusions)	5	4	3	2	1	NA
Search Strategy	Excellent		Average		Poor	
How <u>specific</u> is/are they?	5	4	3	2	1	NA
How <u>likely</u> would it/they yield good data?	5	4	3	2	1	NA
How <u>practical</u> is/are they?	5	4	3	2	1	NA
Self-Assessment Statements	Excellent		Average		Poor	
Willingness to Self-Assess? (Authenticity)	5	4	3	2	1	NA
Open-minded to personal strengths and weaknesses?	5	4	3	2	1	NA
Analytical?	5	4	3	2	1	NA
Self-Confidence?	5	4	3	2	1	NA
Inquisitiveness to multiple points of view?	5	4	3	2	1	NA
Raters' Overall Critical Thinking Assessment	Excellent		Average		Poor	
	5	4	3	2	1	NA

Science Inquiry Academic Self-Confidence Scale (SIASCS): Inquiry Task Self-Confidence Scale Items
Thinking about your academic skills ***in general***, please rate your confidence in your skills in the following areas.
a) ability to work independently
b) ability to work with unfamiliar/new computer software
c) finding information in campus libraries
d) finding information on the Web
e) finding information on library computer data bases
f) posing a research question
g) refining a research question
h) finding an expert for my research
i) interviewing an expert
j) evaluating information on a website
k) evaluating information in a journal
l) evaluating information from E-mail
m) evaluating information from an expert
n) evaluating information in a presentation
o) integrating information from many sources
p) using information to answer a question
q) presenting and selling the answer **orally**
r) presenting and selling the answer **in writing**
s) dividing responsibility when working in a group
t) providing group members with feedback
u) working with difficult people
v) self-evaluation of my academic strengths and weaknesses
w) self-evaluation of my group work skills

Science Inquiry Academic Self-Confidence Scale (SIASCS): Inquiry Task Self-Confidence Scale Items
x) self-evaluation of my time management skills
y) overall, I have (blank) about my skills

Science Inquiry Academic Self-Confidence Scale (SIASCS): Projected Academic Goal Completion Self-Confidence Scale Items
How confident are you, ***right now,*** to complete the following tasks?
a) Complete the math requirements for most science or business **undergraduate** degrees
b) complete the math requirements for most math or engineering **undergraduate** degrees
c) complete a science or business **undergraduate** degree
d) complete a math or engineering **undergraduate** degree
e) ask a professor if you could volunteer in their lab
f) ask a professor for an academic reference
g) call a medical school admissions officer for information about their program
h) call a graduate school admissions officer for information about their program
i) be accepted into **graduate** program
j) be accepted into medical or dental school

Approaches to Learning Scale Items (Lancaster Scale) Items
Meaning Orientation
I try to relate ideas in one subject to those in others, whenever possible.
In reporting practical work, I like to try to work out several alternative ways of interpreting the findings.
I find that studying academic topics can often be really exciting and gripping.
Puzzles or problems fascinate me, particularly where you have to work through the material to reach a logical conclusion.
When I'm tackling a new topic, I often ask myself questions about it which new information should answer.
I am usually cautious in drawing conclusions unless they are well supported by evidence.
When I'm reading an article or research report, I generally examine the evidence carefully to decide whether the conclusion is justified.
I find academic topics so interesting I should like to continue with them after I finish university.
I find it helpful to 'map out' a new topic for myself by seeing how the ideas fit together.
I spend a good deal of my spare time in finding out more about interesting topics which have been discussed in class.
It's important to me to do really well in the courses here.
I usually set out to understand thoroughly the meaning of what I am asked to read.
I often find myself questioning things I hear in class or read in books.
If conditions aren't right for me to study, I generally manage to do something to change them.
My main reason for being here is so that I can learn more about the subjects which really interest me.
I generally put a lot of effort into trying to understand things which initially seem difficult.
I need to read around a subject pretty widely before I'm ready to put my ideas down on paper.
Reproducing Orientation
It's important to me to do things better than my friends, if I possibly can.
I like to be told precisely what to do in essays or other set work
When I'm doing a piece of work, I try to bear in mind exactly what that particular professor seems to want.
Often I find I have to read things without having a chance to really understand them.
I usually don't have the time to think about the implications of what I have read.
I tend to read very little beyond what's required for completing assignments.

Approaches to Learning Scale Items (Lancaster Scale) Items
When I am reading, I try to memorize important facts which may come in useful later.
The best way for me to understand what technical terms mean is to remember the text-book definitions.
I'm more interested in the qualifications I'll get than in the courses I'm taking.
Lecturers seem to delight in making the simple truth unnecessarily complicated.
I prefer courses to be clearly structured and highly organized.
I hate admitting defeat, even in trivial matters.
I find I have to concentrate on memorising a good deal of what we have to learn.
Organized Orientation
I find it easy to organize my time effectively.
Distractions often make it difficult for me to do effective work in the evenings.
I'm usually prompt in starting work in the evenings.
My habit of putting off work leaves me with far too much to do at the end of term.

CHAPTER 5

Study Design of the McMaster Science Inquiry Project Program Evaluation: Academic Years 2000 to 2004

Study Design Rationale

How do you know which elements of a particular intervention made a difference in how a student engaged in learning? Many educational interventions in higher education were successful when championed by a dedicated and charismatic educator, but these successes were not sustained when the champion moved onto other endeavours. The purpose of this dissertation was to examine if certain elements of the McMaster University Inquiry interventions were effective as implemented in the Faculties of Science and Health Science. If these interventions were effective, what elements seem to work best? Is it the small groups that students work with? Or is it that they work together to teach and learn from each other when defining a group project together? Is it the opportunities to work closely with a faculty member in small classes? Or is it the contact with their peers and upper year students? Is it the opportunity to work on university level skills in searching and evaluating information sources? Or is it the opportunity to get feedback promptly from peers and professors on what they are working on? Is it being encouraged to think deeply about subjects they are interested in?

These Inquiry course elements were common to the three interventions reported on in this study. More importantly, these elements were consistent with the Seven Principles of Good Teaching (Chickering and Gamson, 1987) because they were active learning strategies. Most importantly, they were replicable in that any instructor would be able to implement some or all of these in any higher education setting.

Educational field research can be difficult at the best of times. The classroom context makes it hard to ensure that the intervention was administered faithfully and with the same intensity at each administration. Good educational practice dictates that the instructor varies their lessons to meet the needs of their students. Factors outside of the specified intervention, such as the instructor's personality, peer interactions, or institutional changes, can have a substantial impact on the students' performance. Also, when compared to classical experimental designs that may be conducted in hours, field research may be conducted over several months. While many of these factors can have a positive impact on the quality of the students' experiences, these changes can also make it difficult to measure the variable of interest in a research study. As Norman (2003) stated, educational field research has ample documentation of positive treatment effects for various educational interventions but these treatment effects "cannot be subsumed under a single 'treatment' like 500 mg. of PBL t.i.d." (p.185).

A second difficulty was the notion of "controlling for confounding" in the experimental design. Kember (2003) and Norman (2003) both cite difficulties in applying strict experimental design criteria to educational research. Kember noted that students will not tolerate being randomly assigned to a treatment or control group if being in the control group will negatively impact on the quality of their educational experience.

Conversely, when the innovation being tested had a positive impact on the treatment students' educational experience, students assigned to the treatment group were likely to recommend the innovation to those students in the control group: while the student self-report evidence clearly indicated that the intervention was effective, the measure of a 'treatment' difference will not be seen since both groups experienced the intervention.

The design of this study was intended to ameliorate the contextual factor effects of intervention specificity, intensity, and diffusion. In this study we did the opposite of trying to isolate the variable of 'Inquiry' from the other 'noise' in the environment. In fact, we measured the variable 'Inquiry' in multiple contexts so as not to 'hang our hat' on one narrow time point or single intervention which may lead to misleading results. We measured the outcomes of interest in "Inquiry" over:

- three academic years
- at two time points throughout the academic year
- using two types of quasi-experimental control groups (i.e., longitudinal and cross-sectional differences)
- over three separate Inquiry interventions
- with each intervention having several instructors teaching the course
- in two separate faculties.

If the variable 'Inquiry' was having an impact on students' learning, surely we would see the pattern of its impact in this multifaceted approach. In this way, the results of the study can be triangulated to build a composite picture of the Inquiry intervention to measure whether it has had any impact on student learning and also to examine its impact in various contexts.

The following chapter outlines the procedures and methods for the three years of data collection from the students' participation in Inquiry. We chose to look at three consecutive years of data to ensure that the Inquiry intervention was sustainable and not merely "hot house" phenomena that may not be transferable or durable over the long term in field conditions. In addition, we collected follow-up data of examining the differences of graduating students who had taken Inquiry in year one compared to those who had not taken Inquiry in year one.

We wanted to triangulate the data sets by having two experimental designs (Kember, 2002): a longitudinal design for intra-group differences and a cross-sectional design to comparing the post intervention group differences to a group of first year students who did not take Inquiry at the end of the winter term. We wanted to ensure that we were measuring the effects of the intervention and not solely normal developmental milestone changes.

Similarities and Differences between the Three Inquiry Interventions

In terms of class organization, all courses involved small-group inquiry learning where the students could negotiate among their group members their own topic to study, computer discussion groups, and in-class discussion. However, the first year courses were offered in small sections of 20 to 30 students per instructor while the fourth year course had one instructor for the entire class. Also, the first year course had activities specifically designed to teach students how to evaluate information, use the library, work with peers, and self-assess their strengths and weaknesses. The fourth year course assumed that the students had learned these skills previously.

First Year Science Inquiry: Science Inquiry (Science) was a first-year inquiry course focusing on the establishment of the student as an active learner. For the Science Program, the first year Inquiry course served as an introduction to systematic investigation in an area of academic interest. Using a scientific problem, it strived to help students develop skills that will benefit them throughout their university career. Students learned how to formulate questions, gather and interpret evidence, and present their conclusions orally and in writing. Relative to the other large lecture courses science students were taking (ie. Chemistry, Biology, Psychology, Math, etc.), the small class size in Science Inquiry (n ~25) represented an opportunity to work closely with a professor and their fellow students on academic problems.

The student population enrolled in this course were first-year Science students, with academic averages required for entry into the program ranging from 80-83%. Of the approximately 700 students in the Science I program at McMaster University, approximately 70 to 150 elected to enroll in the course. Time requirements include one hour per week in semester one and two hours per week in semester two (i.e., 36 hours of instruction).

The evaluation of students was conducted using a number of tools that may vary among the Inquiry sections because of faculty member preferences. Students were expected to demonstrate competency in areas such as group work, scholarly writing, presentations, discussions and debates, and were given every opportunity to develop and improve these skills throughout the year. There were also opportunities for ongoing peer and self-evaluations with guidance from peer tutors and the faculty member. A key component of the Inquiry evaluation was the collaboration between faculty member and student to develop expectations that are unique to each individual.

The facilitators for Inquiry were from a variety of backgrounds and disciplines within the faculty of Science, such as mathematics, biology, psychology and health sciences, and from McMaster's Center for Leadership and Learning. The teaching team was organized so that there was a three-year rotation of instructors and new instructors could apply their skills and expertise to the Inquiry course with the support of third-year instructors who had taken on a leadership role. All facilitators were academics and were experts on a variety of scientific issues.

First Year Health Sciences Inquiry: The Bachelor of Health Sciences (HSc) Inquiry course strived to introduce and develop the capacity to research and analyse complex problems and to communicate easily and effectively. Six particular elements were the focus of this course: asking good questions; determining what needs to be learned to answer those questions; identifying appropriate resources for learning; using resources effectively and reporting on what was learned; communicating effectively with peers and professors; and the ability to self-evaluate their academic performance.

The HSc Inquiry course was a first-year Inquiry course focusing on the establishment of the student as an active learner. The student population enrolling in this course consisted of first-year HSc students and a select few second-year students transferring into the program. Relatively high academic averages are required for entry into the program (approx. 90% overall) and all students registered in the HSc were required to enroll in Inquiry. Relative to the other large lecture courses students may be taking (eg. Chemistry, Biology, etc.), the small class size (n ~ 20) in Inquiry represents an opportunity to work closely with a professor and their fellow students on academic areas of interest. Taking place over two semesters (i.e., 72 hours of instruction), the course enabled students to apply the principles of scientific inquiry to one or two selected health care issues, simultaneously encouraging the development of critical thinking and communication skills essential for life-long learning.

The evaluation of students were conducted using a number of tools that varied among the Inquiry sections. Students were expected to demonstrate competency in areas such as group work, scholarly writing, presentations, discussions and debates, and were given every opportunity to develop and improve these skills throughout the year. There were also opportunities for ongoing peer and self-evaluations with guidance from peer tutors and the facilitator. A key component of the Inquiry evaluation was the collaboration between facilitator instructor and student to form expectations that are unique to each individual.

The facilitators for Inquiry are a stable group of instructors who come from a variety of backgrounds and disciplines, such as pharmacy, psychoanalysis, psychology, medicine, medical science, social work, and librarianship. All hold professional degrees and are experts in group-work, interested in teaching basic research and interpersonal communication skills which are deemed to be essential to HSc Inquiry.

Fourth Year Science Inquiry: (4th Year) Biology 4V03 was a fourth-year course focusing on the field of virology. The course discussed the viruses of animals, bacteria, and plants, with an emphasis on the molecular biology of virus replication and the diversity of virus-cell interactions. Taking place over thirteen weeks (i.e., 36 hours), the course began by introducing students to viruses in several traditional lecture periods. Following the lectures, the students were placed in small groups in which students work together in taking a self-directed inquiry approach to their learning. Support was provided in tutorial sessions, a suggested textbook - 'Fields Virology', and through the use of a customized version of First Class – an electronic bulletin board serving to facilitate peer support and interaction.

The evaluation criteria for this course involved a number of components, many of which have been modified in accordance with the students' needs in the four years this course was followed. In its most recent year, the primary source of the students' grade (55%) reflects their participation and work in the inquiry groups. Students were also required to hand in weekly journal submissions, summarizing new information and demonstrating its integration into their knowledge base. Students were permitted to choose two entries for final grading (40%). The final component (5%) involved the exploration of a scholarly education paper, for which each inquiry group creates a 200-word summary of its findings and a 400-word description relating the paper to personal experiences at McMaster University. In addition, students were given the opportunity to have a video-taped interview with the instructor, both in the inquiry groups and individually, and may later review the tape for the purposes of a self-evaluation.

The instructor for this course was well known by students and faculty alike as an innovative educator and excellent lecturer. As a professor of pathology & molecular medicine and biology, he strived to create an environment in which individuality and cooperation were encouraged. A focus was placed on allowing students to create their own schedules and custom-designing learning plans to each individual. Together in small groups and as a class on LearnLink, students worked in an active learning environment, coming together and demonstrating collegiality among peers. As a result of his unique style of teaching, the professor's teaching evaluations attest to his success as an instructor and have consistently been well above the departmental average over the past several years.

Research Purpose

The fundamental research questions asked in the McMaster Science Inquiry program evaluation of the Science and Health Science Faculties were:

1. Did Inquiry foster the attributes of critical thinking, academic self-efficacy or confidence, and a preference to delve deeply into subject areas instead of focusing on memorizing solely the textbook materials for the test?
2. If it did foster these attributes by the end of the Inquiry course, were these gains in evidence when these students were ready to graduate?

Table 5.1 outlines the common elements between the three interventions and shows how these common course elements are related to the theoretical constructs under study. The hypotheses for both study designs are outlined in Tables 5.2 and 5.3.

Table 5.1: Common Elements between Inquiry Interventions and the Learning Domains Being Studied

Common Curriculum Elements Between Inquiry Interventions	Theoretical Construct Being Promoted			
	Critical Thinking Skills	Current and Projected Inquiry Task Self-Confidence	Meaning Orientation and Discourages Reproducing Orientation	Organized Orientation
Mini-Inquiry Exercise	x	x	x	x
Small Group Work allowing students to work on problems of their choosing	x	x	x	x
Individual interview with instructor	x	x	x	x
Interviewing Experts	x	x	x	x
Group Presentations/ Final Video taped group meetings	x	x	x	x
LearnLink (First Class Client) electronic course conference system		x		x
Discussion about time management		x		x
Small class sizes		x	x	
In class discussion in large groups	x	x	x	
Evaluating journal articles	x	x	x	
Discussions about critical thinking	x		x	
Learning to evaluate information from web based sources	x	x	x	
Informal Discussions with peers	x	x	x	
Exercises in how to use the library	x	x	x	
Written Self-Evaluations	x	x	x	
Written Group Evaluations	x	x	x	

Figure 5.1: Experimental Design One of Entry and Exit Scale Scores for the First Three Years of Data Collection in all Samples

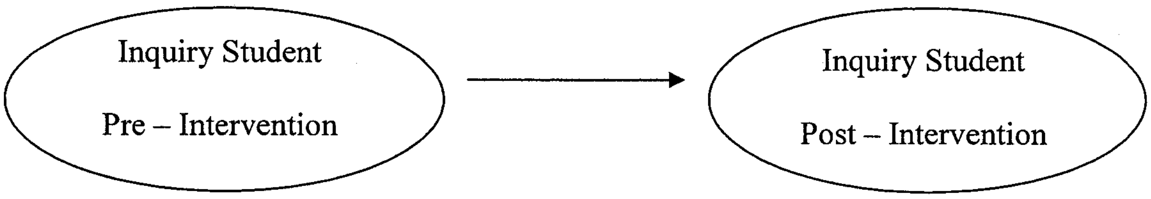


Table 5.2 Longitudinal Comparisons Research Questions and Hypotheses for the First Three Years of Data Collection in All Samples

Mini-Inquiry Benchmarking Exercise Measuring Critical Thinking Skills
Hypothesis for all three years of all samples: That Inquiry students in an inquiry intervention will improve skills development on: 1. appraisal of arguments 2. information search strategy 3. self-assessment of critical thinking skills 4. and the raters' opinion of the students' overall critical thinking skills as measured by their beginning of the course scores compared to their end of course scores.
Approaches to Learning Scores
Hypothesis on years two and three for all samples: that Inquiry students in an inquiry intervention: 5. will show a greater preference for a meaning orientation 6. will show a decreased preference for a reproducing orientation 7. will show a greater preference for an organized orientation as measured by their beginning of the course scores compared to their end of course scores.
Science Inquiry Academic Self-Confidence Scale (SIASCS)
Hypothesis on years two and three for all samples: that Inquiry students in an inquiry intervention will improve on: 8. inquiry task self-confidence attitudes 9. projected inquiry task self-confidence attitudes as measured by their beginning of the course scores compared to their end of course scores.

Figure 5.2: Cross-sectional Experimental Design for First Three Years of the Study

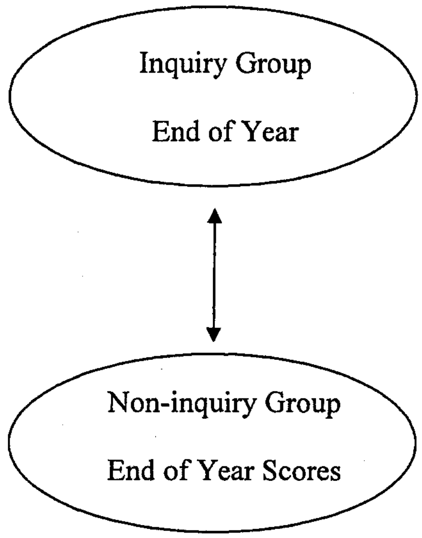


Table 5.3 Research Hypothesis for Cross-sectional Research Design in the First Three Years of the Study with Inquiry and Non-Inquiry Students

Approaches to Studying Preference Scale
Hypothesis for years one to three for all samples: that inquiry students' end of semester scores on: 8. will show a greater preference for a meaning orientation 9. will show a decreased preference for a reproducing orientation 10. will show a greater preference for an organized orientation when compared to the non-inquiry students' end of semester scores.
Science Inquiry Academic Self-Confidence Scale (SIASCS)
Hypothesis on years one to three for all samples: that Inquiry students' end of semester scores will be higher on: 10. inquiry task self-confidence attitudes 11. projected inquiry task self-confidence attitudes when compared to the non-inquiry students' end of semester scores.

Samples from the Initial Data Collection Period

First year science students (Science): All possible students enrolled in first year of the course who had completed course were included in the data analyses. Enrollment in first year science inquiry was restricted to students who were in their first year of university, in the Science Faculty although there were some students who were from Nursing (1%) and Kinesiology (1%). Students who completed the course were retained for data analyses. Total enrollment in the course increased over the three years of study (2000-2001 = 113; 2001-2002 = 102; 2002-2003 = 150). Inquiry students were similar in age, and gender ratio compared to the first year Calculus students' demographic profile reported in Chapter 4

First year Health Science students (HSc): All students enrolled in their first year of the HSc program, including the students who transferred into the program during their second year from another undergraduate program, were required to take the first year Inquiry course: all completed data collection instruments were analyzed for these students. The HSc program was a limited enrollment program where students need to have a minimum of a 90% high school average in their final year of high school. Many of these students were intending to gain admission to a health profession (i.e., medicine, dentistry, veterinary medicine, etc.). The number of students entering the program has increased since its inception (2000-2001) from 80 to 160 students per year. This group has been consistent in their demographic characteristics (e.g., gender, age at admission, and language spoken most often in the parental home) when compared to their admission characteristics.

Fourth Year Molecular Biology Students (4th Year): Fourth year molecular biology students were enrolled in an elective course that had an enrollment of about 100 to 160 students per year. This fourth year course was taught by an experienced instructor who had been teaching inquiry courses in both science and HSc first year inquiry programs. The sample of fourth year students represented many different programs, but given the large sample size (x = 160/310 total number of graduates per year from Faculty of Science) it was representative of the graduating class for age, gender and language most often spoken in the parental home

Non-Inquiry Students: Students who participated in the non-inquiry group or control group were recruited from the first year psychology course. Students were eligible to participate if they were not currently or previously enrolled in a first year science or HSc inquiry course. Students received either course credit or money for their participation. This segment of the study received ethics approval from the McMaster Ethics Review Board.

There were difficulties in recruiting students from the psychology course. In the 2000-2001 year, there was a teaching assistant strike which caused severe disruption to the University: the sample size in this year was 18 students. Also, recruiting students at the end of the year usually resulted in the least motivated students participating since highly motivated students would schedule their participation earlier in the year. Therefore, in the third year of the study, we recruited students to complete all instruments at the beginning of the academic year and encouraged students to return and complete post-test data collection forms for a more heterogeneous, and hopefully representative sample of non-inquiry first year students.

Many students taking the first year psychology course were not first year or first-degree students. In general, they were older students, some of whom had a previous degree and significant work experience. However, all students were reported here because they met the eligibility requirements, and admitting them would maintain our sample size and would give a more conservative estimate of the treatment effect.

Timelines and Setting: On the first three consecutive years, first and fourth year inquiry students (n ~ 300 per year) were asked to complete the Mini-Inquiry Benchmarking Exercise in the first and last month of their respective courses. For the first year of the study, the inquiry students (n ~ 300 per year) were asked to complete a course evaluation survey which contained the Science Inquiry Academic Self-Confidence and the Approaches to Studying Preferences scales at end of the course. In the second and third years of the study, the inquiry students (n ~ 300 per year) were asked to complete pre-test and course evaluation surveys which contained the Science Inquiry Academic Self-Confidence and the Approaches to Studying Preferences scales at the beginning and end of the course.

For the first two years of data collection, the non-inquiry control group, students enrolled in the first year psychology course but not enrolled in first year Inquiry, were asked by their tutorial leaders during regular tutorial sessions at the end of the year to complete a survey with the Approaches to Learning, Science Inquiry Academic Self-Confidence scales, as well as answer some demographic questions.

For the third year of the study, non-inquiry control group students were asked to sign-up on posters in the designated psychology credit sign-up area at the beginning of the first year psychology course for either course credit or financial remuneration. These students were also advised that we wanted them to complete the same set of instruments in the following March for additional course credit or financial compensation. Students, who had completed the September instrument administration, were contacted by e-mail and invited to complete the second session. Additional students were recruited from the sign-up posters in the designated psychology credit sign-up area to increase the power to detect true differences in the cross-sectional comparisons between inquiry and non-inquiry participants. This portion of the study was approved by the McMaster Research Ethics Review Board.

The finalized tools to assess the students' inquiry skills and attributes have been described in detail in Chapter Four and will not be described here. The following tables (Tables 5.4 to 5.6) detail when the instruments were administered for each of the study samples by year of study.

Table 5.4: Instrument Administration for the First Year of the Study

M = Mini-Inquiry	Fall Semester				Winter Semester			
S = Survey including the Approaches to Studying and SIASCS scales	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
First Year Non-Inquiry							S	S
First Year Science Inquiry	M						M	S
First Year Health Science Inquiry	M						M	S
Fourth Year Science Inquiry	M		M	S				

Table 5.5: Instrument Administration for the Second Year of the Study

M = Mini-Inquiry S = Survey including the Approaches to Studying and SIASCS scales	Fall Semester				Winter Semester			
	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
First Year Non-Inquiry							S	S
First Year Science Inquiry	M S						M	S
First Year Health Science Inquiry	M S						M	S
Fourth Year Science Inquiry	M S		M	S				

Table 5.6: Instrument Administration for the Third Year of the Study

M = Mini-Inquiry S = Survey including the Approaches to Studying and SIASCS scales	Fall Semester				Winter Semester			
	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
First Year Non-Inquiry	M S	M S					M S	M S
First Year Science Inquiry	M S						M	S
First Year Health Science Inquiry	M S						M	S
Fourth Year Science Inquiry	M S		M	S				

Data Analyses for the First Three Years of the Study

Mini-Inquiry Benchmarking Exercise: On the mini-inquiry benchmarking exercise, ANCOVA analyses were conducted on each sub-scale to estimate the adjusted mean group differences on the pre-test to post-test scores for each Inquiry course. The balanced model took into account the following factors as main effects: time (beginning or end of the course); course (first year Science, first year Health Science or fourth year Molecular Biology); article (which standardized article); rater (three raters); and the individual differences. The interaction effects were also entered into the model taking into account the nesting of the individual student within their enrollment in each course. The F values are reported to give the reader an indication of the magnitude of factor's impact on the model.

However, it is to be remembered that each inquiry course is a separate intervention and needs to be considered as such. To compare inquiry interventions against each other is not answering the research question of whether Inquiry interventions *per se* foster critical thinking skills, approaches to studying preferences and self-confidence attributes. As there was no interaction between the rater and cohort effects or rater and time effects, the main effects of rater and time were considered to be independent. A separate ANCOVA was used to calculate the post-hoc comparison significance values for each study sample. The revised model had fewer terms (i.e., rater, time and individual differences) and, with fewer terms, the model had sufficient power for each sample to be calculated separately. These parameter estimates, although in a poorer fitting model, were sufficient to be used as post –hoc comparisons in considering if the mean differences between each study sample were statistically significant.

Approaches to Studying and Science Inquiry Academic Self-Confidence Scores: On the Science Inquiry Academic Self-Confidence Scales and Approaches to Learning scales data, paired samples t-test were conducted to compare group means from the beginning to the end of the course scores for the pre-test to post-test studies and independent samples t-test were conducted to compare group means between each inquiry course and the non-inquiry course for the cross-sectional studies. As with the critical thinking scores, each inquiry intervention was compared separately because each inquiry course is considered a separate intervention. The significance levels from these t-tests are reported in the Results tables in Chapter Six. To adjust for multiple inferential tests, the critical value for statistical significance ($p < 0.05$) was divided by the number of t-tests for each sub-score per year.

Concerning All Reported Scores in the First Three Years of the Study: All scores have a maximum value of 100: therefore, mean scores can also be considered to be percentage scores. Effect sizes were calculated by dividing the difference between the two mean scores (i.e., difference between entry mean score and exit mean scores) by the standard deviation of the mean entry score. When comparing effect sizes for significant differences from the null hypothesis (i.e., $H_0 = 0$) the magnitude of these effect sizes is based on the work of Cohen (1992, p. 157). According to Cohen, small effect sizes are negative or positive values between 0.20 and 0.49, medium effect sizes are negative or positive values between 0.50 to 0.79 and large effect sizes are negative or positive values above 0.80. Effect sizes reported for each sub-score were averaged for both longitudinal and cross-sectional data with the first three years of the inquiry intervention groups. The studies were performed over three years to ensure that the treatment differences were durable in field conditions. Average effect size scores are reported for the three years of Inquiry intervention group data and the single years' longitudinal data is reported for the non-inquiry control group.

Follow-up Study Methods with the Upper Year Inquiry and Non-Inquiry Students

The study methods for the follow-up study with the upper year inquiry and non-inquiry students are presented separately.

Samples for Follow-up Studies

Science Inquiry Students and Health Science Students: Students who had previously taken a first year Inquiry course in their respective faculty were contacted in their upper years of studies (i.e., third or fourth year) to participate in the follow-up study. Students were contacted through their e-mail address and asked to arrange a time to complete all three instruments. Appointments were scheduled so that more than one student would be present enabling the participant to have a peer with which they could get peer feedback on their mini-inquiry exercise. Students received financial compensation for their participation and were able to see their first year mini-inquiry papers so that they could compare their progress. These data were collected between November 2003 and February 2004. This segment of the study received ethics approval from the McMaster Ethics Review Board.

Non-Inquiry Students: When examining the same first year students in their third and fourth years of the program, a comparison group was drawn from the fourth year Molecular Biology course which attracts approximately 50% of the graduating Faculty of Science students at McMaster University: these students were verified from previous first year class lists as not having taking a first year Inquiry course. Students who were enrolled in this course and had taken either first year Health Science or Science inquiry were reallocated to the correct intervention group. As part of their in-class work, students were asked to complete a mini-inquiry exercise at the beginning and end of the term. However, only the mini-inquiry exercise used at the end of term was collected for analyses. The Science Inquiry Academic Self-Confidence Scale and Approaches to Studying (Lancaster Scale) were completed as part of the end of term course evaluation survey. This segment of the study received ethics approval from the McMaster Ethics Review Board.

The cross-sectional study design is presented in Figure 5.3 and the hypotheses are outlined in Table 5.7.

Figure 5.3: Cross-sectional Experimental Design for the Follow-up Study with the Upper Year Students

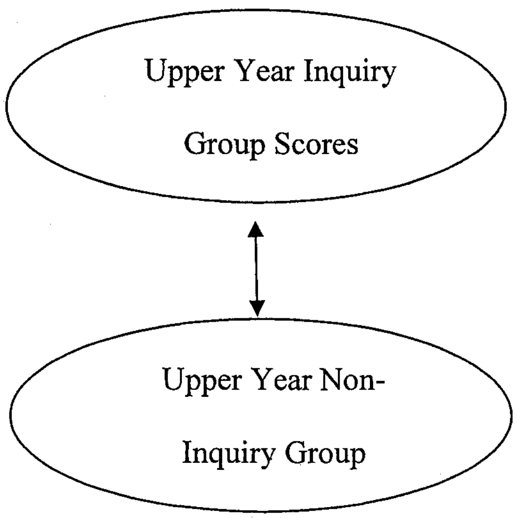


Table 5.7 Research Hypothesis for Cross-sectional Research Design in the Follow-up Study with Upper Year Inquiry and Non-Inquiry Students

Mini-Inquiry Benchmarking Exercise Measuring Critical Thinking Skills
Hypothesis for the follow-up year -- that the upper year students' who took the first year inquiry intervention will have higher scores on: 1. appraisal of arguments 2. information search strategy 3. self-assessment of critical thinking skills 4. and the raters' opinion of the students' overall critical thinking skills when compared to the upper year non-inquiry students' scores.
Approaches to Studying Preference Scale
Hypothesis for the follow-up year -- that the upper year students' who took the first year inquiry intervention: 5. will show a greater preference for a meaning orientation 6. will show a decreased preference for a reproducing orientation 7. will show a greater preference for an organized orientation when compared to the upper year non-inquiry students' scores.
Science Inquiry Academic Self-Confidence Scale (SIASCS)
Hypothesis for the follow-up year -- that the upper year students' who took the first year inquiry intervention will have higher scores on: 8. inquiry task self-confidence attitudes 9. projected inquiry task self-confidence attitudes when compared to the upper year non-inquiry students' scores.

Table 5.8: Follow-up of Graduating Students Year of Study

M = Mini-Inquiry S = Survey including the Approaches to Studying and SIASCS scales	Fall Semester				Winter Semester			
	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Graduating Students not taking a First Year Inquiry			M S	M S				
Graduating Student who took First Year Science Inquiry			M S	M S	M S	M S		
Graduating Students who took First Year Health Science Inquiry			M S	M S	M S	M S		

Data Analyses for the Follow-up study with the Upper Year Students

Mini-Inquiry Benchmarking Exercise: On the mini-inquiry benchmarking exercise, ANCOVA analyses were conducted on all sub-scales to estimate the adjusted mean group differences on the scores for each of the samples. The balanced model took into account the following factors as main effects: course (first year Science Inquiry participation, first year Health Science Inquiry participation or no first year inquiry course participation); rater (three raters); and the individual differences. The interaction effects were also entered into the model. The F values are reported to give the reader an indication of the magnitude of factor's impact on the model.

A separate ANCOVA was used to calculate the post-hoc comparison significance values for each study sample. The revised model had fewer terms (i.e., rater, time and individual differences) and, with fewer terms, the model had sufficient power for each sample to be calculated separately. These parameter estimates, although in a poorer fitting model, were sufficient to be used as post –hoc comparisons in considering if the mean differences between each study sample were statistically significant.

Science Inquiry Academic Self-Confidence Scales and Approaches to Learning Scales: For the follow-up data analyses, cross-sectional comparisons (independent samples t-tests) are reported for the upper year inquiry students and the upper year science students who did not take first year inquiry separately. To adjust for multiple inferential tests, the critical value for statistical significance ($p < 0.05$) was divided by the number of t-tests for each measure.

Concerning All Sub-Scales for the Follow-up Study: All scores have a maximum value of 100: therefore, mean scores can also be considered to be percentage scores. Effect sizes are calculated by dividing the difference between the two mean scores (i.e., difference between entry mean score and exit mean scores) and dividing these scores by the standard deviation of the mean entry score. When comparing effect sizes for significant differences from the null hypothesis (i.e., $H_0 = 0$) the magnitude of these effect sizes is based on the work of Cohen (1992, p. 157). According to Cohen, small effect sizes are negative or positive values between 0.20 and 0.49, medium effect sizes are negative or positive values between 0.50 to 0.79 and large effect sizes are negative or positive values above 0.80.

References

Chickering, A. W., & Gamson, Z. F. (1987). Seven Principles for Good Practice in Undergraduate Education. AAHE Bulletin March 1987.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* pp. 155-159. Reprinted in Methodological issues & strategies in clinical research (2nd ed.) Alan Kazdin (Ed.). Washington DC: APA

Kember, D. (2003). To control or not to control: the question of whether experimental designs are appropriate for evaluating teaching innovations in higher education. Assessment & Evaluation in Higher Education, 23(1), pp 89-101.

Norman, G. (2003). Editorial – The Effectiveness and the Effects of Effect Sizes. Advances in Health Science Education, 8, pp 183-187.

CHAPTER 6

Results of the McMaster Science Inquiry Project Program Evaluation Academic Years 2000 - 2004

Results from the First Three Years for All Inquiry Interventions and the Non-Inquiry Group Comparisons

For tables describing each of the study samples in the first year of the study, see Tables 6.1 to 6.3. Details on each scores' data by each year of the intervention, with the grouped means, their respective mean differences, effect sizes, and whether or not these differences are statistically significant, are given in detail in the tables provided in Appendix 6.1 following this chapter.

Critical Thinking Scores Results

For the information on the adjusted mean differences from the ANCOVA analyses for the Critical Thinking scores at entry and exit in the first three years of the study, see Tables 6.10 to 6.14. From the ANCOVA data from the first three years of the mini-inquiry benchmarking data, the F values for the main effect of **time** for the overall model including rater, study sample (course), article, individual differences were as follows:

1) Appraisal of Arguments scores for:
    i) 2000-2001 df (1, 273) $F = 37.87$, $p = 0.0001$;
    ii) 2001-2002 df (1, 240) $F = 3.91$, $p = 0.049$; and,
    iii) 2002-2003 (with non-inquiry control group) df (1, 328) $F = 0$, $p = 0.988$.

2 Information Search Strategy scores for:

i) 2000-2001 df (1, 273) $F = 33.86$, $p = 0.0001$;

ii) 2001-2002 df (1, 240) $F = 24.85$, $p = 0.0001$; and,

iii) 2002-2003 (with non-inquiry control group) df (1, 328) $F = 29.73$, $p = 0.0001$;

3 Self-Assessment scores for:

i) 2000-2001 df (1, 273) $F = 46.12$, $p = 0.0001$;

ii) 2001-2002 df (1, 240) $F = 7.15$, $p = 0.008$; and,

iii) 2002-2003 (with non-inquiry control group) df (1, 328) $F = 0$, $p = 0.939$;

4 Overall Critical Thinking Assessment for:

i) 2000-2001 df (1, 273) $F = 42.80$, $p = 0.0001$;

ii) 2001-2002 df (1, 240) $F = 13.27$, $p = 0.0001$; and,

iii) 2002-2003 (with non-inquiry control group) df (1, 328) $F = 1.90$, $p = 0.169$.

The overall model scores were a conservative estimate of the within subject differences for time because the model includes all Inquiry intervention sample comparisons. Also, the 2002-2003 F values should be interpreted with caution since they include comparisons between the inquiry interventions and the non-inquiry control group: as a result, the difference between the entry and exit scores may be blunted because one of the groups was not in the inquiry intervention.

In the revised model of rater, time and individual differences, used to calculate the post-hoc significance (p values), determined whether the beginning to end of intervention comparisons for each treatment group were significant. The significant post hoc comparisons are in bolded print in Appendix 6.1, Tables 6.14 to 6.18.

Table 6.1: Age, Gender and Language Spoken at Home for All Samples at the end of Each Course

Study Year One			
	Age (Mean, Standard Deviation)	Gender (Percentage Females)	Language Spoken in Parental home not English (%)
Control Group n = 16	19.2 (1.54)	56%	13%
Science Inquiry n = 80	19.6 (2.11)	76%	29%
HSC n = 66	19.1 (0.42)	82%	9%
4th Year Inquiry n = 103	22.1 (1.40)	59%	22%
Study Year Two			
Control Group n = 26	20.2 (3.80)	76%	15%
Science Inquiry n = 87	19.1 (0.70)	62%	26%
HSC n = 88	19.2 (0.74)	69%	21%
4th Year Inquiry n = 81	22.1 (1.59)	62%	19%
Study Year Three			
Control Group n = 47	23.6 (7.40)	75%	21%
Science Inquiry n = 73	19.1 (0.77)	57%	20%
HSC n = 114	19.1 (0.72)	58%	20%
4th Year Inquiry n = 129	22.1 (1.07)	64%	24%

Table 6.2: Students Work, Volunteer and School Commitments relative to Course Load

	Percentage Worked less than 5 hours per week during school year	Percentage Volunteered less than 5 hours per week during school year	Percentage Taking a full course load
Study Year One			
Control Group n = 16	87	12	87
Science Inquiry n = 80	96	100	92
HSC n = 66	98	100	95
4th Year Inquiry n = 103	94	5.6	80
Study Year Two			
Control Group n = 26	92	92	93
Science Inquiry n = 87	97	97	96
HSC n = 88	98	100	95
4th Year Inquiry n = 81	99	100	77
Study Year Three			
Control Group n = 47	57	96	56
Science Inquiry n = 73	77	0	95
HSC n = 114	86	84	100
4th Year Inquiry n = 129	67	74	48

Table 6.3: Student Responses to the Survey Question "Have skills developed in Inquiry helped you with other courses taken in first year, future years and your career?"

Study Year One			
Have skills developed in Inquiry helped you with your … ?			
	First Year Courses? (Percentage Yes)	Upper year Courses? (Percentage Yes)	Career? (Percentage Yes)
Science Inquiry n = 80	88	96	85
HSC n = 66	96	99	99
4th Year Inquiry n = 103	84	95	91
Study Year Two			
Science Inquiry n = 87	91	96	82
HSC n = 88	99	99	98
4th Year Inquiry n = 81	89	95	91
Study Year Three			
Science Inquiry n = 73	92	99	81
HSC n = 114	95	97	97
4th Year Inquiry n = 129	75	84	84

- Responses were ranked yes, not sure or no

Approaches to Studying and Science Inquiry Academic Self-Confidence Scales

For the information on the mean differences from the paired samples t-tests analyses comparing entry and exit scores for the Approaches to Learning and Inquiry Self-Confidence scores in the second and third years of the study, see Tables 6.15 to 6.19 in Appendix. For the information on the mean differences from the independent samples t-test analyses comparing cross-sectional differences for the Approaches to Learning and Inquiry Self-Confidence scores in the first three years of the study, see Tables 6.20 to 6.24.

Effect Size Results for All Scales

Overall, there were 138 separate outcome measurements statistics derived from this program evaluation. For clarity's sake, the longitudinal (entry compared to exit studies) and cross-sectional (treatment vs control) statistics for the first three years of the Inquiry intervention were aggregated into average effect sizes for each of samples under study. With these summary statistics, it was easier to discuss the trends in the data. Summary tables of all mean effect sizes were presented in Tables 6.4 through 6.5. Summary figures showing the trends within each study sample were shown in Figures 6.1 through 6.4.

Table 6.4 Summary of Mean Effect Sizes, Ranges for the Longitudinal Critical Thinking Measures for the First Three Years of Inquiry Students

Mean (Minimum and Maximum)	First Year **Non Inquiry Group** Effect Size	Mean Effect Size of First Three Years of **Science**	Mean Effect Size of First Three Years of **Health Science**	Mean Effect Size of First Three Years of **4th Year**	Mean Effect Size of **First Three Years** of **All Inquiry**
Number of Studies	1	3	3	3	12
Appraisal of Arguments	**-0.30***	**0.40***	**0.49***	**0.29***	**0.42***
		(0.15 –0.96)	(0.19 –0.77)	(0.25 –0.36)	(0.15 –0.96)
Information Search Strategy	**0.55†**	**0.86‡**	**0.92‡**	**0.49***	**0.70†**
		(0.61 –1.03)	(0.32-1.13)	(0.12-0.78)	(0.12-1.13)
Self-Assessment	**-0.39***	**0.35***	**0.50†**	**0.43***	**0.39***
		(0.34 - 0.37)	(0.08-0.57)	(-0.30-1.17)	(-0.30-1.17)
Overall Critical Thinking Assessment	**-0.17**	**0.55†**	**0.61†**	**0.21***	**0.43***
		(0.36-0.88)	(0.22-0.67)	(-0.25-0.53)	(-0.25-0.88)

* small effect size, † medium effect size , ‡ large effect size

Table 6.5 Summaries of Mean Effect Sizes, Ranges, and Study Type for Cross-sectional and Longitudinal Approaches to Studying and Self-Confidence Measures for the First Three Years of Inquiry Students

Mean (Minimum and Maximum)		First Year **Non Inquiry Group** Effect Size	Mean Effect Size of First Three Years of **Science**	Mean Effect Size of First Three Years of **Health Science**	Mean Effect Size of First Three Years of **4th Year**	Mean Effect Size of **First Three Years** of **All Inquiry**
Meaning Orientation	**Mean**	**-0.03**	**0.18**	**0.27***	**0.59†**	**0.55†**
	Longitudinal (2 years)		(-0.09-0.22)	(0.17-0.34)	(0.19-0.21)	(-0.09-0.34)
	Cross-Sectional (3 years)		(0.23-1.16)	(0.47-1.30)	(0.33-1.69)	(0.23-1.69)
Reproducing Orientation	**Mean**	**0.26***	**-0.43***	**-0.87‡**	**-0.66†**	**-0.42***
	Longitudinal (2 years)		(-0.19 – -0.08)	(-0.33 – -0.09)	(-0.38 – -0.33)	(-0.38 – -0.08)
	Cross-Sectional (3 years)		(-0.6 – 0.13)	(1.10 – -0.22)	(-1.66 – -0.34)	(-1.66 – 0.13)
Organized Orientation	**Mean**	**-0.24***	**0.24***	**0.32***	**0.32***	**0.31***
	Longitudinal (2 years)		(0.33-0.42)	(-0.05 – 0.27)	(0.00 – 0.17)	(-0.05 – 0.42)
	Cross-Sectional (3 years)		(0.07-0.40)	(0.31-0.58)	(0.23-0.69)	(0.07-0.69)
Inquiry Task Self-Confidence	**Mean**	**0.45***	**1.16‡**	**1.05‡**	**0.70†**	**1.02‡**
	Longitudinal (2 years)		(1.23-1.35)	(1.22-1.23)	(0.57-0.78)	(0.57-1.35)
	Cross-Sectional (3 years)		(0.77-1.56)	(0.64-1.46)	(0.38-1.30)	(0.38-1.56)
Projected Academic Goal Self-Confidence	**Mean**	**0.08**	**0.74†**	**0.86‡**	**0.61†**	**0.76†**
	Longitudinal (2 years)		(0.48-0.68)	(0.33-0.42)	(0.11-0.13)	(0.11-0.68)
	Cross-Sectional (3 years)		(0.78-1.03)	(0.65-1.10)	(0.69-1.14)	(0.65-1.14)

* small effect size, † medium effect size , ‡ large effect size

Figure 6.1: Effect Sizes for the Non-Inquiry Students Longitudinal Data Only for All Measures

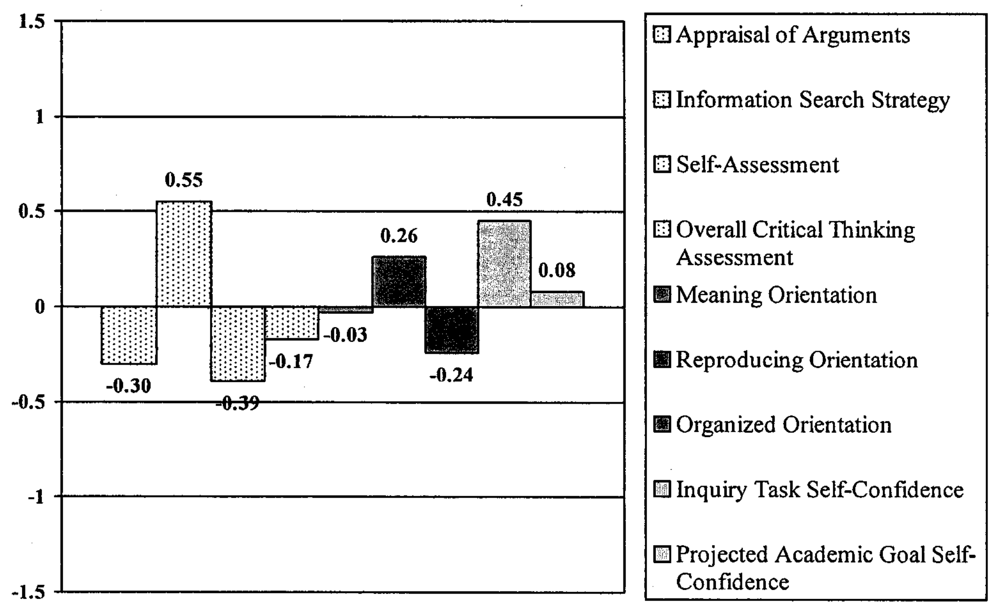


Figure 6.2: Mean Effect Sizes for First Year Science Inquiry Students for All Three Years of Outcome Measures

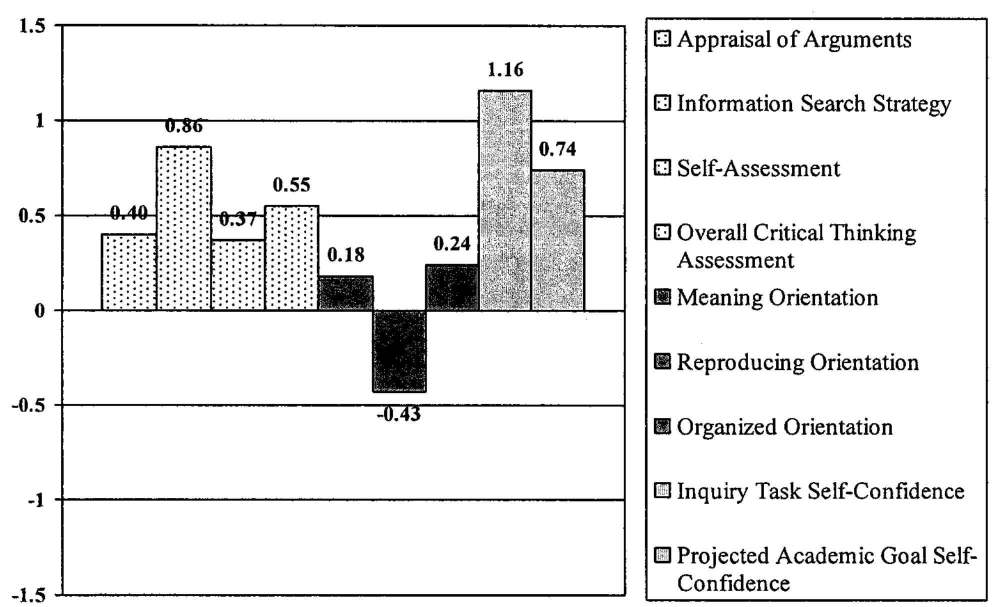


Figure 6.3: Mean Effect Sizes for First Year Health Science Inquiry Students for All Three Years of Outcome Measures

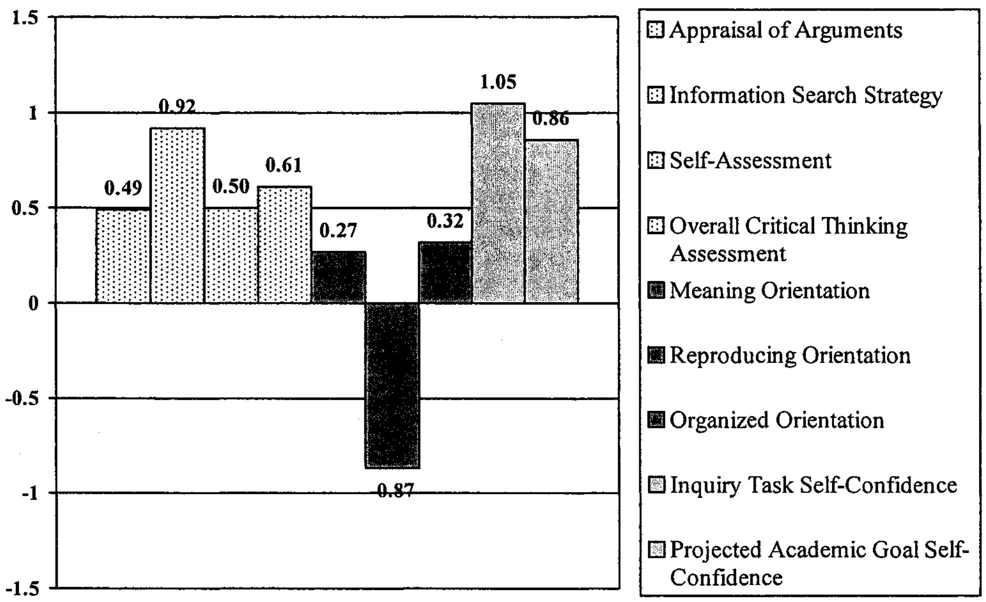
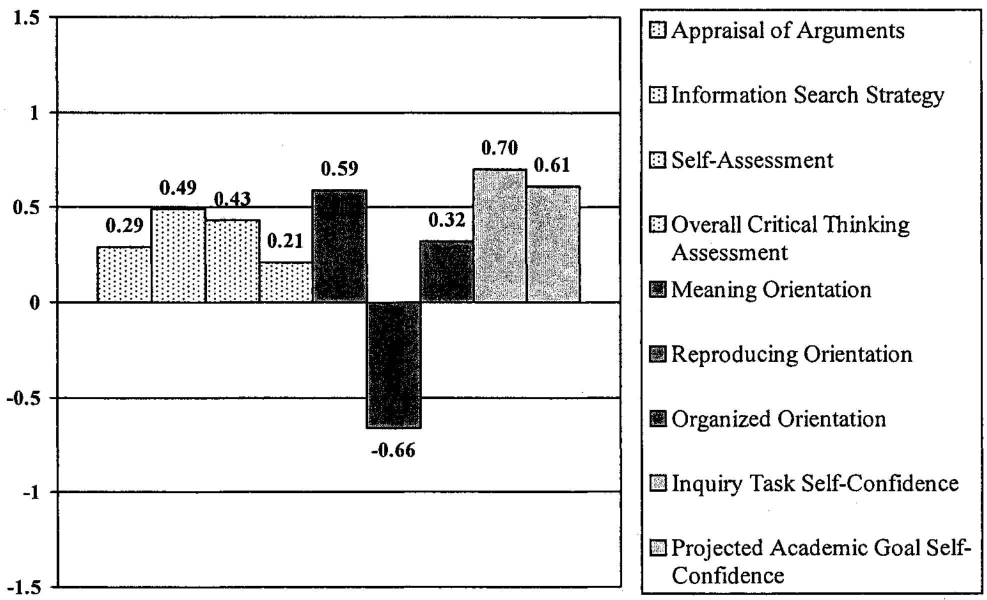


Figure 6.4: Mean Effect Sizes for Fourth Year Science Inquiry Students for All Three Years of Outcome Measures



When examining the average effect sizes for all inquiry interventions, there appeared to be the largest gains in information seeking behaviour from the critical thinking scale (0.49 to 0.92). It was not surprising to see the greatest change in these skill areas since they were the most concrete or straightforward for the students to show measurable progress. Also, skills such as learning how to use electronic journal index databases, evaluate information from many different sources, and how to find journal articles in a library, were most characteristic of the intervention. Other critical thinking sub-scores for the Appraisal of Arguments, Self-Assessment and Overall Critical Thinking Assessment, show small to medium overall mean effect sizes (0.37 to 0.61) for first year students and the 4th year students show small effect sizes (0.21 to 0.43). Fourth year students were more likely to show improvement in their self-assessment scores relative to first year students. Conversely, raters were more likely to rate first year students higher on their second mini-inquiry benchmarking exercise than 4th year students.

Critical Thinking skills were important to be introduced in the first year and followed up in future years. Library skills may be learned more readily while other skills show some improvement with the appraisal of arguments showing the smallest yet still significant gains. This result was not surprising given that learning how to critique ideas and research methodology of others was a complex skill that was highly dependent on knowing the context of the information presented. However, when these gains were compared to the effect sizes seen with the first year students who did not take Inquiry (effect sizes -0.30 for appraisal of arguments, -0.39 for self-assessment and –0.17 for overall critical thinking assessment), then the importance of these skills becomes obvious.

Most interesting, the largest and most consistent gains were seen in the Inquiry Skills Self-Confidence sub-scores. With an overall mean effect size ranging between a medium effect size of 0.61 and a large effect size of 1.16, it was clear that the inquiry intervention resulted in the science, health science and 4th year students having more confidence to approach academic tasks such as interviewing an expert, making presentations, or dealing with difficult group members. Relative to non-inquiry students (0.08), all inquiry student groups made large gains on their projected academic goal self-confidence scores (0.61 to 0.86). These gains may be due to the fact that the inquiry intervention focuses on having close contact with professors and peer tutors, working with their peers in groups on tasks of their choosing, and having large class debates and discussions. These tasks were more likely to have an impact on their students' academic "sense of self" because they were designed to persuade students to rethink their role as a university student and were less focussed on cognitive skill development. However, the longitudinal scores for projected academic goal orientation told a slightly different story: there were not statistically significant differences on these scores showing that 4th year students did not make gains. It may be that by fourth year, students had become more confident in their skills. Tasks, such as asking professors for work in their research labs or contacting a graduate school for information, were no longer "projected", but actual tasks these students needed to complete for their future goals of attending professional or graduate school or in finding future employment.

The Approaches to Studying Preference sub-scale scores did not show marked improvement in the average effect sizes for all inquiry interventions. Effect sizes were small for the meaning orientation preferences in all first year samples. Interestingly, the average effect size of the upper year students was highest for this group (effect size of 0.59) suggesting that the upper year students may have a greater need for delving deeper into content areas than first year students. The preference for reproducing orientation showed an overall mean negative effect size with the Inquiry interventions showing that students did not gain a preference for memorization from any of the inquiry interventions. It seems to encourage a preference for developing a meaning and organized orientations in first year students since their meaning and organized orientation scores were more likely to be statistically significant. Fourth year students appeared to have a greater dislike for a reproducing orientation but did not show a greater organized orientation preference.

The profile of the first year students who do not take inquiry as a skills development course in their first year shows that they struggled with adapting to the new expectations of being able to engage in academic research at university. Without the direct interaction with faculty members and peers on academic matters, these students appeared to lose ground on the areas of critical thinking and being organized in their work. While these students made gains on the more direct and proximate skills sets of Information Search Strategy and Inquiry Task Self-Confidence, the magnitude of these gains is about half of the average gains made by the first year Science and Health Science students (effect sizes of 0.55 for non-inquiry compared to 0.86 for first year science and 0.92 for health science on the Information Search Strategy sub-score; 0.45 compared to

1.16 and 1.05 for Inquiry Task Self-Confidence sub-scores). Most important, it seems that the university environment was an unwelcoming place of memorizing for tests where professors "delight in making the simple truth unnecessarily complicated" (Approaches to Learning Reproducing Orientation Scale item) given the positive scores in the Reproducing Orientation (effect size 0.26 for non-inquiry students) and the relatively low score in the Projected Academic Goal Self-Confidence (effect size of 0.08 for non-inquiry students). As their first year ends, these student may feel that they have not had opportunities to engage in learning. They may view university course work as solely a means to an obtaining a degree without engaging in the process of learning.

Results from the Follow-up Comparisons between the Science and Health Science Inquiry Interventions and the Non-Inquiry Control Group

For the demographic information of the non-inquiry, science and health science students in the follow-up study, see Tables 6.6 to 6.7.

Critical Thinking Scores Results

For the information of the non-inquiry, science and health science students from the ANCOVA analyses of the Critical Thinking scores from the follow-up study, see Table 6.9 and Figures 6.5 and 6.6 From the ANCOVA data from follow-up study on the mini-inquiry benchmarking data, the F values for the main effect of **cohort** for the overall model including rater, cohort, article, individual differences were as follows: *Appraisal of Arguments* scores was df (2, 220) F = 17.28, p = 0.0001; *Information Search Strategy* scores was df (2, 220) F = 39.86, p = 0.0001; *Self-Assessment* scores was df (2, 220) F = 21.08, p = 0.0001; and *Overall Critical Thinking Assessment* was df (2, 220) F = 30.30, p

= 0.0001. The overall model scores were a conservative estimate because the model included both Inquiry intervention sample comparisons. In the revised model of rater, cohort and individual differences, used to calculate the post-hoc significance (p values), determined whether the intervention to non-intervention comparisons for each treatment group were significant. The significant post hoc comparisons are in bolded print in Table 6.9.

Table 6.6: Upper Year Table Students Age, Gender and Language Spoken at Home

Upper Year Student at Follow-up			
	Age (Mean, Standard Deviation)	Gender (Percentage Females)	Language Spoken in Parental home not English (%)
Science Inquiry n = 61	22 (0.7)	78	27
BHSC n = 107	22 (0.8)	66	20
Non Inquiry n = 52	22 (1.4)	65	25

Table 6.7: Upper Year Table Students Work, Volunteer and School Commitments relative to Course Load

Upper Year Student at Follow-up			
	Percentage Worked less than 5 hours per week during school year	Percentage Volunteered less than 5 hours per week during school year	Percentage Taking a full course load
Science Inquiry n = 61	66	52	92
HSC n = 107	69	65	90
Non Inquiry n = 52	61	44	62

Table 6.8: Upper Year Student Responses at Follow-up to the Survey Question "Did skills developed in Inquiry helped you with other courses taken in first year, future years and your career?"

Upper Year Student at Follow-up			
Did skills developed in Inquiry helped you with your .....?			
	First Year Courses? (Percentage Yes)	Upper year Courses? (Percentage Yes)	Career? (Percentage Yes)
Science Inquiry n = 61	72	90	42
HSC n = 104	77	96	88
Non Inquiry n = 52	na	na	na

- Responses were ranked yes, not sure or no

Approaches to Studying and Science Inquiry Academic Self-Confidence Scales Results

For the information of the non-inquiry, science and health science students from the independent samples t-test analyses of the Approaches to Learning and Inquiry Self-Confidence scores from the follow-up study, see Table 6.9 and Figures 6.5 to 6.6.

Table 6.9: Follow-up Comparisons between Upper Year Students who did take Inquiry and Upper Year Science Students who did not take a First Year Inquiry Course on Critical Thinking, Approach to Studying and Inquiry Skills Self-Confidence

Critical Thinking Skills (Means, Mean Differences and Effect Size)					
		Inquiry Mean	Non-Inquiry Mean	Mean Difference	Effect Size
Appraisal of Arguments	Science	73.49	74.76	-1.27	-0.11
	HSc	**83.12**	**74.76**	**8.37**	**0.73†**
Information Search Strategy	Science	**73.65**	**63.20**	**10.45**	**0.78†**
	HSc	**83.73**	**63.20**	**20.53**	**1.52‡**
Self Assessment	Science	68.25	64.85	3.40	0.26*
	HSc	**77.99**	**64.85**	**13.15**	**1.02‡**
Overall Critical Thinking Assessment	Science	68.72	65.69	3.03	0.23*
	HSc	**80.78**	**65.69**	**15.09**	**1.13‡**
Student Approaches To Learning (Means, Mean Differences and Effect Size)					
		Inquiry Mean	Non-Inquiry Mean	Mean Difference	Effect Size
Meaning Orientation (Deep Learning)	Science	**83.28**	**91.27**	**-7.99**	**-1.26‡**
	HSc	**83.31**	**91.27**	**-7.96**	**-1.25‡**
Reproducing Orientation (Surface Learning)	Science	**70.59**	**84.97**	**-14.38**	**-1.79‡**
	HSc	**60.46**	**84.97**	**-24.51**	**-3.05‡**
Organized Orientation	Science	65.57	65.67	-0.10	-0.01
	HSc	68.18	65.67	2.51	0.33*
Student Confidence to Approach Inquiry Skills Development (Means, Mean Differences and Effect Size)					
		Inquiry Mean	Non-Inquiry Mean	Mean Difference	Effect Size
Inquiry Task Self-Confidence	Science	**83.27**	**78.68**	**4.59**	**0.58†**
	HSc	**85.86**	**78.68**	**7.18**	**0.90‡**
Projected Academic Goal Completion Self-Confidence	Science	**82.43**	**76.65**	**5.78**	**0.52†**
	HSc	**83.87**	**76.65**	**7.22**	**0.65†**

- All sub-scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores.
- Effect Sizes calculated by dividing the difference between the two scores (i.e., difference between HSc and Non Inquiry scores) and dividing these scores by the standard deviation of the Non Inquiry score.
- To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score. Therefore, the critical value for p is 0.01 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

Figure 6.5: Effect Sizes for Follow-up Science Inquiry Students when compared to Non-Inquiry Students for All Measures

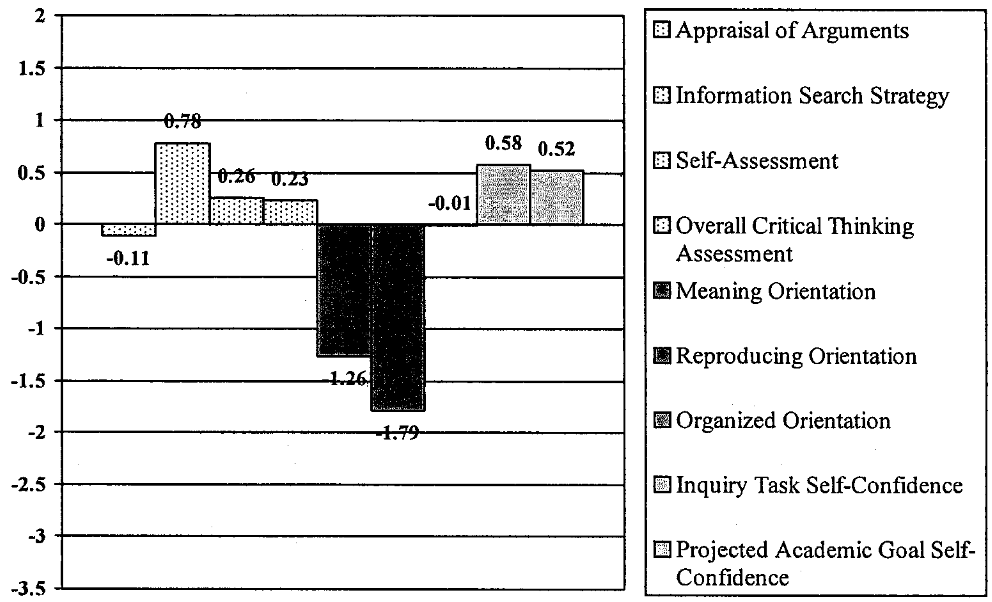
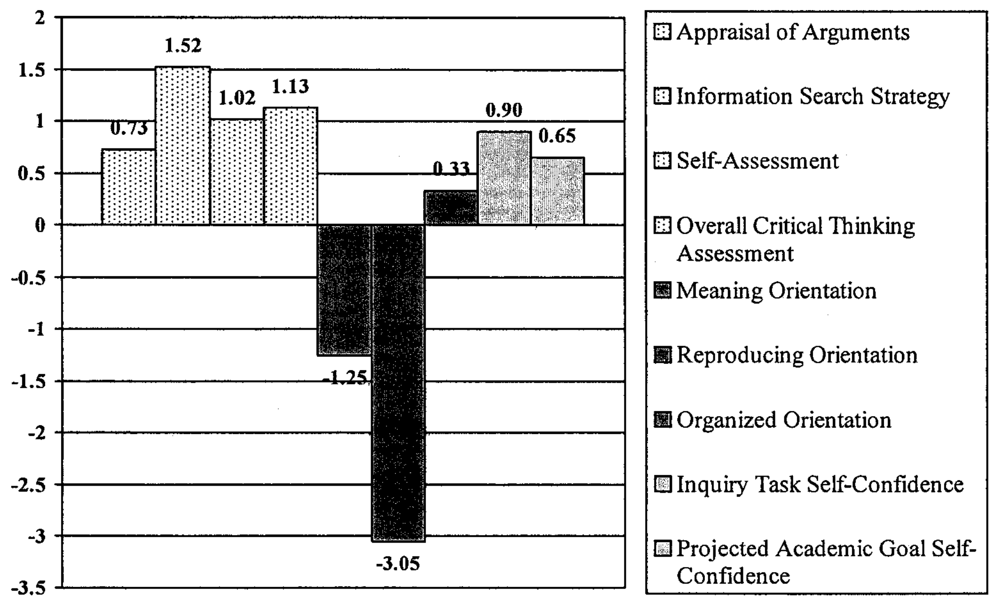


Figure 6.6: Effect Sizes for Follow-up Health Science Inquiry Students when compared to Non-Inquiry Students for All Measures



As was apparent from examining the medium to large effect sizes in all critical thinking scores shown on the preceding graphs for the follow-up studies, health science students had higher scores in all domains relative to graduating students who have not taken inquiry in their first year. In comparison, science inquiry students showed a medium effect size gain in the Information Search strategy (0.78) and small effect sizes for the two critical thinking sub-scores (0.26 for Self Assessment; and, 0.23 for Overall Critical Thinking Assessment); and a small negative effect size for the Appraisal of Arguments (-0.11). Since the health science students continued to have inquiry course work during all four years of their program, it appeared that the more inquiry exposure the students had, the more likely they were to retain their improvement in their critical thinking skills.

In the Inquiry Task Self-Confidence and Projected Academic Goal completion sub-scales, both the health science and science inquiry appeared to have made gains relative to graduating students who did not take inquiry in their first year of university. However, the Health Science students appeared to have higher scores relative to the science inquiry students and these differences may be also related to the health science having additional courses in their second to fourth years of their program.

The large negative effect size on the preference for meaning orientation was perplexing for the health science (-1.26) and science (-1.25) students relative to the extremely large negative effect sizes for the reproducing orientation scores (-3.05 for health science and -1.79 for science inquiry). While it is clear that the health science and science students have developed a meaning orientation to their inquiry-based studies based on their 83% scores on their meaning orientation preference scores for both groups, it may be that they did not attach the same meaning to competitive learning environments that the non-inquiry students experience. According to Knapper (2004), high scores on both meaning and reproducing orientation preferences were indicative of students who had a strategic learning orientation. A strategic learning orientation indicates that the students have learned to adapt their learning preferences to their environment and not developed their own internalized learning preferences. Students from the first year inquiry courses, with the opportunities to engage in self-directed learning with topics that were of interest to them, may not have developed this strategic orientation and appeared to have developed more stable, internalized learning preferences.

Alternative interpretations was that students with a high preference for reproducing learning were more likely to have difficulties with their studies and were less likely to complete their degrees (Trigwell, 2002; Kember, 1987; Kember et al., 1999; Byrne et al., 2004). However, since the follow-up students were in their final year of study, they were likely to complete their degrees. It may be that the non-inquiry value opportunities to engage in self-selected group learning activities because they have had fewer opportunities to engage in them and were more disenfranchised from their academic institution. In focus groups with science inquiry and health science students conducted after their degrees were completed, science students tended to describe gaining access to thesis or other self-directed activities as guerrilla warfare. Health Science students had difficulties empathizing with these statements because they felt that self-directed learning opportunities were easily accessible to them (Trim, unpublished data, 2004).

Similar, but smaller effect sizes were seen all scales when comparing the upper year science students who had only first year inquiry course compared to health science students who had their first year inquiry course and additional upper year inquiry courses. Upper year health science courses have a specific content focus, but students continued to work in small groups on a project they define, and there was ample opportunity to get feedback from their instructor given the small class sizes. Also, while health science students' organized sub-scale effect size increased from 0.31 to 0.33, science students decreased from 0.24 to 0.01. Science students did not appear to retain gains on the critical thinking skills, showing a decrease in their appraisal of arguments, self-assessment and overall critical thinking skills while the more concrete information search strategy was maintained. Similarly, science students did not retain their gains in the inquiry task self-confidence and projected academic goal self-confidence. It would appear that exposure to these skills in each year of their university curriculum has reinforced critical thinking skills for the health science students while the science students, without this reinforcement and opportunities to transfer, have not.

Synthesis of Results

By focusing on the strongest data, a noticeable pattern in the results developed. The following methods yielded the results in Table 6.13. Because of the similarity in t-test and ANCOVA values for the first year science and health science students, their results were counted together. The fourth year inquiry and follow-up students were treated as separate samples because of the dissimliarity in their t-test and ANCOVA values. Only longitudinal results for the Approaches to Studying and Inquiry Self-

Confidence scales were used for the first and 4th year samples because they would show the individuals' change over the course of the intervention. A simple count was taken of the number of significant longitudinal t-values and post-hoc ANCOVA values[1] that were subsequently converted to a percentage.

The pattern developed by examining Table 6.13 was that Inquiry Task Self-Confidence, Information Search Strategy, Overall Critical Thinking Assessment, and Self-Assessment Skills were the four areas that all students made progress and retained their gains from Inquiry. Projected Academic Goal Self-Confidence and Appraisal of Arguments were important gains for first year inquiry students that were retained by them in their graduating year.

However, it did not appear that 4th year Inquiry students made gains in their Projected Academic Goal Self-Confidence since it was not statistically significant in the longitudinal results although it was significant in 100% in the cross-sectional comparisons. It should be noted that the cross-sectional comparisons were with 1st year non-inquiry students. This was a useful result in creating a ceiling of what may be expected of Inquiry, but it would not tell us about the 4th year students' progress. Fourth year Inquiry students did not make gains in their Appraisal of Argument skill either. As has been previously argued, it may be that ascertaining the reliability and validity of arguments is a complex skill that requires several courses to make significant progress.

None of the students groups appeared to change their preference for a meaning orientation in the longitudinal comparisons or the follow-up cross-sectional comparisons. The reproducing orientation appeared to be a domain that was most relevant to 4th year and the follow-up Inquiry students in the Approaches to Studying Scale but it was not

[1] The significance level was relaxed to $p<0.05$.

relevant to the first year student. Conversely, the organized orientation was relevant to first year students but not to 4th year or follow-up students. These results may be interpreted as first year students were learning to cope with the study demands through the Inquiry course; but by their graduating year, all students had learned to be organized from their university course work *per se* and not specifically from Inquiry. Also, students in first year were expected to take courses that required many multiple choice test and text book learning. They may not have disengaged from a reproducing learning mode, since these skills were required of them. However, by their graduating year, Inquiry students intensely disliked a reproducing orientation relative to students who did not take Inquiry their first year. While no student samples showed consistent gains in their meaning orientation, cross-sectional comparisons showed that the first year students held onto their current meaning orientation while non-inquiry students showed a decrease in their meaning orientation after first year.

Table 6.10: Percentage of Times Statistical Test were Significant in Hypothesized Direction for the Longitudinal Results of All First Year and 4th year Inquiry Students and Cross-Sectional Results for the Follow-up Inquiry Students

Number of Times Statistical Test was Significant in Hypothesized Direction	Longitudinal Results All First Year Inquiry Interventions	Longitudinal Results 4th year Inquiry Interventions	Cross-Sectional Results Follow-up Inquiry Students
100%	Inquiry Task Self-Confidence	Inquiry Task Self-Confidence	Inquiry Task Self-Confidence
	Information Search Strategy		Information Search Strategy
	Overall Critical Thinking Assessment		Reproducing Orientation
			Projected Academic Goal Self-Confidence
75% or less	Projected Academic Goal Self-Confidence	Information Search Strategy	
	Self-Assessment	Self-Assessment	Self-Assessment
	Appraisal of Arguments	Overall Critical Thinking Assessment	Overall Critical Thinking Assessment
	Organized Orientation	Reproducing Orientation	Appraisal of Arguments
25% or less	Meaning Orientation	Appraisal of Arguments	
	Reproducing Orientation		
0%		Meaning Orientation	Meaning Orientation
		Organized Orientation	Organized Orientation
		Projected Academic Goal Self-Confidence	

References

Knapper, C. (2004) Personal Communication.

Appendix 6.1

Table: 6.11: Critical Thinking Skills: Appraisal of Arguments (Means, Mean Differences and Effect Size)

		Entry Mean	Exit Mean	Mean Difference	Effect Size
Study Year One	Science Inquiry n = 84	**58.33**	**71.81**	**13.47**	**0.96‡**
	HSC n = 97	**69.49**	**80.28**	**10.79**	**0.77†**
	4^{th} Year Inquiry n = 103	**73.64**	**77.09**	**3.45**	**0.25***
Study Year Two	Science Inquiry n = 77	46.11	49.89	3.78	0.15
	HSC n = 72	**61.95**	**67.05**	**5.10**	**0.19**
	4^{th} Year Inquiry n = 105	58.00	60.91	2.91	0.26*
Study Year Three	Non-Inquiry Group n = 31	55.24	50.99	-4.25	-0.30*
	Science Inquiry n = 136	**51.19**	**58.71**	**7.53**	**0.61†**
	HSC n = 100	66.60	70.26	3.66	0.27*
	4^{th} Year Inquiry n = 73	61.84	66.44	4.60	0.36*

- Mean Difference was based on difference between adjusted post-test and pre-test scores for each cohort.

-- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.

- Significant P-values are in **bolded print**

- * small effect size, † medium effect size, ‡ large effect size

Table: 6.12: Critical Thinking Skills: Information Search Strategy (Means, Mean Differences and Effect Size)

		Entry Mean	Exit Mean	Mean Difference	Effect Size
Study Year One	Science Inquiry n = 84	**57.88**	**74.98**	**17.10**	**1.03‡**
	HSC n = 97	**67.13**	**85.89**	**18.76**	**1.13‡**
	4th Year Inquiry n = 103	**71.93**	**84.84**	**12.91**	**0.78†**
Study Year Two	Science Inquiry n = 77	**39.21**	**56.10**	**16.89**	**0.61†**
	HSC n = 72	**70.15**	**79.74**	**9.59**	**0.32***
	4th Year Inquiry n = 105	**60.11**	**68.54**	**8.43**	**0.58†**
Study Year Three	Non-Inquiry Group n = 31	**47.09**	**56.21**	**9.11**	**0.55†**
	Science Inquiry n = 136	**59.76**	**74.39**	**14.63**	**1.01‡**
	HSC n = 100	**77.23**	**88.28**	**11.06**	**0.68†**
	4th Year Inquiry n = 73	70.17	71.98	1.80	0.12

- Mean Difference was based on difference between adjusted post-test and pre-test scores for each cohort.

-- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.

- Significant P-values are in **bolded print**

- * small effect size, † medium effect size, ‡ large effect size

Table: 6.13: Critical Thinking Skills: Self-Assessment of Critical Thinking Skills (Means, Mean Differences and Effect Size)

		Entry Mean	Exit Mean	Mean Difference	Effect Size
Study Year One	Science Inquiry n = 84	**75.29**	**79.31**	**4.01**	**0.37***
	HSC n = 97	**74.64**	**81.22**	**6.58**	**0.57†**
	4th Year Inquiry n = 103	**54.73**	**69.40**	**14.67**	**1.17‡**
Study Year Two	Science Inquiry n = 77	44.61	54.11	9.51	0.36*
	HSC n = 72	67.90	70.25	2.35	0.08
	4th Year Inquiry n = 105	**54.11**	**59.21**	**5.11**	**0.41***
Study Year Three	Non-Inquiry Group n = 31	58.87	52.84	-6.02	-0.39*
	Science Inquiry n = 136	**69.89**	**74.97**	**5.08**	**0.34***
	HSC n = 100	59.32	65.58	6.26	0.47*
	4th Year Inquiry n = 73	**65.43**	**61.13**	**-4.30**	**-0.30***

- Mean Difference was based on difference between adjusted post-test and pre-test scores for each cohort.
-- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**
- * small effect size, † medium effect size, ‡ large effect size

Table 6.14: Critical Thinking Skills: Raters' Overall Assessment of Critical Thinking Skills (Means, Mean Differences and Effect Size)

		Entry Mean	Exit Mean	Mean Difference	Effect Size
Study Year One	Science Inquiry n = 84	**51.19**	**66.19**	**15.00**	**0.88‡**
	HSC n = 97	**67.84**	**80.07**	**12.23**	**0.67‡**
	4th Year Inquiry n = 103	**70.95**	**76.6**	**5.65**	**0.33***
Study Year Two	Science Inquiry n = 77	**42.97**	**52.46**	**9.49**	**0.36***
	HSC n = 72	**68.1**	**74.29**	**6.19**	**0.22***
	4th Year Inquiry n = 105	**57.35**	**63.61**	**6.26**	**0.53‡**
Study Year Three	Non-Inquiry Group n = 31	53.01	50.6	-2.42	-0.17
	Science Inquiry n = 136	**53.38**	**62.47**	**9.09**	**0.71‡**
	HSC n = 100	**69.99**	**76.13**	**6.14**	**0.43***
	4th Year Inquiry n = 73	65.75	62.41	-3.34	-0.25*

- Mean Difference was based on difference between adjusted post-test and pre-test scores for each cohort.

-- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.

- Significant P-values are in **bolded print**

- * small effect size, † medium effect size, ‡ large effect size

Table 6.15: Comparison of Beginning to the End of the Semester Scores for Inquiry and Control Students on Meaning Orientation

2001 – 2002					2002 - 2003				
	Mean Score Beginning	Mean Score End	Mean Difference	Effect Size		Mean Score Beginning	Mean Score End	Mean Difference	Effect Size
Control Group = not collected					Control Group = 32	75.1	74.9	-0.3	-.03
Science Inquiry n = 64	77.5	79.2	1.6	.22*	Science Inquiry n = 37	78.5	79.9	1.4	-.09
HSC n = 77	**74.7**	**79.4**	**4.8**	**.34***	HSC n = 35	80.9	80.3	-0.7	.17
4^{th} Year Inquiry n = 71	77.8	79.2	1.4	.19	4^{th} Year Inquiry n = 73	76.7	78.3	1.7	.21*

- Mean Difference was based on difference between post-test and pre-test scores for each cohort.
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**
- *small effect size, † medium effect size, ‡ large effect size

Table 6.16: Comparison of Beginning to the End of the Semester Scores for Inquiry and Control Students on Reproducing Orientation

2001 – 2002					2002 - 2003				
	Mean Score Beginning	Mean Score End	Mean Difference	Effect Size		Mean Score Beginning	Mean Score End	Mean Difference	Effect Size
Control Group = not collected					Control Group = 32	66.1	68.3	2.2	.26*
Science Inquiry n = 64	69.6	68.8	-1.2	-.08	Science Inquiry n = 37	71.2	69.5	-1.6	-.19
HSC n = 77	**67.0**	**63.8**	**-2.9**	**-.33***	HSC n = 35	68.1	68.1	-0.9	-.09
4th Year Inquiry n = 71	**65.9**	**62.7**	**-2.8**	**-.38***	4th Year Inquiry n = 73	68.5	65.3	-3.3	-.33*

- Mean Difference was based on difference between post-test and pre-test scores for each cohort.
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**
- *small effect size, † medium effect size, ‡ large effect size

Table 6.17: Comparison of Beginning of the Semester to End the of Semester Scores for Inquiry and Control Students on Organized Orientation

2001 – 2002					2002 - 2003				
	Mean Score Beginning	Mean Score End	Mean Difference	Effect Size		Mean Score Beginning	Mean Score End	Mean Difference	Effect Size
Control Group = not collected					Control Group = 32	60.5	56.7	-3.8	-.24*
Science Inquiry n = 64	**60.6**	**65.8**	**5.2**	**.33***	Science Inquiry n = 37	**64.1**	**70.5**	**6.4**	**.42***
HSC n = 77	**64.0**	**68.8**	**4.8**	.27	HSC n = 35	68.9	68.1	-0.8	-.05
4th Year Inquiry n = 71	65.4	68.2	3.1	.17	4th Year Inquiry n = 73	65.9	66.0	0.1	.00

- Mean Difference was based on difference between post-test and pre-test scores for each cohort.
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**
- *small effect size, † medium effect size, ‡ large effect size

Table 6. 18: Comparison of Beginning of the Semester to End the of Semester Scores for Inquiry and Control Students on Inquiry Task Self-Confidence

2001 – 2002					2002 - 2003				
	Mean Score Beginning	Mean Score End	Mean Difference	Effect Size		Mean Score Beginning	Mean Score End	Mean Difference	Effect Size
Control Group = not collected					Control Group = 32	**70.2**	**74.7**	**4.5**	**.45***
Science Inquiry n = 64	**71.6**	**84.2**	**12.6**	**1.35‡**	Science Inquiry n = 37	**74.5**	**88.0**	**13.5**	**1.23‡**
HSC n = 77	**71.3**	**82.6**	**11.3**	**1.23‡**	HSC n = 35	**73.3**	**82.7**	**8.4**	**1.22‡**
4th Year Inquiry n = 71	**74.9**	**80.0**	**5.9**	**.57†**	4th Year Inquiry n = 73	**74.5**	**80.7**	**6.2**	**.78†**

- Mean Difference was based on difference between post-test and pre-test scores for each cohort.
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**.
- *small effect size, † medium effect size, ‡ large effect size

Table 6.19: Comparison of Beginning of the Semester to End the of Semester Scores for Inquiry and Control Students on Projected Academic Goal Self-Confidence

2001 – 2002					**2002 - 2003**				
	Mean Score Beginning	Mean Score End	Mean Difference	Effect Size		Mean Score Beginning	Mean Score End	Mean Difference	Effect Size
Control Group = not collected					Control Group = 32	**62.1**	**63.3**	**1.2**	**.08**
Science Inquiry n = 64	**72.3**	**80.0**	**7.7**	**.68†**	Science Inquiry n = 37	**75.7**	**82.6**	**6.9**	**.48***
HSC n = 77	**70.4**	**76.3**	**6.1**	**.42***	HSC n = 35	76.1	80.9	4.8	.33*
4th Year Inquiry n = 71	**76.0**	**77.3**	**1.3**	**.13**	4th Year Inquiry n = 73	76.5	78.0	1.5	.11

- Mean Difference was based on difference between post-test and pre-test scores for each cohort.
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
- Significant P-values are in **bolded print**
- *small effect size, † medium effect size, ‡ large effect size

Table 6.20: Comparison of End of Semester Scores for Inquiry Students to Control Students on Meaning Orientation

Study Year One				
	Mean Score Treatment	Mean Score Control	Mean Difference	Effect Size
	Non-Inquiry Group = 16			
Science Inquiry n = 80	**77.01**	**69.54**	**7.47**	**1.16‡**
HSC n = 66	**77.88**	**69.54**	**8.34**	**1.30‡**
4th Year Inquiry n = 103	**80.41**	**69.54**	**10.87**	**1.69‡**
Study Year Two				
	Non-Inquiry Group = 26			
Science Inquiry n = 87	**79.64**	**75.25**	**4.40**	**0.54†**
HSC n = 88	**79.97**	**75.25**	**4.72**	**0.58†**
4th Year Inquiry n = 81	79.66	75.25	4.41	0.54†
Study Year Three				
	Non-Inquiry Group = 47			
Science Inquiry n = 73	78.19	76.30	1.89	0.23*
HSC n = 114	80.11	76.30	3.81	0.47*
4th Year Inquiry n = 129	78.96	76.30	2.66	0.33*

- Mean Difference was based on difference between inquiry cohorts and control group scores
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
-To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score (i.e, 0.05/3). Therefore, the critical value for p is 0.017 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

Table 6.21: Comparison of End of Semester Scores for Inquiry Students to Control Students on Reproducing Orientation

Study Year One				
	Mean Score Treatment	Mean Score Control	Mean Difference	Effect Size
	Non-Inquiry Group = 16			
Science Inquiry n = 80	70.57	74.59	-4.02	-0.60†
HSC n = 66	**67.16**	**74.59**	**-7.43**	**-1.10‡**
4th Year Inquiry n = 103	**63.38**	**74.59**	**-11.20**	**-1.66‡**
Study Year Two				
	Non-Inquiry Group = 26			
Science Inquiry n = 87	69.92	68.61	1.31	0.13
HSC n = 88	64.23	68.61	-4.39	-0.44*
4th Year Inquiry n = 81	**62.47**	**68.61**	**-6.15**	**-0.62†**
Study Year Three				
	Non-Inquiry Group = 47			
Science Inquiry n = 73	68.85	69.53	-0.68	-0.07
HSC n = 114	67.47	69.53	-2.07	-0.22*
4th Year Inquiry n = 129	66.28	69.53	-3.26	-0.34*

- Mean Difference was based on difference between inquiry cohorts and control group scores
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
-To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score (i.e, 0.05/3). Therefore, the critical value for p is 0.017 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

Table 6.22: Comparison of End of Semester Scores for Inquiry Students to Control Students on Organization Orientation

Study Year One				
	Mean Score Treatment	Mean Score Control	Mean Difference	Effect Size
	Non-Inquiry Group = 16			
Science Inquiry n = 80	63.13	60.00	3.13	0.22*
HSC n = 66	65.83	60.00	5.83	0.42*
4th Year Inquiry n = 103	69.61	60.00	9.61	0.69†
Study Year Two				
	Non-Inquiry Group = 26			
Science Inquiry n = 87	65.35	64.23	1.12	0.07
HSC n = 88	69.15	64.23	4.92	0.31*
4th Year Inquiry n = 81	67.90	64.23	3.67	0.23*
Study Year Three				
	Non-Inquiry Group = 47			
Science Inquiry n = 73	65.04	58.51	6.53	0.40*
HSC n = 114	**67.81**	**58.51**	**9.30**	**0.58†**
4th Year Inquiry n = 129	**66.90**	**58.51**	**8.39**	**0.52†**

- Mean Difference was based on difference between inquiry cohorts and control group scores
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
-To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score (i.e, 0.05/3). Therefore, the critical value for p is 0.017 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

Table 6.23: Comparison of End of Semester Scores for Inquiry Students to Control Students on Inquiry Task Self-Confidence

	Mean Score Treatment	Mean Score Control	Mean Difference	Effect Size
Study Year One				
	Non-Inquiry Group = 16			
Science Inquiry n = 80	**83.10**	**66.95**	**16.15**	**1.56‡**
HSC n = 66	**82.06**	**66.95**	**15.11**	**1.46‡**
4th Year Inquiry n = 103	**80.36**	**66.95**	**13.41**	**1.30‡**
Study Year Two				
	Non-Inquiry Group = 26			
Science Inquiry n = 87	**84.08**	**76.03**	**8.05**	**0.77†**
HSC n = 88	**82.75**	**76.03**	**6.71**	**0.64†**
4th Year Inquiry n = 81	**80.00**	**76.03**	**3.97**	**0.38***
Study Year Three				
	Non-Inquiry Group = 47			
Science Inquiry n = 73	**85.85**	**75.18**	**10.67**	**0.98‡**
HSC n = 114	**82.41**	**75.18**	**7.23**	**0.66†**
4th Year Inquiry n = 129	**80.72**	**75.18**	**5.54**	**0.51†**

- Mean Difference was based on difference between inquiry cohorts and control group scores
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
-To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score (i.e, 0.05/2). Therefore, the critical value for p is 0.025 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

Table 6.24: Comparison of End of Semester Scores for Inquiry Students to Control Students on Projected Academic Goal Self-Confidence

Study Year One				
	Mean Score Treatment	Mean Score Control	Mean Difference	Effect Size
	Non-Inquiry Group = 16			
Science Inquiry n = 80	**75.63**	**63.00**	**12.63**	**0.96‡**
HSC n = 66	**75.06**	**63.00**	**12.06**	**0.92‡**
4th Year Inquiry n = 103	**77.94**	**63.00**	**14.94**	**1.14‡**
Study Year Two				
	Non-Inquiry Group = 26			
Science Inquiry n = 87	**79.13**	**65.69**	**13.43**	**0.78†**
HSC n = 88	**76.91**	**65.69**	**11.22**	**0.65†**
4th Year Inquiry n = 81	**77.68**	**65.69**	**11.99**	**0.69†**
Study Year Three				
	Non-Inquiry Group = 47			
Science Inquiry n = 73	**79.19**	**64.13**	**15.07**	**1.03‡**
HSC n = 114	**80.30**	**64.13**	**16.17**	**1.10‡**
4th Year Inquiry n = 129	**78.25**	**64.13**	**14.12**	**0.96‡**

- Mean Difference was based on difference between inquiry cohorts and control group scores
- Mean Differences based on scores totalling 100: Therefore, mean differences on scores were also considered to be percentage of difference between scores of the two groups under comparison.
-To adjust for multiple inferential tests, the p value was divided by the number of test per year for each sub-score (i.e, 0.05/2). Therefore, the critical value for p is 0.025 in this instance. Significant p -values are in **bolded print.**
- * small effect size, † medium effect size, ‡ large effect size

CHAPTER 7

Discussion and Conclusion

> "Perfect class to improve skills, better than memorizing text work. Group learning is realistic to the work force. Critical thinking is valuable to a successful school career." (First Year Health Science Inquiry Student)

> "Inquiry was good experience. Small group work was refreshing. Research performed helped me familiarize myself with the library. Discussion stimulated thought outside of class." (First Year Science Inquiry Student)

> "Learned a lot of applicable skills. Learned from me peers, bonded with the class in group work. Peer tutors tried hard to interact and made personal connections. Excellent course." (First Year Science Inquiry Student)

> "A great experience. Allowed me to interact with faculty and exposed me to different learning styles. Large group discussion was stimulating but small group projects were more productive. Learned a lot in terms of group dynamics, feedback and communication." (First Year Health Science Inquiry Student)

"Enjoyed the problem-based courses. The materials and skills learned in inquiry will benefit me forever. Happy that I took the course. Thanks." (Fourth Year Science Inquiry Student)

There were hundreds of comments from student course evaluations like the examples quoted above, offered spontaneously on the end-of-year course evaluation surveys: the ratio of positive to negative comments was about ten to one. It was clear that the students enjoyed taking Inquiry courses as much as did the students in the initial qualitative study (Chapter 3). They also valued the exposure to different and sometimes new ways of thinking and behaving which students felt changed their academic attitudes and skills. These self-reported student gains were not to be taken lightly since these alone would be sufficient evidence to justify continuing offering Inquiry courses.[2]

However, the fundamental research questions asked in the McMaster Science Inquiry program evaluation of the Science and Health Science Faculties were:

3. Did Inquiry foster the attributes of academic self-efficacy or confidence, critical thinking, and a preference to delve deeply into subject areas instead of focusing on memorizing solely the textbook materials for the test?
4. If it did foster these attributes by the end of the Inquiry course, were these gains in evidence when these students were ready to graduate?

Developing Self-Confidence as a University Student

It might be that the students' positive comments about the Inquiry course were indicative of something more fundamental. In his model of academic self-efficacy, Albert Bandura posits that students will be influenced to model their behaviours based on their experiences in their immediate environment. The influence of positive role models and peer relationships modelling realistic academic behaviours and expectations will lead to successful academic outcomes because they provide students with opportunities to internalize new behaviours and goals. Students develop the self-confidence to pursue these new behaviours because they have models to emulate and understand their reward. Equally, if the students are not exposed to any academic role models and peer

---

[2] See Theall, Abrami and Mets (eds) (2001) *The Students Ratings Debate: Are They Valid: How Can we Best Use Them? In the New Directions for Institutional Research series, vol (109)* for a summary of the reliability and validity of using student ratings for course evaluation purposes. In this volume, Theall et al. present data which argues that student evaluations of their learning and teachers are useful indicators despite their subjective nature.

relationships, or to negative academic role models and peer relationships, they will internalize those cognitive maps and begin to emulate and find rewards for these behaviours. The large gains in Inquiry Task Self-Confidence and Projected Inquiry Task Self-Confidence in the first year science and health science student may indicate that these students had opportunities to develop these new internalized maps leading them to engage in university level research and collegial behaviours with confidence.

A useful metaphor to think about the Inquiry courses could be that they provide students with an opportunity to rehearse appropriate academic behaviours in a safe environment. Ross and Nesbitt (1991) described the social psychology phenomena of *channel effects* as specific behavioural instruction on how to comply with suggested behaviour change can produce large effects while little or no behavioural instruction can produce changes in attitudes and knowledge but not in behaviours. "When trying to get people to change familiar ways of doing things, social pressures and constants exerted by the informal peer group represent the most potent restraining force that must be overcome and, at the same time, the most powerful inducing force that can be exploited to achieve success" (Ross and Nesbitt, p. 9). Inquiry, with its small class size and close contact with professors may be this "powerful inducing force" to help students learn at university.

Chua and Kustra (2005) interviewed graduating McMaster University health science students asking if and when they felt they had internalized self-directed learning practices. In the model she developed from her research findings, she described the first year inquiry course as *entry* year where the students were "faking it" but still were not able to trust that the university environment accepted that they could be self-directed in their learning. In their second or *transition* year, the students reported that taking a less demanding "bird" course was a defining moment for them because they realized that all courses were quite demanding and there was no point in taking a course that you were not truly interested in. Chua and Kustra described the upper year students as *expressing* and *solidifying* their self-directed learning motivation. In their third and graduating year students had to recruit their own course supervisor and set their own objectives, were now ready to fully trust that they were expected to implement the skills they learned in their first year course and engaged in self-directed learning in earnest.

The findings from the follow-up study show that the health science students appeared to make greater gains in all but two of the measured outcomes relative to the science inquiry students at graduation. This finding was consistent with the findings described by Chua and Kustra (2005). Whereas health science students continued to have required Inquiry courses in all years of their program, there were few vehicles for science students to continue to practice and transfer their inquiry skills in their second and third year of their program. As a result, this showed that Inquiry was a set of complex skills requiring many years of practice. It also showed that students benefited from repeated opportunities to practice and transfer these skills.

Becoming a Student Motivated to Learn

Not all students enjoyed the Inquiry course. Many of the negative survey comments found on the final course evaluation surveys were similar to the contradictory student statements from the qualitative study (Chapter 3). Both groups of students reported that they felt the course would benefit from the instructor providing "more structure" while strongly endorsing having the freedom to work in student groups on a topic of their choosing as a mechanism to learn Inquiry skills. Part of this contradiction may have been from the students' perceiving the instructor as "disorganized" while the instructors' view was that they were letting go of the reins so that the students could define their own learning objectives.

However, a more troubling, but less evident, explanation of students' not enjoying the Inquiry courses was described by Cuneo and Harnish (2002). They described a small group of students who struggled with self-directed learning environments as the "lost generation" because they experienced anxiety, did not feel safe from undue criticism, or meaningfully contribute when debate and critique of course materials was expected. Active learning environments, like Inquiry courses, were not safe places for students who were unsure of their critical thinking skills and knowledge; they likely do not promote self-confidence building among students who want to be told what was expected of them (Cuneo and Harnish, 2002). Students unsure of their knowledge will experience panic and anxiety over classroom discussions and group environments. Such students, who score high on reproducing orientation, generally tend to perform poorly academically (Cuneo and Harnish, 2002; Trigwell et al, 1999; Kember and Harper, 1987).

Good students adapt their learning style to the specific task at hand (Kember, 1987; Byrne et al. et al, 2004; Trigwell et al., 1999). Students' awareness of their learning environment was dependent on how their instructor set the course tone through various structures such as the course outline, assignments and evaluation procedures. Heavy workloads and a lack of opportunity to engage in topics of their choosing moved students to a surface/reproducing orientation to learning (Trigwell et al., 1999; Kember 1987). Typically, students with a preference for deep/meaning learning can shift to recalling factual information when required while students with a surface preference will recite remembered summaries from other experts on material but struggle to articulate their opinions (Kember, 1987). Additionally, students with a reproducing orientation were not likely to complete their course of study (Kember, 1987, p.9; Byrne et al. et al., 2004, p. 451) and their medical examination results were lower (McManus et al., 1998, p. 349).

A meaning orientation has been linked to higher quality learning outcomes. A deep approach was found to be associated with "perceptions of high quality teaching, some independence in choosing what was to be learned, and a clear awareness of the goals and standards required in the subject" (Trigwell et al., 1999, p. 58). Students who claimed to be experiencing "good teaching" consistently correlate their learning experience with a meaning orientation (Trigwell et al., 2002, p.66). Good teaching was defined as consistent with the "Seven Principles of Good Teaching" such as giving helpful feedback, being empathetic to students' difficulties, giving clear explanations, making subjects interesting, motivating and setting high expectations of students and showing an interest in what students have to say (Trigwell et al., 1999, p, 66). Good teachers focused on student needs, ensure that there was class time for self-directed learning, provoked debate, discussion and peer to peer interactions, and assessed students on their conceptual knowledge (Trigwell et al., 1999, p. 67).

Students with a strategic approach to learning had a patchy and variable understanding of the material and used whatever strategy required achieving highest marks possible in competition with other students (Byrne et al. et al., 2004, p. 450; McManus et al. 1998, p.346). Students who were unable to adapt to a meaning approach may become serial learners and were unable to integrate their learning into an integrated whole. The opposite problem was "globetrotting" students who were over-ready to draw conclusions without full consideration of the facts (Kember, 1987).

There have not been consistent results in the approaches to learning literature on the effect of deep and strategic learners particularly when considering cultural and learning discipline factors (Byrne et al., 2004). In the 1990 scale revisions, Byrne et al. (2004) noted that the deep/meaning approach remained the same while the strategic/achieving subscale now included items on the "alertness to assessment demands' and the surface/reproducing orientation remained the same except the name was changed to "surface apathetic" with more emphasis on ineffective studying . In their validation study for accounting students, Byrne et al., noted that there were cross loadings on the factor "*monitoring effectiveness*" for deep and strategic approaches. As Entwistle and McCune (2004) stated (cited in Byrne et al., 2004), these two approaches to learning were conceptually linked and cross loading was not a weakness of the scale. Kember et al. (1999) "found that the best fit was to a two-factor model with deep/meaning and surface/reproducing latent factors. The [strategic/]achieving subscale acted as indicators to both of these latent factors rather than separately to a third achieving approach factor as in Biggs' original model" (p. 2).

Students Becoming Competent Critical Thinkers

During this study, I have often heard faculty members describe the skills learned in Inquiry as "learning to ask good questions". However, I do not believe that the findings from this study support this characterization of Inquiry. We have good evidence that first year students learned to find information, make moderate gains in their ability to assess their critical thinking skills and improve in critical thinking skills overall but weaker evidence that they are able to evaluate arguments. When the skill set of the Inquiry courses is narrowly construed as the appraisal of arguments, the impact of the course on the students' later university skill development may be seriously hampered.

Critical thinking has been hailed by universities, education researchers and governments as vital to the "new economy". It was deemed essential for graduating university students to have "critical thinking skills" to sift through and make sense of the sheer volume of information available in the "information age" (Wolfe & Gertler, 2001). While there was wide spread agreement that critical thinking was an important skill in higher education, there was no consensus as to its meaning. The origin of the modern definition of critical thinking started with John Dewey. Dewey's original term was "reflective thought" which he defined as the "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds to support it and the further conclusions to which it tends" (p.6). His 1913 book, How We Think, outlined five steps in reflective thought: perceiving a felt difficulty; determining its location and definition; suggesting possible solutions; development by reasoning of the bearings of the suggestion; and further observation and experiment leading to its acceptance or rejection (p. 72). Dewey emphasized that "reasoning requires some experimental observation behind it" (p.72) and felt that the scientific method (i.e., inductive and deductive reasoning, hypothesis testing, etc.) was the essence of critical thinking (Hitchcock, 1992). According to Hitchcock (1992) "What many people now identify as critical thinking - the scrutiny of arguments and assertions produced by others – is at best a minor part of reflective thinking ...[and] is hardly mentioned in Dewey's book" (p. 1). Therefore, it becomes important to see appraisal of arguments as one of several skills integral to critical thinking.

Broader Implications of the Study

Encouraging life long learning requires changing the structure of teaching organizations as well as individual faculty members' teaching methods (Knapper and Cropley, 2000). Clearly we are not doing enough if students continue in programs that are largely lecture-based, weighed down by textbooks that used outdated models and facts. Inquiry courses were introduced at McMaster University to be "skill-driven rather than content-driven, on the skills required to perform effectively at university and well beyond university. These generalizable skills help students hone skills equally useful for advanced levels of academic research" (McMaster CLL Website). These courses were intended to develop critical thinking, library, research skills; evaluating and communicating academic information, reinforcing self-direction in academic pursuits, appreciating and becoming effective at teaching and learning from other students.

Halpern (1989) quoted B.F. Skinner on the value of education: "It has been said that an education [was] what survives when a man has forgotten all he has been taught. Certainly few students could pass their final examinations even a year or two after leaving school or college. What has been learned of permanent value must therefore not be the facts and principles covered by examinations but certain other kinds of behaviour often ascribed to special abilities" (p. 33). Halpern continued to say that she believes Skinner was referring to the need to teach specific skills in *how* to think instead of content oriented courses based on the thoughts of others. Woodhouse (2001) posits that education could be an “opportunity to recognize the intrinsic value of knowledge before they [were] fully exposed to the ravenings of the corporate market” (p.31). With this exposure to education, students would become capable citizens, able to question beliefs and habits taken for granted although tenuously supported by fact. “Yet in the modern global economy, information technology is used to increase both work and the stockholder value of business corporations which undermines the type of critical thinking that upholds democratic ideals of questioning and reason”(p. 34). Increasingly, governments and businesses are placing increasing pressure on academic institutions to produce graduates who will excel in the world of work (e.g., Conference Board of Canada, Ontario Ministry of Education, etc.). Allen et al. (1999) described the university environment as one where there was an “increasing need for university programs to provide ‘proof of service’ [which] creates a need for communication courses and activities to document the improvements generated” (p. 18). But “academic skills, decoupled from any disciplinary base, are really no different from “personal management skills,’ which enable a person “to get, keep and progress on a job and to get the best results” (Woodhouse, 2001, p. 33). The only difference was that personal management skills trained individuals to adopt the correct attitudes and behaviours, so as to “contribute to the organisation’s goals of maximizing corporate stockholder value” (p.33). But the prime purpose of teaching and scholarship in education was to “enlarge the public stock of knowledge” through research and to “contribute to the practice of open debate about the truth and falsehood of contesting ideas” (pp.33-34).

Creating active learning opportunities where students feel that they have the appropriate level of autonomy and support to learn what it means to study in a university or professional environments is essential. Attending a research-intensive university environment can turn into positive experience for undergraduates by exposing them to the skills involved with being a successful academic. These attributes and skills include: mentorship, collegiality and community service; the autonomy and integrity to pursue research questions from their inception to presenting the final results to the scrutiny of their peers; and being able to develop scholarly work that is ethical, moral and relevant to the larger community. All students should leave the convocation stage with their degrees convinced that the pursuit of intellectual ideas has merit and relevance to them and to their community.

References

Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. Communication Education, 48, pp. 18-30.

Byrne, M., Flood, B., & Willis, P. (2004). Validation of the approaches and study skills inventory for students (ASSIST) using accounting students in the USA and Ireland: a research note. Accounting Education, 13(4), pp. 449-459.

Chua, C. L. M., & Kustra, E. (2005) A study of academic motivation in graduating BHSC students. Undergraduate Thesis. Hamilton, ON: Bachelor of Health Science, Faculty of Health Sciences, McMaster University.

Cuneo, C. J., & Harnish, D. G. (2002). The lost generation in e-learning: Deep and surface approaches to online learning. 83rd Annual Meeting of the American Education Research Association Division C: Investigations of online learning environments. New Orleans: April 2nd, 2002.

Dewey, J. (1991). How we think. Buffalo, NY: Prometheus Books.

Halpern, D. F. (1989). Thought and knowledge: An introduction to critical thinking (2nd ed.).

Hitchcock, D. The concept of critical thinking. 1992. Unpublished Work

Kember, D., & Harper, G. (1987). Approaches to studying research and its implications for the quality of learning from distance education. Journal of Distance Education/Revue de l'enseignement à distance. Date of Access: 11/13/2005. URL: http://cade.icasp.org/vol2.2/8_Kember_and_Harper.html.

Kember, D., Wong, A., & Leung D. Y. P. (1999). Reconsidering the dimensions of approaches to learning. British Journal of Educational Psychology, 69(3), pp. 323-343.

Knapper C., & Cropley A. (2000). Lifelong learning in higher education (3rd ed.). Sterling, VA: Kogan Press.

McManus, I. C., Richards, P., Winder, B. C., & Sproston, K. A. (1998). Clinical experience, performance in final examinations, and learning style in medical students: prospective study. BMJ, 316, pp. 345-350.

Ross L., & Nisbett, R. (1991). The person and the situation: Perspective of social psychology. New York: McGraw-Hill.

Theall, Abrami & Mets (eds) (2001) The Students Ratings Debate: Are They Valid: How Can we Best Use Them? New Directions for Institutional Research Series, vol (109).

Trigwell, K., Prosser, M., & Waterhouse F. (1999). Relations between teachers' approaches to teaching and students' approaches to learning. Higher Education, 37, pp. 57-70.

Wolfe, D., & Gertler, M. (2001). The new economy: An overview. Ottawa, ON: Social Sciences and Humanities Research Council of Canada.

Woodhouse, H. (2001). In praise of idleness: Bertrand Russell's critical thinking about the global market. Inquiry; Critical Thinking Across the Disciplines, 2001 Win; Vol 20, pp. 27-35.