Implementing Engagement Improvements through Targeted Interventions: Final Report: Intervention Processes, Impacts and Implications

Prepared by Chris Conway, Queen's University for the
Higher Education Quality Council of Ontario

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

Disclaimer:

The opinions expressed in this research document are those of the authors and do not necessarily represent the views or official polices of the Higher Education Quality Council of Ontario or other agencies or organizations that may have provided support, financial or otherwise, for this project.

Cite this publication in the following format:

Conway, C. (2010). *Implementing Engagement Improvements through Targeted Interventions: Final Report: Intervention Processes, Impacts and Implications.* Toronto: Higher Education Quality Council of Ontario.

Published by:

The Higher Education Quality Council of Ontario

1 Yonge Street, Suite 2402
Toronto, ON Canada
M5E 1E5
Phone: (416) 212-3893
Fax: (416) 212-3899
Web: www.heqco.ca
E-mail: info@heqco.ca

Funding for this research was provided by the Higher Education Quality Council of Ontario. The opinions, findings, conclusions and recommendations expressed are those of the author and do not necessarily reflect the views of the Higher Education Quality Council of Ontario.

Acknowledgements

This report reflects the commitment and creativity of dozens of individuals at numerous Ontario universities and several other organizations:

Dr. James Downey, past President, and Dr. Ken Norrie, Vice-President (Research) of the Higher Education Quality Council of Ontario, provided encouragement and support to the initial project concept and provided project funding by the Council;

Teams of faculty members, instructional designers, service providers and institutional researchers at each of the participating universities designed, implemented and reported on their engagement interventions;

Dr. Jillian Kinzie, Associate Director of the NSSE Institute for Effective Educational Practice, and Mr. Todd Chamberlain, Associate Director of NSSE Survey Operations, contributed to the project design and supported the NSSE- and CLASSE-related aspects of project implementation;

Dr. Nathan Marti of Abacist Analytics in Austin, Texas provided expert statistical and assessment design support;

Ms. Kim Akerblom and Mr. Brian Lewis of the Office of Institutional Research and Planning at Queen's University provided extensive administrative and analytical support respectively throughout the project;

Thousands of undergraduate students at the participating universities completed NSSE, CLASSE and other surveys and participated in numerous focus groups.

Any errors or omissions in this report remain the sole responsibility of the author.

Explanatory Note Regarding Project Reports

This is the final report of the Implementing Engagement Improvements Through Targeted Interventions project. Reports were produced at the end of Phase One, and at two points earlier in Phase Two. This report is intended to stand alone: although it does not cover all the materials in the earlier reports in detail, it provides web references for the critical contents of the earlier reports. The contents of the previous reports are as follows:

Phase One Report and Phase Two Proposal (October 29, 2007)

- Overview of NSSE
- Rationale, Origins and Goals of the Project
- Current NSSE Implementation Practices and Assessment Issues
- Description of the Process for Soliciting and Approving Site Projects (proposal submission form, orientation workshop materials, proposal assessment criteria)
- List and Descriptions of 13 Proposed Site Projects
- Phase Two Budget Proposal, Schedule and Deliverables

Phase Two Progress Report #1 (March 31, 2008) and #2 (January 5, 2009)

- Reports on Proposal Design and Development
- Reports on Mid-Point Intervention Preparation
- Update on Project Assessment Designs and Data Sources
- Discussion of CLASSE Survey Instrument
- Pre-Intervention Data and Information Submissions from Project Sites
- Arrangements for Targeted (Post-Test) Administration of NSSE 2009
- Administrative and Implementation Issues to December 2008

Executive Summary

The National Survey of Student Engagement (NSSE) has become a widely used tool for exploring institutional practices and student behaviours that are known to be associated with good learning outcomes. The survey has focused attention on comparative engagement performance and the factors underlying engagement differences among institutions, and it has helped university faculty and service personnel identify opportunities for engagement improvement. All Ontario universities administer NSSE on a regular basis as part of their accountability agreements with the Provincial Government, and their benchmark scores are publicly available. However, moving from process-based to outcomes-based engagement accountability (i.e., demonstrating engagement improvement over time) implies that institutions will be successful in implementing effective improvement strategies, and that NSSE will be able to detect their impacts.

The goal of this project was the design, implementation, NSSE-based assessment and documentation of engagement-related interventions at several Ontario universities in order to strengthen the foundation for engagement implementation and assessment practice. More specifically, the project's four objectives were:

- To establish an inventory of effective intervention "field practices" including those related to data collection, survey administration, intervention design, assessment design and analysis methodology;
- To share (among project participants and more widely) intervention practices and experiences to support improved implementation and assessment efforts in the longer term;
- To conduct formal statistical analyses using best available data and assessment methods to measure the effects of the interventions on both NSSE scores and other key experience and outcome indicators;
- To inform policy discussions related to the appropriate accountability applications of NSSE.

This project was undertaken in two phases. The first phase involved development of an overall project plan, orientation of potential university participants to the project and its participation requirements, the development of intervention and assessment design proposals by interested universities, proposal approval by a project steering committee, and overall project costing. The second phase (which was funded and that commenced after overall project feasibility and budget had been established) involved refinement of the approved intervention and assessment designs, planning and preparation activity at each of the university sites, intervention implementation (including administration of NSSE and other surveys and the collection and submission of other data) generally throughout the 2008/09 academic year, the central analysis and assessment of each intervention and the project overall, and the production of this final report.

The participating universities and the foci of their interventions were:

- Carleton University: Training, mentoring and support to teaching assistants in selected courses across five Faculties in order to improve scores on selected NSSE items;

- University of Guelph: Student peer-led supported learning groups in selected first-year "high-risk" courses to enhance learning skills and improve the first-year experience;
- University of Ottawa: A Faculty-wide learning community implemented through common timetabling, and enrichment/support sessions and other social and academic opportunities to assist first-year student integration into a large Faculty;
- Queen's University (Psychology Department): Students in a very large introductory course attended small group enrichment sessions dealing with research and professional practice to compensate for otherwise limited student-faculty interaction opportunities;
- University of Western Ontario: Integration of science literacy skills development into a first-year Biology course through seminar redesign and on-line support in response to the Faculty's academic plan commitment to improved teaching and learning;
- Ryerson University: Incorporation of a core writing skills component into a required first-year course in each of eight professional programs, delivered via specially designed tutorial sessions, in order to address identified writing skills deficiencies;
- Wilfrid Laurier University: A senior student peer-led program to improve information literacy, research skills and writing skills delivered through two first-year writing-intensive courses, to facilitate the transition to university research and writing;
- University of Waterloo: Courses were redesigned by faculty members following their participation in Waterloo's Teaching Excellence Academy to improve focus on learning objectives and assessment methods and to reflect the University's academic plan and revised degree expectations;
- University of Windsor: First-year students in the School of Business received a substantially enhanced advising program via regular contacts with faculty members and student mentors to deal with academic plans and progress, Q+A's, student activities, and course/program issues, to improve social and academic integration;
- Queen's University (Electrical Engineering Department): Fourth-year students accessed on-line real-time extended-hours tutorial support (in addition to conventional classroom support) integrated across three Fall term and three Winter term courses to address concerns over academic support and curricular integration.

The intervention projects were developed to address a number of design and assessment difficulties that often confront the evaluation of university initiatives. These included the construction of control and experimental groups, targeting to prevent diluted measurement of intervention impacts, propensity matching of control and experimental groups to overcome self-selection bias, intensity of involvement measures to permit more precise measurement of intervention impacts, assessment designs to control for background noise, and the collection of additional supporting data (e.g., grades, retention status) at the student record-level to permit more detailed analysis than that possible with NSSE alone.

The project achieved its objectives with a reasonable level of success. The intervention development, implementation and assessment process at each of the university sites has

been documented, hopefully in a manner that will be helpful to those contemplating their own engagement interventions, and that will encourage constructive advice on design and implementation practices from others. Several effective (and some less effective) data, design and assessment approaches were identified. Problems encountered, and where uncovered, ways of avoiding or addressing them, have been described. The sensitivity of NSSE and other measurement tools to intervention impacts was assessed and documented with the expectation that other practitioners will seek to improve on the methodology. The key finding in this respect is that NSSE item and benchmark measures are generally unable to detect the effects of the relatively modest (and undoubtedly imperfect) interventions involved, but that the course-based version of the survey (CLASSE) and other measurement tools showed significant promise in single- and multiple-course interventions. This finding is not a criticism of NSSE (which has demonstrated value in numerous applications) or necessarily of the interventions themselves; it addresses the "fit" between NSSE and the scale and scope of these particular interventions. This finding argues against rapid movement toward outcomes-based engagement accountability generally, and with respect to the ongoing development of Ontario's multi-year accountability agreements. A number of potentially useful research and practice suggestions are provided to improve ongoing implementation activity.

Table of Contents

List of Figures and Tables

1. Project Background

1.1 Student Engagement as an Emerging Quality Indicator

The quality of postsecondary education, evidence of the extent to which quality exists, the consequences of varying levels of quality and the prerequisites for achieving it have been at the centre of university planning and management for decades. While there is widespread agreement that a "quality education" is one that maximizes student learning and growth, consensus is lacking on the best ways to define, measure and compare educational quality.

This absence of consensus has provided numerous organizations with opportunities to develop a variety of different quality assessment methods. Maclean's Magazine in Canada and US News and World Report in the U.S. both produce university rankings based primarily on such input measures as student entering grades, faculty-student ratios or class sizes, faculty research grants, operating budgets, library resources and institutional reputation. The Globe and Mail's *Canadian University Report* (formerly the University Report Card) avoids explicit rankings but permits side-by-side university "grades" comparisons based on student responses to a satisfaction survey, library expenditures, athletic scholarships and programs, tuition and accommodation costs, and student entering grades. The Times Higher Education Supplement in the UK publishes *World University Rankings* based on peer and employer reviews, student-faculty ratios, citations of faculty research, and the proportion of international faculty and students. Shanghai Jiao Tong University produces an *Academic Ranking of World Universities* based on the proportion of alumni and faculty winning Nobel prizes, citations of faculty research, and faculty publication volume. In Canada, Research Infosource publishes *Canada's University Innovation Leaders* based on a standardized measure of total sponsored research income. In addition, numerous media organizations develop rankings of specific schools and Faculties (most notably Schools of Business). In Ontario, a small portion of each university's funding is driven by its (implicit quality) performance in terms of the employment rate of its baccalaureate graduates and the graduation rate of its undergraduate students.

For their part, universities have traditionally argued the existence of quality by using a number of input, process and outcome measures. Student-faculty ratios, class sizes, research productivity, student satisfaction and graduate outcomes (employment and advanced education) have all played a role in external reporting (though the limitations of such measures are often acknowledged and the context and level of aggregation in which they are presented often differ from media rankings exercises). Universities generally prefer to utilize institutional processes as evidence of quality, including academic program reviews and professional accreditations, Senate and Board scrutiny of program issues, and jurisdiction-wide quality appraisals (e.g., accreditation reviews or, in Ontario, formal sector-wide undergraduate and graduate program approval and review processes).

It is surprising, then, that over the past several years, discussions of postsecondary quality have increasingly come to focus on the concept of student engagement, and that growing

agreement exists within universities and governments that student engagement is perhaps the central (though certainly not the only) tool with which to frame discussions of university quality. There appear to be several reasons for this shifting focus. Instructional and classroom practice, curriculum and course delivery, service provision, technology applications and the linkages between home, community and campus are clearly related to students' academic and social experiences and therefore to student learning, but are noticeably absent in the ranking and quality assessment methods described above. Second, the primary focus of media attention on quality *input* measures has been shown to be approximate at best and misleading and inaccurate at worst. Third, universities are different from one another – some would say unique – and the standardization of quality measures across institutions with varying missions, sizes, locales and student bodies appears to many to be an oversimplification. Fourth, universities generally find that conventional input measures provide little if any direction to their quality improvement efforts (other than implying that more resources and an improved reputation would somehow solve the problem).

But the primary explanation for growing interest in student engagement as a quality measure has been the development and widespread administration of the National Survey of Student Engagement (NSSE). A significant body of literature indicates that student engagement is highly correlated with positive learning outcomes: when students are active participants in their education, when they interact with faculty and other students, when they participate in supplemental learning experiences like community service learning and international study, and when they experience the supportiveness of their university to both their academic and social needs, the research demonstrates that learning is enhanced (i.e., that knowledge acquisition, skills development and personal growth occur). As indicated above, if a quality learning environment is one that maximizes student learning and growth, then the level of, and improvements to student engagement within the learning environment must be considered key indicators of quality.

1.2 An Overview of the National Survey of Student Engagement

NSSE measures institutional practices and student behaviours across numerous dimensions of the student experience that are known to be associated with positive learning outcomes. It is normally administered to students in their first and final year of study in first-entry undergraduate programs. The instrument contains about 100 questions dealing with student engagement experiences and exposure to institutional engagement practices, perceptions toward the academic and social environment on campus and off, and student program and demographic characteristics. Forty-two of the engagement-related questions deal with, and can be assembled into five benchmarks of educational practice:

- Level of Academic Challenge: institutional emphasis on/time spent preparing for class; amount of reading and writing performed; course emphasis on analysis, synthesis/organization, making judgments, applying theories; working to meet expectations;

- Active and Collaborative Learning: asking questions/participating in in-class discussions/making presentations; working with classmates out of class; tutoring other students; community-based project involvement; discussions with others out of class;
- Student-Faculty Interaction: discussing grades/assignments/ideas/career with faculty; non-coursework interactions with faculty (committees, research projects, etc.); receiving prompt feedback on performance;
- Enriching Educational Experiences: co-curricular involvement; practicum/internship/co-op involvement; community service/volunteer work; foreign language coursework; study abroad; independent study;
- Supportive Campus Environment: campus environment support to succeed academically/cope with non-academic responsibilities/thrive socially; quality of relationships with students/staff/faculty.

Some of these items and others on the instrument deal with deep learning (higher-order, integrative and reflective learning), personal and social gains, general educational gains, practical competence gains, satisfaction, and time usage/demands. Additional questions ascertain respondent age, gender, domestic/international status, ethno-cultural status, prior postsecondary history, full-/part-time enrolment, academic performance to date, program major, parental educational attainment and year of study.

The instrument was developed in 1998 and piloted in 1999 at the University of Indiana Bloomington under the direction of Dr. George Kuh with initial funding from the Pew Charitable Trusts. Since 2000, it has been administered at least once at over 1,300 universities in the US and Canada. Extensive statistical testing has shown the NSSE instrument to be statistically valid and reliable. (Full documentation of the history, theoretical foundations and psychometrics of NSSE, and copies of the instruments are available on the NSSE web site.) Several similarly-intentioned instruments have been developed to augment NSSE for Faculties of Law (LSSSE), community colleges (CCSSE), secondary school students (HSSSE), university faculty members (FSSE) and pre-entry university applicants (BCSSE, or BUSSE in Canada). Independent development and piloting has been undertaken on two course-specific versions of the instrument (CLASSE-Student and CLASSE-Faculty).

Canadian participation in NSSE began in 2004 with 11 institutions administering a slightly modified Canadian version of the survey. A French language version was developed and first administered in 2005. As of the 2008 administration, 61 Canadian universities (including satellites and affiliates) have administered the survey at least once; all Ontario institutions have administered NSSE at least twice as a component of an accountability framework developed with the Ministry of Training, Colleges and Universities. A handful of Canadian universities have also administered BCSSE and/or FSSE. Though administered in both paper and web-based formats in the US, only the latter version of NSSE is available in Canada.

Canadian NSSE participants submit a file to NSSE containing email contact information and selected program characteristics for first- and final-year student populations, from which NSSE selects a random sample (and in some cases targeted over-sample selected by the participating university). Students receive an email invitation and subsequent reminder customized by their university, and three generic follow-up emails. On

completion of the survey, participants receive the record-level response file and reports on the response means, frequencies and benchmark scores (including standard statistical information to facilitate comparisons) for their own university, their consortium (if applicable) and several institution-defined groups of comparator institutions. A wide range of information dealing with the interpretation, dissemination and application of survey results is also provided through both the NSSE web site and the publication of an annual report summarizing that year's administration and results.

1.3 Engagement Implementation Practice

The ultimate rationale for administering NSSE is, of course, to use the results to direct improvements in quality. However, a review of US implementation practice undertaken in late 2007 (and contained in the Phase One report) suggests that NSSE implementation has often not progressed to explicit or formal institutional improvement activity. US institutions appear to fall into one of the following implementation stages:

- Phase 1: Analysis of survey results, performing data drilldowns, conducting peer comparisons, data sharing, engagement issue/problem identification;
- Phase 2: Dissemination of results to internal and external audiences, introducing vocabulary change, effecting limited organizational learning, committing to further activity, conducting/communicating additional research;
- Phase 3: Integration of NSSE into institutional processes such as budgeting, plan benchmarking and monitoring, developing KPI's, accreditation reports, the formation of NSSE-related committees and task forces, supporting academic program reviews;
- Phase 4: Informal implementation of service, program and curricular responses substantially based on NSSE findings but without formal assessment;
- Phase 5: Formal implementation of service, program and curricular responses substantially based on NSSE findings and including formal assessment;
- Phase 6: Continuous improvement through repeated implementation-assessment-retesting, widespread incorporation of best practices, substantial cultural/organizational change.

Considerable summary documentation exists demonstrating the (often selective) use of *single point in time* NSSE item and benchmark scores in accreditation reviews, marketing and alumni/applicant relations; and the use of *time-series* NSSE results to *informally* guide the design of student services, instructional development workshops and enrolment management. These kinds of activities generally fall into Phases 1 to 4. NSSE has assembled numerous examples of this implementation activity in each of its annual reports and on its web site. Almost all institutions have progressed at least through Phases 1 and 2, with a decreasing number reaching Phases 3 and 4. While an institution's position along the continuum is certainly affected by the length of time it has been involved with NSSE, it nonetheless appears that many institutions are "stalled" at dissemination/ integration (Phases 2 and 3).

Published examples of the applications of NSSE results in examining institutional quality over time and evaluating the role and impact of explicit institutional efforts to achieve improvement (Phases 5 and 6) are scarce: very little implementation activity in the US and Canada could be found to occur at this level. Where "sustained" implementation activity has occurred, it has generally been without formal evaluation. There exists little *detailed* documentation of implementation experience and therefore limited opportunity for information sharing and adoption of promising practice, a situation that may in fact be inhibiting formal evaluation efforts and progress toward continuous improvement.

In 2009, near the end stage of this project, NSSE published two reports dealing with NSSE implementation practice. *Using NSSE to Assess and Improve Undergraduate Education* provides case studies of over a dozen universities and their NSSE implementation focus on the first year experience, communication of results, integration of NSSE and institutional data, institutional assessment and student affairs. Second, NSSE's 2009 Annual Report, *Assessment for Improvement: Tracking Student Engagement Over Time*, provides a 10-year window on selected trends in NSSE results over time, and discusses the implications of these trends. While both reports suggest progress in NSSE implementation practice over the past several years, the approach taken in this project appears to remain unique.

1.4 The Ontario Context: Postsecondary Funding and Accountability Procedures

Ontario's public universities have long advocated for the funding increases they consider necessary to achieve quality improvements. Except for a few isolated and relatively minor increases in per-student operating grants, grant funding increases in Ontario in the two decades prior to 2005/06 had been based only on enrolment growth, with the result that per-student inflation-adjusted operating grants had been in steady decline. Only a portion of this decline had been offset by increases in tuition and fees, most often to the maximum level permitted by a series of government regulatory regimes. The Ontario Government commissioned a review of the provincial postsecondary sector led by former Provincial Premier Bob Rae that resulted in the publication in 2005 of *Ontario: A Leader in Learning – Report & Recommendations*. The review made numerous recommendations concerning coordination among postsecondary institutions (particularly colleges and universities), the need for operating grant increases, overall accessibility and access for historically under-represented groups, changes to government student financial assistance programs, and institutional and provincial roles and responsibilities within a new accountability framework. Many of the Rae Review recommendations were at least partially addressed in the 2005 Province of Ontario Budget, and the multi-year *Reaching Higher* plan that the budget introduced (including the creation of the arm's-length Higher Education Quality Council of Ontario (HEQCO) with a mandate to conduct independent research, and advise the government on postsecondary issues; HEQCO is this project's funding agency).

The primary component of the new accountability framework was the introduction of bilateral Multi-Year Accountability Agreements (MYAA's) between the Province and each of Ontario's universities and colleges. The first round of agreements covered the 2005/06 – 2008/09 period, and included the initial specification of (and annual report backs on)

university strategies, targets and performance on the participation of under-represented groups and the achievement of quality improvements in relation to mission and plans; development and implementation of a student access guarantee; and faculty complement counts. For universities specifically, the MYAA's include student retention reporting (generally using the Consortium on Student Retention Data Exchange model); and participation in the NSSE survey and a graduate student satisfaction/experiences survey (known in Canada as the Canadian Graduate and Professional Student Survey or CGPSS developed by the G-13 group of Canadian research universities). One of HEQCO's primary objectives is to make recommendations to the Province on the structure and role of MYAA's in the evolving accountability framework, and on the value and role of NSSE within those MYAA's as a quality measurement tool.

1.5 Project Origins, Objectives and Development

In April 2007, HEQCO organized a workshop on NSSE entitled "What Have Ontario Universities Learned?" at which several university presenters discussed their approaches to, and progress on, NSSE analysis, dissemination and implementation. In July 2007, HEQCO released its first *Review and Research Plan*, in which NSSE was identified as having significant potential applications in quality assessment and accountability (while acknowledging the uncertainties discussed above). On the heels of the Rae Review, the *Reaching Higher* plan and the introduction of MYAA's, it appeared that at least within the Ontario context, a formal approach to NSSE implementation, and the organizational learning, assessment experience and policy contributions that might flow from it, would provide greater certainty about the role of NSSE in guiding quality improvement and enhancing accountability at Ontario universities. As indicated above, U.S. and Canadian implementation experiences also demonstrate that documentation of NSSE-driven quality practices and their results is scarce (or at least difficult to assemble and pursue). However, it is clear that participation declines over the spectrum of implementation approaches; that institutional efforts to improve quality through a focus on improvements in NSSE results are hindered by a lack of practice models and hence considerable uncertainty; and that linkages between the design and assessment of such efforts are at best uncertain (i.e., it is unclear what types of interventions have what effects and at what cost, and how they can be structured to maximize the possibility of measuring whatever their effects might be).

This project is the ultimate result of an unsolicited research proposal to HEQCO by the author that became increasingly detailed before and during its initial phase (see below). The goal of the project is the design, implementation, assessment and documentation of a series of engagement-related interventions at several Ontario universities in order to achieve four primary objectives:

- Objective 1: Documentation Establish a "manual" of appropriate data/measurement practices and intervention implementation protocols/procedures for various engagement improvement activities that are both relevant in the Ontario context and that might be generalizable (with adaptations) across the university sector;

- Objective 2: Promising Practice Create and share (among project participants and more widely) an inventory of implementation practices/experiences to support implementation and assessment efforts in the longer term;
- Objective 3: Assessment Using the best available data, intervention designs and statistical procedures, conduct formal analysis of the extent to which intervention effects can be detected and measured using NSSE and other tools;
- Objective 4: Quality and Accountability Policy Inform policy discussions related to the accountability and reporting applications of NSSE within and apart from the MYAA's, and to quality improvements generally.

The project was undertaken in two phases, each separately funded by HEQCO. In the first phase (June 2007 – October 2007) several activities were undertaken in order to assess the feasibility of a second (implementation and assessment) phase:

- Representatives of all 19 universities in Ontario were invited to submit intervention project proposals using a standard proposal submission form;
- An inventory of US implementation practice was conducted;
- Support was provided to universities as they developed their proposals including a workshop at which intervention development, assessment design and other key issues were discussed;
- Sixteen proposals submitted by 13 universities were assessed by the project steering committee using a standard assessment template; 13 projects at 11 universities were approved and incorporated into the Phase Two proposal;
- Initial arrangements were concluded with NSSE staff for a limited administration of NSSE 2009 consistent with each of the approved designs.

Based on Phase One results, HEQCO committed to funding the second phase of the project (November 2007 – November 2009), which included:

- Documentation and assessment of Phase One activities;
- Refinement of intervention designs, data sources, implementation procedures, budgeting and staffing at each of the participating universities;
- Submission of pre-intervention data and information (including NSSE 2008 administration);
- Assembly and setup of all databases;
- Intervention implementation (generally September 2008 – April 2009);
- Submission of post-intervention data and information (including NSSE 2009 targeted administration);
- Determination of the value of NSSE and other tools in intervention assessments;
- Project reporting (progress reports and this final report).

1.6 Report Outline

Section Two summarizes the processes followed and the issues encountered in Phase One of the project, and provides an assessment of this phase. Section Three presents an

overview of the various intervention design and assessment issues that were either anticipated at the outset, encountered during intervention implementation, or uncovered through the assessments. Section Four deals with the assessment of the "fit" between NSSE and the interventions. Because both intervention design and assessment methodology are dependent on context and data availability, the background and context for each intervention are briefly provided, along with both a formal statistical assessment of each intervention, and a qualitative assessment of the overall project. Section Five presents the author's views on the implications of the intervention projects and assessment results for intervention design, implementation and assessment practice, and for university quality assessment and accountability policy (with specific reference to the role of NSSE and other measurement tools in the MYAA process).

References, background materials, copies of survey instruments and similar items are available on the internet; the URL's for these items are provided in Section Six.

2. Phase One: Intervention Design and Preparation

2.1 Phase One Description and Objectives

The purpose of Phase One of this project was to establish the case for the full-scale design, implementation and funding of engagement-related interventions at multiple sites. As indicated above, this objective was fulfilled through a series of activities:

- An inventory of U.S. NSSE implementation practice was undertaken to demonstrate existing documentation and assessment approaches, and to identify the continuum of implementation activity within which to locate and justify Phase Two (summarized in 1.3 above);
- Representatives of universities considering participation in the project were given an orientation to the project, intervention assessment data sources, and design and assessment requirements; and were provided support as they developed (and ultimately refined) their proposals (as discussed in more detail below);
- Intervention project proposals from Ontario universities were solicited, and were assessed using a standardized template; a subset of these proposals was selected by a review committee at Queen's University for the project, subject to confirmation of funding;
- A detailed budget and schedule for the 13 recommended projects was developed, submitted to HEQCO, and ultimately approved.

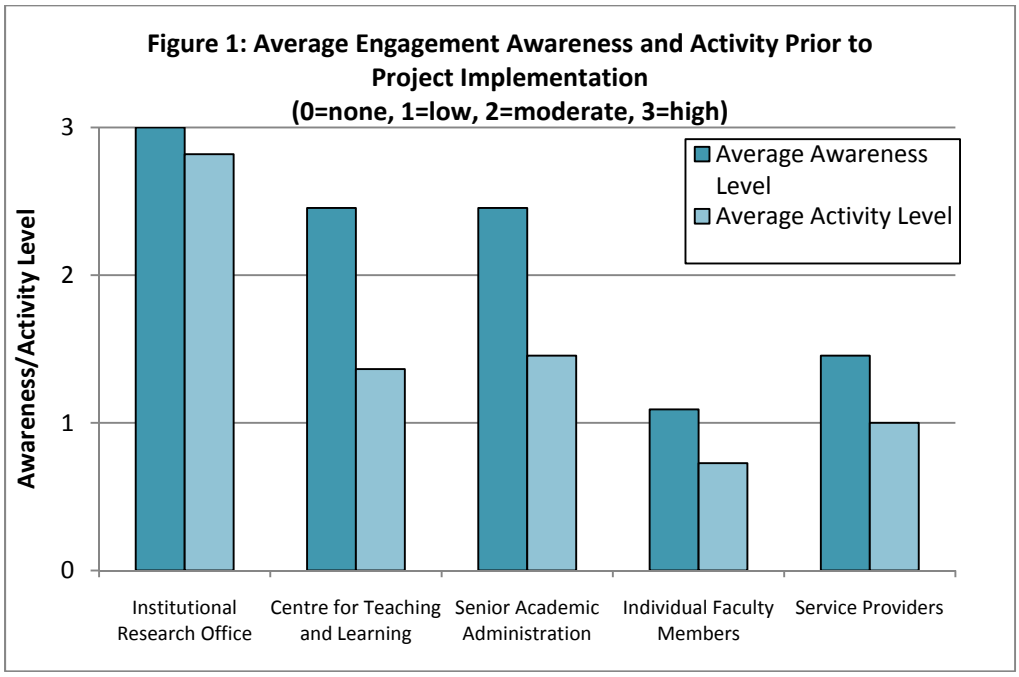
Early in Phase Two, representatives of participating sites were surveyed twice to identify the issues and problems encountered during Phase One proposal design/development and initial intervention planning. The first survey of participating sites attempted to establish a few baseline measures of NSSE/engagement activity: engagement awareness and activity levels within various institutional offices, number of previous NSSE administrations, and self-assessed location of each institution (current and projected) along the continuum of NSSE practice presented in 1.3 above. It also generated information about the proposal development process: key drivers in the decision to prepare a proposal, factors affecting the choice of intervention, administrative arrangements for proposal development, usefulness of information sources, and actual and anticipated difficulties prior to implementation. The second survey collected entirely qualitative information at the mid-point of intervention planning: progress and problems to date, additional assessment data sources identified, changes in intervention or assessment design contemplated, and additional support required.

2.2 Phase One Evaluation

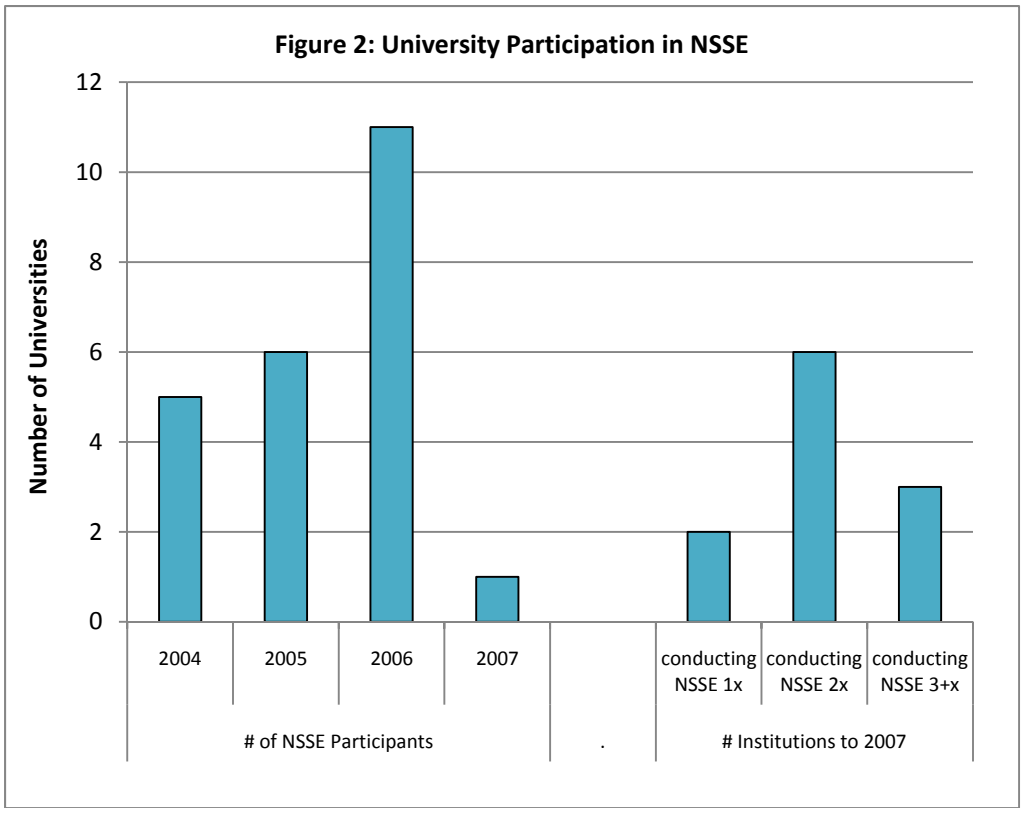
2.2.1 Pre-Project Engagement Awareness, Activity and NSSE Participation

This section and the two following draw in part on the first of the two surveys noted in 2.1 above. Prior to commencement of project activity, awareness of NSSE and engagement,

and activity levels related to engagement varied significantly across university stakeholder groups (Figure 1). Institutional Research offices were consistently considered the most aware and most active, followed by Centres for Teaching and Learning and senior academic administration (both with "moderate" to "high" awareness and "low" to "moderate" activity levels), service providers and individual faculty members (who were seen as having "low" to "moderate" awareness and "low" or "no" activities). With the exception of Institutional Research Offices (with their consistently high awareness *and* activity levels), other stakeholders are generally reported as having awareness levels higher than their activity levels, suggesting that activity lags behind what for some would be relatively recent awareness of NSSE.

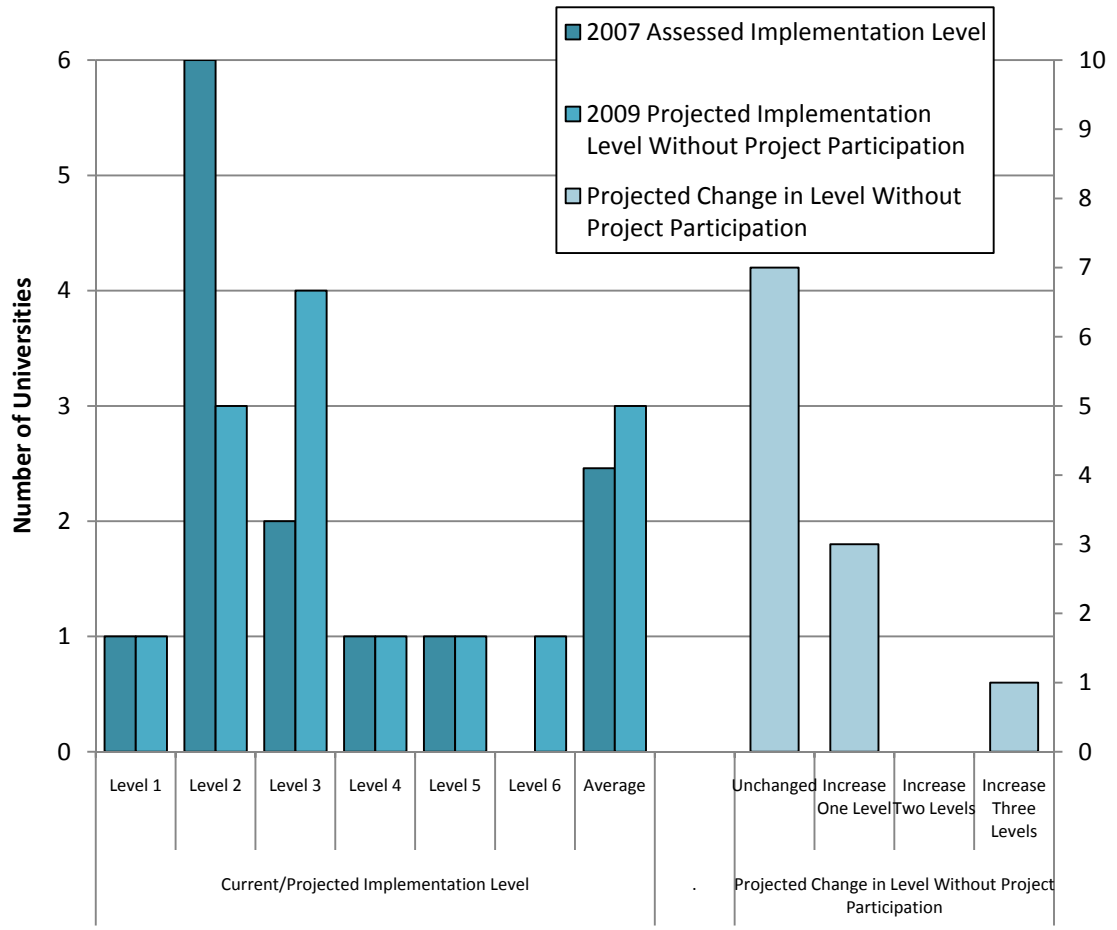


In 2004, five of the 11 participating universities administered NSSE; six administered in 2005; all 11 in 2006; and one in 2007. (All participants also administered NSSE in 2008 and incorporated over-sample as a baseline measure for intervention assessment.) Across the entire four-year period 2004 to 2007, two of the universities had administered NSSE only once (in 2006); six had administered twice, and three universities had administered three or more times (Figure 2).



It appears that self-assessed engagement-related activity is at least loosely related to the number of NSSE administrations. More specifically, institutional representatives were asked to locate their university along the implementation continuum presented above in 1.3 – both at the current time and as projected in two years assuming no involvement in the interventions project (Figure 3). While some progress (e.g., from Level 2 to Level 3) was projected to occur, half of the universities predicted they would not move forward along the continuum.

Figure 3: Current Implementation Level and Projected Level Without Project Participation



2.2.2 Factors Affecting Participation and Proposal Development

While institutional responses to the awareness and activity level questions were quite similar, responses to questions about factors affecting the decision to submit an intervention proposal and issues encountered in proposal development varied widely (Table 2 and Figure 4).

The impetus behind the decision to prepare a proposal originated with a number of factors. The most important were the relevance of project participation to planning and performance measurement, the availability of intervention funding and the proposal submission form, the participation of other universities, and the relevance of the intervention to engagement implementation activities already underway.

Participants ranked the proposal submission form, support from the principal investigator, the participant workshop, and within-university support as the most useful sources of support and information during proposal development. The inventory of U.S. implementation practice was considered much less useful, and only a minority of participants relied directly on support provided by NSSE staff. The greatest difficulty reported by project participants in proposal development was establishing the assessment design (i.e., defining and isolating an intervention target group and determining the appropriate measurement strategies). This issue was generally reported as having caused "moderate" difficulty, while others generally posed no or only minor difficulties (e.g., identifying internal expertise, providing planning context, linking the intervention to specific NSSE items).

Table 1: Factors Affecting Participation Decision and Proposal Development (Raw Responses)

Rate the importance of each of the following as contributing factors in your decision to prepare an intervention proposal

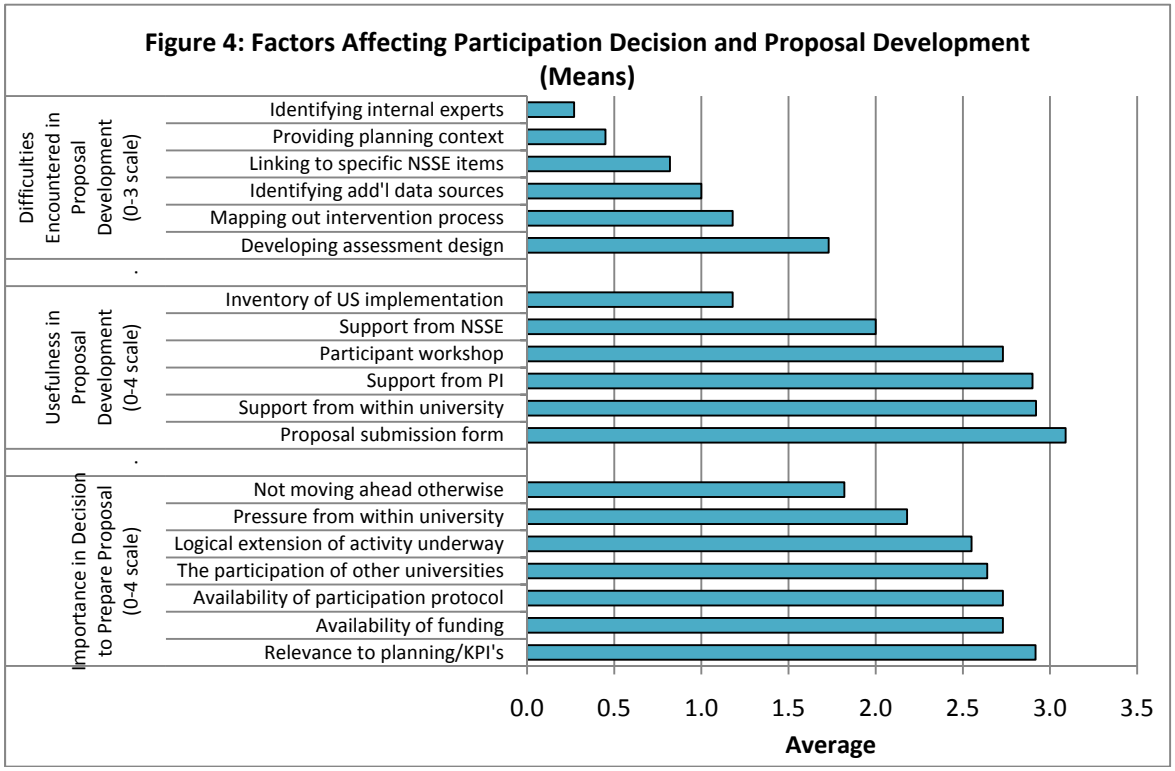
	none/very low	low	moderate	high	very high
Availability of funding		1	4	3	3
The participation of other universities		1	3	6	1
The availability of a protocol to assist in intervention design		2	5	4	
Pressure from within university		3	4	3	1
Relevance to planning/KPI's			4	1	6
Seemed not to be moving ahead otherwise	2	2	3	4	
Logical extension of implementation work already underway	1		5	2	3

Rate the usefulness of each of the following in preparing your proposal once the intervention had been decided upon

	none/very low	low	moderate	high	very high
Proposal submission form			1	8	2
September 11 workshop		1	4	3	3
Inventory of US implementation	3	3	5		
Support from PI			3	5	2
Support from within your university	1		2	5	4
Support from NSSE		2	1		1
Other (colleagues, research, working group)				2	2

Rate the difficulty you encountered completing each of the following components of the proposal submission form

	none	a little	moderate	a lot
Providing planning context	8	1	2	
Identifying add'l data sources	4	4	2	1
Developing the assessment design		5	4	2
Mapping out intervention process	2	6	2	1
Linking to specific NSSE items	6	1	4	
Identifying internal experts	9	1	1	



2.2.3 Preparation and Implementation Issues

Institutional respondents provided written responses describing the process through which their interventions were decided upon, and the processes used to develop the intervention from an idea into a formal proposal. Again, institutional approaches to intervention selection and development varied widely in terms of organizational structure, delegation of responsibility, involvement of senior academic administrators, and relationship to engagement efforts already underway. The general process can be described as follows:

- In some cases, a pre-existing structure (committee, department, designated individual) provided a vehicle for the management of "student experience" issues. In others, a structure was created in response to the invitation to submit the proposal (e.g., ad hoc committee, meetings among researchers and senior academic officials).
- Where structures preceded the invitation to submit proposals, they appear to have served as project "champions" themselves or to have facilitated assignment of intervention responsibilities to such champions. Because the invitation to submit was sent to institutional research representatives, they necessarily championed the proposal through its early stages; however, the majority of activity related to proposal development was shared amongst faculty members, teaching and

learning staff and academic administrators with institutional researchers playing a more limited role as time progressed (but continuing to serve a liaison role with the Principal Investigator (PI) and generating the data files necessary for documentation and assessment).

- In about half the institutions, engagement activity (sometimes preliminary and in some cases substantial) was already underway in response to previous NSSE results, and they were able to modify their current activities in order to conform to the proposal requirements (generally by incorporating the assessment component). These institutions had performed more additional analyses (particularly Faculty- and program-level drilldowns) and were able to utilize this information in intervention design.
- The majority of participants described at least one actual or potential barrier to intervention design and eventual implementation including constructing an assessment design consistent with available data, instructor buy-in to the project generally (where necessary) and for an in-class assessment survey, achieving a sufficiently high student participation rate in the intervention, and survey response rates.

2.2.4 Diversity, Complexity and Scale of Interventions

The funding available encouraged the development of, and resulted in, "modest" intervention designs. While initial budget requests from participants were significantly higher, final project budgets averaged less than $14,000. Budget reductions were achieved in two ways: by scaling back project activity, and by substantial actual and in-kind contributions from the participants themselves. The interventions targeted as many as 1,500 and as few as 100 students, necessitating careful targeting and monitoring of project participants and the achievement of acceptable survey response rates to prevent reduced statistical power in the assessments. The interventions are relatively straightforward as the result of their limited focus and reach, and are able to be implemented and administered with small project teams (typically 2-4 people).

The intervention proposals also reflect the priorities at each of the participating institutions and the NSSE results on which they were predicated. Of the original 13 projects, 10 were focused on first-year students (a group for which Ontario and Canadian universities generally show much lower Student Faculty Interaction (SFI) and Active and Collaborative Learning (ACL) engagement levels than their U.S. counterparts), one exclusively on fourth-year students, and two on both. Two projects involved single course interventions, eight involved multiple course interventions, and three were directed toward non-course activities. A more equal balance between course- and non-course-based interventions and first- and fourth-year students would have been desirable. On the other hand, the course-based projects resulted in easier subject targeting and were able to rely in several cases on the CLASSE instrument as an assessment tool, and more closely addressed issues of current concern within Ontario (e.g., the first-year experience).

2.3 Phase One Output - Intervention Project Proposals

As noted above, 13 intervention projects were approved for Phase Two funding and all proceeded through the planning and initial implementation stages. However, three of the interventions were terminated prior to completion:

- One project dealing with targeted career services for special needs students began with a relatively small target population. It encountered a lower-than-expected intervention participation rate and a high attrition rate, and as such, did not sustain a target group of sufficient size to permit formal assessment.
- A second project involving the development and augmentation of a mathematics help centre encountered administrative and data collection difficulties that made it impossible to implement the intended assessment design.
- A third project proposing a course-based learning community and delivery enhancements was affected by a lengthy labour dispute on campus that compromised the integrity of the intended pre- and post-intervention measures and left too little time for implementation.

A fourth project encountered participation rate and survey response rate difficulties well into its implementation. While it could not be assessed in the same manner as the other projects, its context and background are presented along with a modified set of analyses.

An overview of the ten intervention projects is provided below to demonstrate the wide range of interventions developed at the participating institutions.

- Carleton University: Development of a TA Mentorship Model
  Teaching assistant mentors were assigned to five academic units (one per Faculty) in order to provide training and support to the units' teaching assistants (TAs), and in particular, those TA's assigned to first-year courses. Virtually all first-year students take one or more courses involving the mentored TA's. The objective of the intervention was to improve the TA-related student experience in response to pre-existing concerns (corroborated by NSSE core questions and the Ontario NSSE consortium questions).

- University of Guelph: Supported Learning Groups in High Risk Courses
  Senior student peers provided out-of-class group study and review sessions in six single-semester high-risk first-year courses having either high dropout/failure rates and/or low grades, using an approach modeled after the University of Missouri (Kansas City) "Supplemental Instruction Program". The initiative addressed Guelph's objective to improve the first-year experience through increased retention and the development of learning skills.

- University of Ottawa: Faculty of Social Sciences Course-Based Learning Community
  First-year students self-selected to participate in a program in which they were assigned to groups having (nearly) common course timetables, attend extra weekly meetings and workshops, and receive senior student mentoring support. The academic, cultural and study skills orientation of the program was intended to

address new student integration into a large Faculty and large university as prescribed in the strategic plan.

- Queen's University: Enhanced Student-Faculty Interaction in a Large Introductory Course
  To compensate for limited student-faculty interaction opportunities in Introductory Psychology (with section enrolment exceeding 300) and to better integrate research issues into course content, students self-selected to participate in the "Discovery Project" – a series of small group activities with renowned faculty members and practitioners introducing students to real world research, laboratory visits, psychological experiments and professional practice.

- University of Western Ontario: Improvement of Science Literacy Through Course Re-Design
  The Biology Science Literacy Initiative integrated the development of science literacy skills into the first-year Biology curriculum, using two large Biology courses. Seminar activity and on-line supports focused on information retrieval and evaluation, information integration, science writing and critical analysis of science writing. The project reflects the University's strategic plan focus on the student experience and the Faculty's academic plan commitment to improved teaching and learning.

- Ryerson University: Improving Writing Skills in Selected Academic Programs
  One required first-year course in each of eight programs across an entire Faculty was selected to house a range of curricular changes and service enhancements (writing tutorials, rapid turnaround of writing assessments) to improve writing skills competencies and highlight writing skills as a learning outcome distinct from course content. Writing competency had previously been identified as a Faculty objective based on NSSE results and is integrated into Ryerson's priority-setting and accountability processes.

- Wilfrid Laurier University: Peer Learning Program for Literacy, Research and Writing Skills
  WLU offered a peer (senior student) delivered learning program designed to improve the information literacy, research skills and writing skills of students in two introductory writing-intensive courses. The program consisted of a 3-day peer student training program prior to the start of classes, and the delivery of skills development sessions throughout the Fall term. The intervention reflected concern over student ability to make the transition to university and university-level writing.

- University of Waterloo: Curricular Re-Design via a Teaching Excellence Academy
  Waterloo's TEA provides participating faculty members with intensive training on course design: learning objectives and techniques, assessment methods, accommodating student characteristics and course evaluation. Two introductory and one senior course taught by faculty completing the TEA were selected for the intervention. The basis for the project lies in Waterloo's strategic plan and provincially established degree level expectations.

- University of Windsor: Intrusive Faculty-Wide First-Year Advising
  First-year students in the School of Business received a significantly enhanced advising program consisting of regular contacts with faculty members and senior student mentors to deal with academic plans and progress, Q+A's, student activities, and course/program issues. The intervention reflects concern over the level of social and academic integration of students as reflected in NSSE results and in the expectations-experiences gap between BUSSE and NSSE.

- Queen's University: Computer Enhanced Tutorial and Academic Support Integrated Across Courses
  Fourth-year students in the Electrical Engineering program were provided access to on-line real-time extended-hours tutorial support services (in addition to conventional classroom support) integrated across three Fall term and three Winter term courses that share a common foundation, to facilitate both individual and group study. The project addressed a general concern over the level of academic support and curricular integration, and program-specific concerns with NSSE results.

3. Intervention Design and Assessment Issues

3.1 Pure and Quasi-Experimental Designs

Pure experimental designs require "identical" control and experimental groups achieved through random subject selection, and a distinct "treatment" that is applied to the experimental group and whose effect can be measured independent of all external influences. In fact, further conditions sometimes apply to pure experimental designs to ensure against the intrusion of factors that may distort measured results (e.g., double-blind administration, placebo control and multiple trial repetition). When such conditions are met, it is generally appropriate to attribute differential outcomes across control and experimental groups to the treatment itself.

Such conditions are rarely if ever achieved in an applied social research environment, and certainly not in university settings involving "experiments" on students. Random selection of students to control and experimental groups often poses serious research ethics concerns (e.g., the effective denial of potential benefits to the control group members and the need for appropriate compensatory treatment). As such, students are often invited to self-select to participate in a revised service or program, leaving in doubt their comparability to those who do not self-select. The diverse and ever-changing university academic and social environment is hardly a stable one in which the effects of curricular and service changes can be easily isolated and measured – particularly over time. Fortunately, a number of quasi-experimental assessment designs, measurement tools and statistical techniques are available that attempt to approach the "laboratory ideal" by accommodating one or more of these real-world complexities, particularly when applied to research on university student behaviours.

3.2 Survey Non-Response

Student satisfaction and experience surveys are central to many aspects of university assessment, and the NSSE survey in particular is a key element in both this project and quality assessment generally. Student surveys typically achieve response rates in the 20 per cent - 50 per cent range and thus create a number of assessment issues:

- The comparability of responders and non-responders (i.e., non-response bias) affects the generalizability of survey results to the entire student population;
- Varying survey response rates over time within the same population can create difficulty in the interpretation of trends;
- Experimental and control groups – and therefore the scope of the project – must be large enough to result in net survey sample sizes that are sufficiently large to permit rigorous assessment;
- The clearly desirable practice of relying on multiple surveys in an assessment can ironically result in a decline in survey response rate due to survey fatigue (for the second and subsequent surveys) and a reduced net sample overall (because fewer students respond to all surveys than to each of the individual surveys);

- Specific item non-response by survey responders may necessitate the elimination of the entire response from the assessment.

The response to all these potential problems is to construct sufficiently large control and experimental groups, achieve the highest possible survey response rates, conduct comparisons where possible of responders and non-responders to identify potential non-response bias, and exercise caution when inferring the results to the entire student (as opposed to the entire respondent) population.

3.3 Intervention Participation Rate

The identification of a target group does not necessarily define experimental participants in a university setting, given the often voluntary nature of academic and non-academic services and the right of students to self-select for participation. For example, if the content and delivery of a specific course are modified and all students encounter these modifications, the participation rate is effectively 100 per cent - the target group is the experimental group. However, if a service to a particular category of students is made available on a voluntary self-selection basis, then the target group and the actual participant (experimental) group are clearly different. Proper assessment requires that actual participants be tracked in cases where targeted students may experience the experiment in different ways (or not experience it at all).

Three key issues must be addressed. First, some students will "exit" the experiment (if the experiment is structured to permit such departures). Failing to account for attrition – particularly early attrition – is likely to distort outcomes measurement by assigning participant status to all students initially identified as participants regardless of their subsequent attrition. As a result, the tracking of attrition is necessary, at least in the form of end-of-project status checks. Second, ongoing participants will possibly be involved in the experiment at varying levels of intensity. If a service experiment consists of numerous components – say, weekly voluntary enrichment sessions in a semester-long course – a "participant" could attend as few as one or as many as 12 sessions. As such, an intensity of participation measure is appropriate for experiments that permit varying levels of involvement. In analysis terms, this suggests both a single experimental group (ignoring intensity of involvement) and multiple experimental groups (each reflecting a point/range on a graduated intensity of involvement scale). Third, one intervention project discussed in the following section actually consists of six simultaneous experiments: students were able to participate in multiple experiments, and at varying levels of involvement within each. This permits an analysis not just of student intensity of involvement but also of scale of institutional effort. The tracking and data collection required to accommodate attrition and intensity of involvement/effort may be onerous, but contribute to improved measurement and assessment.

Experimental participation rate is more than an analytical complexity that needs to be controlled in assessment activity. The characteristics of self-selectors relative to non-self-selectors can provide an important basis on which to assess the recruitment success of voluntary participation experiments and possibly to improve the design of recruitment strategies.

3.4 Participant Self-Selection Bias and Propensity Matching

Given the voluntary nature of participation in many university academic and service experiments and the right of students to self-select, assessments must incorporate the possibility that members of a self-selected experimental group differ from non-self-selectors in a way that is directly related to the focus of the experiment. For example, it is inappropriate to compare post-experiment engagement survey results for students who self-selected for a course enrichment activity against a control group of non-self-selectors, when pre-experiment engagement itself may have been a key factor in self-selection. Such an analysis would be unable to differentiate between experimental and predisposition effects and would therefore fail to satisfy the central purpose of the assessment.

While control and experimental groups may differ in their composition due to self-selection, matched groups can be simulated in several ways. The most basic is to weight the records in either group based on one or more characteristics that are known (or believed) to drive self-selection. For example, if female students in the 80.0 per cent - 84.9 per cent entering grade average range represent 5 per cent of the experimental group and 10 per cent of the control group, such experimental group students could be assigned a weight of 2.0 (to simulate a 10 per cent contribution). Corresponding weight adjustments would be applied to experimental group males and females across all grade averages. Such an approach assumes that gender and grade average are the appropriate mechanism for group matching. Another approach is to incorporate self-selection drivers such as gender and grade average as covariates in a multiple regression model. The outcome measure would then be expressed as a function of multiple factors: participation in the intervention, gender and grade average.

A more sophisticated approach, and the one employed in this project, involves the "pairing" of each experimental group member with a "highly-similar" control group member, on the basis of characteristics that are thought to influence self-selection and/or outcomes predisposition. All interventions involved either potential participant self-selection bias or required confirmation of control and experimental group similarity. Propensity matching was performed using a SAS macro developed by the Mayo Clinic. The macro utilizes an algorithm that creates a "distance matrix" by computing the weighted sum of the absolute differences between control and experimental group records on the basis of the matching variables, and matches each control group record with the experimental group record having the smallest difference (i.e., the greatest similarity). The maximum difference on a match can be specified to ensure against weak matches and weights can be used to give greater importance to certain matching variables. Multiple control group members can be matched with each experimental group member to increase the statistical power of analytical procedures.

The matching process proceeds in several steps:

- Matching variables were selected on an "as available" basis for each project, and included two or more of gender, admission average, admission type, year of study, part-time/full-time status, Faculty/ program of enrolment, domestic/international status and/or age;

- Control and experimental groups were initially compared on all matching variables to establish a baseline for evaluating the success of subsequent matching: logistic regressions (intervention participation status as a function of matching variable) were conducted one matching variable at a time to avoid multicollinearity and produce a conservative estimate of group differences;
- The algorithm was applied to match control and experimental group records;
- Logistic regressions were repeated after matching to test the elimination of significant differences across the groups on the matching variables. In the event that there were differences after the match, weights used in the matching process were adjusted to give greater importance to matching variables that differed between control and experimental groups post-match.

Where all significant variation between groups is eliminated, assessment can be undertaken using simple regression, where the outcome is expressed solely as a function of intervention participation. Propensity matching attempts to simulate random subject assignment, but it guarantees group similarity only with respect to the matching variables used. To the extent that certain variables are known to influence self-selection and engagement, and to the extent that these can be utilized, propensity matching is a relatively powerful device in assessment design.

3.5 Dilution Effects

The measurement of impacts resulting from engagement interventions in a university setting is subject to two kinds of dilution effects. First, if a small experimental group is contained within a much larger population, and if it cannot be isolated for measurement purposes, then however large the impact of the experiment, it will likely be masked. For example, even a highly effective student service enhancement in which 100 of 3,000 first-year students participate is unlikely to show an effect if measurement occurs across the entire first-year population. This problem can be managed by targeting the intervention to permit clear identification of the experimental group (in conjunction with addressing the participation rate issues of attrition and intensity of involvement discussed above).

Second, if an experimental group is subjected to an experiment whose effects are confined to a small proportion of the total student experience during the experimental period, those effects may not be detectable. For example, an enhancement within a single one-semester course during a school year in which students are registered in 10 such courses is unlikely to show an effect if measurement focuses on the academic year in total. This issue is critical in the current context. NSSE explores various aspects of the student's experience over an entire academic year. While a primary purpose of this project is to assess the ability of NSSE to capture the short-term engagement impacts of relatively modest interventions, other measurement tools might, in theory, be better suited to the purpose. As such, intervention designs and assessment strategies incorporated numerous other measurement tools, including CLASSE (a course-specific version of NSSE currently in pilot testing) for course-specific interventions, and numerous other surveys, focus groups and educational outcomes data (e.g., grades) for most interventions (see 3.9 below).

3.6 Sampling Error

Sampling is subject to statistical error that is reduced, but not eliminated as sample size increases. Even the construction of large matched control and experimental groups (whether by random selection or propensity matching) is therefore an inexact science. Survey results are often presented with the proviso that the results are considered accurate within a certain range (about 5 per cent) most (or 95 per cent) of the time. A maximum error in the 5 per cent range may be minor in some contexts, but is much more problematic in the current context. For example, a random sample of n=1,000 is considered accurate within 3 per cent, 95 per cent (or most) of the time. In such a situation, however, a measured (post-experiment) difference between a control and experimental group of 2 per cent (without other forms of corroboration) will not permit a firm conclusion as to whether the difference is due to an experimental effect or random error: experimental effects larger than those attributable to sampling error are required to achieve such reasonable certainty. As a result, it is common to rely on a "preponderance of evidence" approach, in which multiple tests all pointing to the same conclusion may provide a level of confidence in the results that a single test cannot.

3.7 Background Noise

Assessment is often based on control and experimental groups (samples) drawn from candidate populations – either from successive populations one of which is subject to the experiment, or from within a single population, a portion of which is subject to the experiment). In the former case, between the time of the first and second samples, other exogenous events (non-experiment related) may occur that are responsible – rather than the experiment – for measurement differences between the control and experimental groups. The possibility that such "background noise" exists must be accommodated when interpreting experimental results employing measures taken over two or more points in time. Generally speaking, quasi-experimental designs employing surveys over time do not permit the complete elimination of the background noise problem. The approach used in the following assessments was to observe general (and where available, specific item) trends over time for the university overall (and where available, for relevant drilldowns) and to interpret intervention results cautiously where warranted. A more formal correction for background drift was implemented in one of the interventions.

3.8 Experimental Designs Used in Intervention Assessments

The best possible quasi-experimental assessment design is therefore one that utilizes matched groups (either propensity-matched groups in a cross-sectional design, or matched groups over time while acknowledging possible changes in the background environment); that targets and isolates the experimental group; that incorporates and measures varying levels of intervention involvement; and that utilizes a variety of measurement tools to maximize the probability of detecting the experimental effect without dilution.

Two general assessment designs were utilized. Successive cohort designs were originally envisaged to be, and ultimately became, the primary approach used. For most intervention projects, this involved normal full administration of NSSE (in some cases in combination with other surveys) in 2008 with a 100 per cent sampling rate for students in a specified group whose following year successors would become subjects of interventions in 2008/09; and a specially targeted administration of NSSE (and possibly other surveys) in 2009 with a 100 per cent sampling rate for students who were in fact subjects of the intervention in 2008/09. The first cohort is the control group; the second cohort is the experimental group. NSSE research has indicated that for reasonably large student cohorts, successive cohort designs are appropriate because of the stability of student characteristics and survey response behaviour over the short term (i.e., one or two years). However, two additional safeguards were employed in the analysis phase. The control and experimental groups were compared to determine whether they were in fact similar on key characteristics, and propensity matching was performed. Second, all available NSSE results from administrations up to and including 2008 were examined – particularly for survey items that constituted the dependent variables (i.e., expected impacts of the interventions) – to provide a basis for the subjective incorporation of background noise into the analysis.

Cross-sectional designs were also employed, either as the primary intervention assessment device, or in combination with a successive cohort design. Cross-sectional designs required a single administration of NSSE (and/or another survey) simultaneously for both control and experimental group members. Generally, the survey was administered post-intervention, and the design required propensity matching of experimental subjects with selected members of the generally larger experimental group in order to control for self-selection bias. In one intervention project, both pre-experimental and post-experimental surveys were undertaken on both the control and experimental groups (not on successive cohorts of both groups, but within both groups, pre- and post-experiment).

3.9 Supplementary Data Sources

NSSE is not, nor does it purport to be, a quality assessment panacea. It has proven value first as an engagement benchmarking tool at both the university-wide level and (through drilldowns) at the Faculty, academic program and student subgroup level as well. Second, it has provided a clear focus for generating engagement-related discussions and supporting engagement-improving activities at literally hundreds of universities. However, its value as a tool to measure engagement changes resulting directly from specific service and academic experiments (particularly for relatively small experiments and over the short term) has not been widely explored. As such, it is appropriate to incorporate a number of additional data sources into intervention assessments – both as alternative intervention outcome or impact measures, and in order to strengthen the analysis generally.

The quality and quantity of supplementary data vary by institution and with the nature and objectives of the intervention itself. Some data were available at the student record level and were merged with NSSE response data (primarily though not exclusively to permit propensity matching). Other data were available in aggregate form only and provided

general support and context to the assessments. Each of the assessment designs employed one or more of the following supplementary data sources:

- Time-series university-wide NSSE results for all available administrations (items and benchmarks) were used to assess the stability of the background environment in which the interventions were undertaken;
- Strategic, academic and service and/or operational plans, policy statements and performance indicators and targets provided context for the development and implementation of the interventions;
- The results of previous surveys provided both general context and supported intervention design (e.g., institution-specific satisfaction/experience surveys, exit surveys, service-specific surveys, and results from the Canadian University Survey Consortium (CUSC));
- CLASSE was employed in a number of course-specific interventions, and was modified to varying degrees to reflect the specifics of each project;
- Several other intervention-specific surveys were undertaken;
- Intensity of involvement measures were developed for several of the interventions in which varying levels of participation were possible;
- Faculty members and service providers were interviewed before, during and after several of the interventions;
- Focus groups were undertaken on students, faculty members and service providers;
- Three rounds of questionnaires were administered (immediately following proposal approval, at the mid-point of intervention planning/design, and near the completion of the project) to assess the views of administrators, faculty members and other intervention project participants;
- A variety of demographic and academic data items at the student record level were collected for both the control and experimental groups, including secondary school grades and grade average, semester or cumulative grade average at university, specific university course grades/status measures, retention/attrition behaviour following the intervention, basis of university admission (i.e., immediate entry from secondary school vs. delayed entry with or without postsecondary activity), and full-/part-time status, gender and program major (from the student records system rather than the NSSE survey itself).

3.10 Qualitative Assessment

Two primary tools were available for qualitative assessments of the interventions. The first was the series of three participant questionnaires referred to above. The questionnaires provide insight into the process, management and implementation of the interventions themselves, and are discussed in a separate section for all projects combined. The key value of the qualitative assessment is the insight it provides to future project design and assessment. The second tool consisted of additional qualitative information collected by each of the participating universities that contributes to both process and outcome assessment (e.g., focus groups, interviews, supplementary surveys). While this report identifies many of these additional information sources, it does not attempt to analyze them or incorporate them into formal statistical assessment: this

activity is better undertaken by the participating universities themselves. In a few cases, qualitative (and some quantitative) data collection and analysis continue to be undertaken but are not yet available for this report.

3.11 NSSE Validity and Assessment Design

As noted above, NSSE is widely perceived to be both valid and reliable, which is to say it is seen to measure what it purports to measure (validity) and it does so with a high level of consistency (reliability) over successive administrations. Dr. Stephen Porter of Iowa State University presented a paper at the 2009 meeting of the Association for the Study of Higher Education (ASHE) containing a number of analyses on NSSE data and concluding that on certain examined items, NSSE's validity and reliability were questionable. Assuming Porter's statistical findings to be correct as presented, they are largely immaterial to this project. Porter's findings demonstrate the gulf that occasionally exists between the science (statistics) and art (interpretation and implementation) of survey research. Within-institution (rather than institution-wide or cross-institutional) analysis of similarly constructed (matched) samples avoids most of the potential problems Porter identifies. Only unmatched (or questionably matched) samples that are subject to significant cognitive or pedagogical bias present serious validity and reliability concerns. NSSE data analysis and implementation practice can adopt a number of strategies to avoid validity and reliability concerns, including careful peer set construction, program- and student subgroup-level drilldowns (i.e., targeting), time-series analysis of NSSE response data, and concentrating on only major cross-group engagement differences (which are likely to be meaningful regardless of potential validity and reliability limitations).

3.12 Summary

As indicated above, the intervention and assessment designs attempted to satisfy as many of the above criteria as possible. None were able to satisfy all, but each reflects the reality and limitations of applied research in a university setting. A summary of the assessment designs is provided in Table 2 below.

Table 2: Summary of Assessment Designs by Project													
		Assessment Design			Design Enhancements			Measurement Tools (in addition to NSSE)		Dependent Measures (in addition to NSSE)			
University	Intervention Project	Successive Cohort	Pre/Post Measure	Cross-Sectional Post Measure	Propensity Matching	Intensity of Involvement	Qualitative Assessment	CLASSE (modified)	Other Surveys	Grades	Attrition/ Continuation	Academic Program Choice	
Carleton	TA Mentoring Program	X		X	X	X	X			X			
Guelph	Supported Learning Groups	X	X		X	X	X	X		X			
Ottawa	FSS+			X	X					X	X		
Queen's	Psychology Discovery Project	X		X	X	X	X	X	X	X			X
Western Ontario	Science Literacy Initiative	X		X	X	X	X	X	X	X			
Ryerson	Writing Skills	X			X		X		X	X	X		
Wilfrid Laurier	Peer Learning Program	X			X		X			X			
Waterloo	Teaching Excellence Academy	X		X	X		X			X			
Windsor	Intrusive Advising	X			X	X							
Queen's	Engineering On-Line Support	X		X	X	X		X		X			

4. Intervention Assessment Results

4.1 Intervention Assessment Overview

This section provides a description of each intervention with specific reference to the assessment design issues discussed above, and presents the results of the formal statistical analysis to which each was subject. An essential point made earlier is worth repeating here. A primary purpose of this project overall, and of each of the interventions it examines, is to assess whether changes in NSSE/CLASSE scores and other selected academic outcomes can be detected and identified as being results of the interventions. As such, it is the fit between NSSE, CLASSE and academic performance measures and the structure and scale of the interventions that is being assessed. For example, a finding that small intervention scale is detrimental to measuring engagement effects, or that successive cohort designs are more (or less) problematic than cross-sectional designs, is intended to address the applicability of NSSE in assessing the types of interventions undertaken, and is not intended as commentary on the design or effectiveness of the interventions themselves.

Further details on the interventions are available from the contacts listed in the final section of this report. The interventions are discussed as a group with respect to selected qualitative aspects of assessment at the end of this section. As noted above, the results of some interviews, focus groups and follow-up student monitoring and analysis in some interventions are not available for this report, and given the importance of "local context" should be reported by those directly involved anyway.

4.2 Carleton University (Development of a TA Mentorship Model)

4.2.1 Intervention Description and Background

Student perceptions about the role of teaching assistants (TA's) in undergraduate instruction had previously been measured through Carleton's results on the Canadian University Survey Consortium (CUSC) survey, student responses to the Ontario consortium question results on NSSE 2006, and a series of focus groups with students. In response to higher-than-average dissatisfaction with teaching assistants, Carleton developed a pilot TA mentoring program intended to ensure that TA's had the skills and resources necessary to provide effective instructional support. Training was made available to newer and less experienced TA's by more experienced senior teaching assistants that consisted of several components:

- One-on-one advice with TA mentors;
- Teaching skills workshops;
- Individualized classroom technique evaluation;
- Access to resources on WebCT;
- Course in pedagogy communication skills (to those assessed as requiring it);
- Library training;

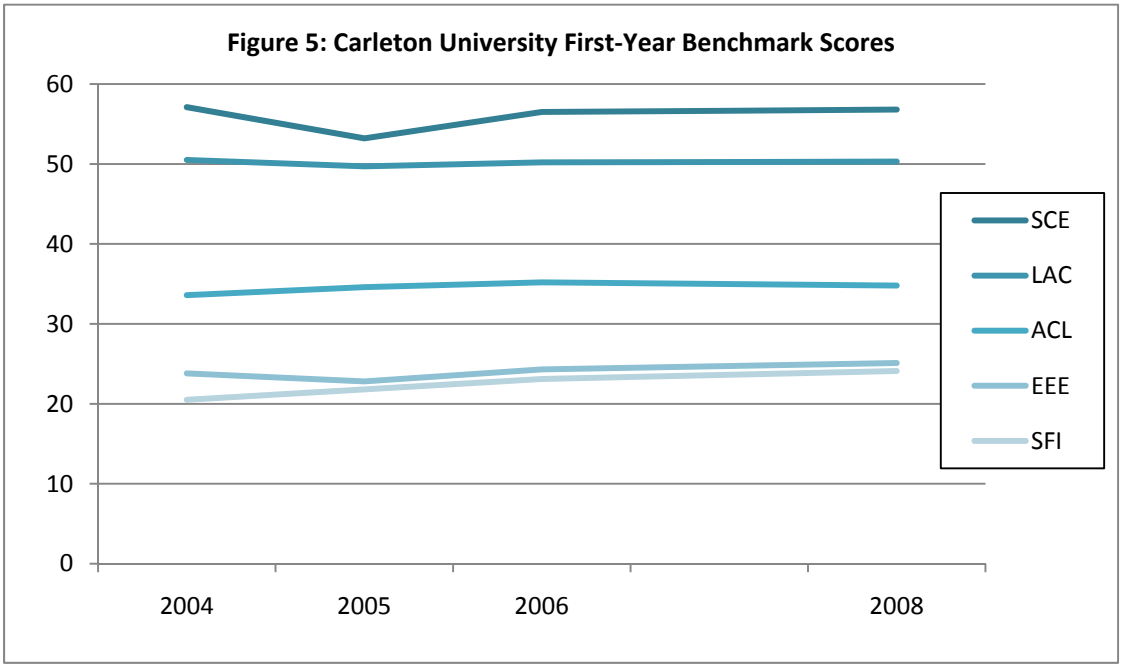
- Creation of department-level learning communities providing TA's with professional development opportunities.

A total of 124 TA's participated in the mentorship program (excluding mentors). A portion of the TA's were involved in the approximately 130 lecture/lab/seminar sections of the 27 introductory-level courses taken by first-year students in Business, Mathematics, Political Science, Psychology and Systems Computing. The majority of first-year students register in at least one course in each of these five academic programs. TA's were assigned a subjective "intensity of training" rating by their TA mentor based on exposure to the mentoring program components. The courses to which TA's were ultimately assigned, combined with the course registration profile of first-year students, permitted the construction of a "intensity of exposure" measure for each student (the number of courses involving mentored TA's in which the student was registered). The participation of teaching assistants in the program was voluntary (though strongly encouraged). There existed varying degrees of program uptake by TA's across departments, likely due to uneven levels of publicity of the program and differing "cultures" across departments. While TA self-selection bias undoubtedly exists, the objective of the program remains the incremental improvement in TA effectiveness measured across the students coming into contact with mentored TA's; this can be tested regardless of possible TA participation bias. Student self-selection is not an issue, as all students responding to NSSE could be assigned both an exposure status and an intensity of exposure measure based on their course registration profile.

Carleton had commenced the TA mentoring program prior to the start of this research project, and began implementation in the 2007/08 academic year. As a result, CLASSE was not available as a potential assessment tool. Carleton's objective in participating in this project was to subject the program to a more detailed assessment than might otherwise have been possible.

4.2.2 Context Provided by NSSE Administrations

Carleton administered NSSE in 2004, 2005, 2006 and 2008, and has also administered FSSE and BCSSE in previous years. The more recent administrations of NSSE have utilized a 100 per cent sampling rate, permitting Carleton to perform Faculty- and program-level drilldowns on its results. As noted above, the development and ongoing evaluation of the TA mentoring program were predicated on various NSSE, FSSE, BCSSE and focus group results. Results for Carleton's four NSSE administrations are quite stable across all five first-year benchmarks, and are almost constant for the critical 2006 and 2008 administrations. The Level of Academic Challenge (LAC) benchmark score increased by 0.2 per cent from 2006 to 2008; Active and Collaborative Learning (ACL) by -1.2 per cent; Student-Faculty Interaction (SFI) by 4.3 per cent; Enriching Educational Experiences (EEE) by 3.3 per cent and Supportive Campus Environment (SCE) by 0.5 per cent.



Carleton identified a number of NSSE items and benchmarks that reflected the goals of the TA mentoring program and that formed the basis for this assessment. About half of the NSSE items fall within the ACL, LAC and SCE benchmarks; the remainder lies outside the five benchmarks. With respect to the individual NSSE items against which the intervention was assessed, only two of the 12 item means changed by more than a small amount between 2006 and 2008 (-5.1 per cent and +2.5 per cent); the other 10 changed by 1.3 per cent or less. Responses to the Ontario consortium question of key interest (student identification of TA's as a key issue for the university to address in the classroom) declined from 35 per cent in 2006 to 28 per cent in 2008 (across all first-year responses). In addition, overall academic performance (grade average) was identified as an experimental outcome.

Table 3: Carleton University Results for Key NSSE Items 2006 and 2008			
	Mean Item Scores		
Core NSSE Items	**2006**	**2008**	**% Change**
Asked questions … contributed to class discussions	2.35	2.23	-5.11%
Came to class unprepared	2.30	2.28	-0.87%
Worked with classmates on assignments outside class	2.39	2.45	2.51%
Coursework emphasis on applying theories or concepts	3.01	2.98	-1.00%
Examined the strengths/weaknesses of your own views	2.48	2.48	0.00%
Tried to understand another view	2.72	2.75	1.10%
Learned something that changed your view/understanding	2.84	2.83	-0.35%
Quality of relationships with faculty members	4.85	4.85	0.00%
Institutional emphasis on spending time on academics	3.03	3.06	0.99%
Institutional emphasis on supporting you academically	2.96	2.98	0.68%
Evaluate entire educational experience at institution	3.05	3.09	1.31%
If starting over, would you attend same institution	3.20	3.20	0.00%
Ontario Consortium Item (student selection of item as one of up to 2 selections permitted from a list of 10)	**% Checking Response**		
	2006	**2008**	**% Change**
Institution needs to address issue of teaching assistants	35%	28%	-20%

4.2.3 Assessment Design

The intervention assessment employs both post-measure cross-sectional and post-measure successive cohort designs. A 100 per cent first-year NSSE sample was surveyed in Spring 2006 prior to the implementation of the TA mentoring program, and again in Spring 2008 at the end of the program's first year of operation. The two administrations achieved response rates of 45 per cent and 49 per cent. The 2008 NSSE response records were merged with (a) course registration and TA course assignment information to identify candidate control and experimental group populations and permit construction an intensity of exposure measure, and (b) demographic and academic data from the student records system to facilitate propensity matching. The 2006 NSSE response records were merged with identical demographic and academic data only.

For the cross-sectional design, 2008 respondents were assigned to the control group (intensity of exposure = 0) or experimental group (intensity of exposure > 0). For the successive cohort design, all 2006 respondents were assigned to the control group. For both designs, propensity matching based on full-/part-time status, age, year of study, student type (direct vs. delayed from secondary school) and gender was undertaken to

construct control and experimental groups that did not differ on these matching variables (see Table 4). Post-match, sample sizes were 1102 for each of the control and experimental groups (in the cross-sectional design), and 1754 for each group (in the successive cohort design).

Table 4: Carleton University Propensity Matching Results

		Pre-Match				Post-Match			
Design	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Sq	p-value
Cross-Sectional	Part-time/Full-time	0.56	0.13	18.17	0.000	0.00	0.18	0.00	1.000
	Age	-0.06	0.01	44.02	0.000	-0.01	0.01	0.26	0.609
	Direct/Non-Direct Entry	-1.76	0.08	467.87	0.000	0.00	0.10	0.00	1.000
	Year of Study	-0.89	0.07	171.73	0.000	0.00	0.09	0.00	1.000
	Gender	0.04	0.07	0.35	0.550	0.00	0.09	0.00	1.000
		(n=2058 experimental, n=1716 control)				(n=1102 experimental, n=1102 control)			
Successive Cohort	Part-time/Full-time	-0.20	0.12	2.60	0.107	0.00	0.16	0.00	1.0000
	Age	0.04	0.01	23.06	0.000	0.00	0.01	0.00	0.9830
	Direct/Non-Direct Entry	1.12	0.06	344.91	0.000	0.00	0.07	0.00	1.0000
	Year of Study	0.26	0.06	19.78	0.000	0.00	0.08	0.00	1.0000
	Gender	-0.01	0.06	0.04	0.840	0.00	0.07	0.00	1.0000
		(n=2058 experimental, n=3149 control)				(n=1754 experimental, n=1754 control)			

4.2.4 Assessment Results

Table 5 presents the results of the initial series of bivariate regressions for both designs, using a basic participation/non-participation measure (i.e., without considering intensity of student involvement).

Table 5: Carleton University Regression Results Round 1										
	Cross-Sectional Design					Successive Cohort Design				
Dependent Variable	R^2	B-Est	SE	t-score	p-value	R^2	B-Est	SE	t-score	p-value
Asked questions/contributed to class discussions	0.003	0.098	0.037	2.64	0.008	0.003	-0.098	0.029	-3.42	<.0001
Came to class unprepared (reversed response scale)	0.000	-0.018	0.035	-0.52	0.601	0.001	-0.048	0.029	-1.68	<.0001
Worked with classmates on assignments outside class	0.000	-0.024	0.040	-0.58	0.562	0.000	0.039	0.031	1.23	0.219
Coursework emphasis on applying theories or concepts	0.000	0.030	0.039	0.77	0.443	0.000	-0.012	0.031	-0.39	0.693
Examined the strengths/weaknesses of your own views	0.000	0.032	0.039	0.82	0.415	0.000	0.016	0.030	0.52	0.601
Tried to understand another view	0.000	-0.021	0.037	-0.56	0.574	0.000	0.030	0.030	1.00	0.313
Learned something that changed view/understanding	0.000	0.034	0.035	0.95	0.340	0.000	-0.011	0.028	-0.37	0.712
Quality of relationships with faculty members	0.000	0.045	0.062	0.73	0.464	0.000	0.022	0.049	0.45	0.653
Institutional emphasis on spending time on academics	0.000	0.025	0.033	0.74	0.458	0.000	0.017	0.026	0.65	0.514
Institutional emphasis on supporting you academically	0.000	0.026	0.036	0.72	0.469	0.000	0.034	0.029	1.16	0.247
Evaluate entire educational experience at institution	0.000	-0.006	0.032	-0.18	0.861	0.000	0.030	0.025	1.14	0.253
If starting over, would you attend same institution	0.000	0.023	0.036	0.64	0.524	0.000	0.029	0.027	1.03	0.303
Academic Challenge Benchmark	0.003	1.440	0.588	2.45	0.014	0.000	0.397	0.454	0.87	0.382
Institution needs to address issue of teaching assistants	0.000	-0.002	0.020	-0.10	0.921	0.006	0.072	0.016	4.42	<.0001
Overall grade average	0.003	0.734	0.443	1.66	0.098	0.000	-0.074	0.279	-0.26	0.792

The majority of the core NSSE engagement items show no experimental effect. The LAC benchmark – which contains the "coursework emphasis on applying theories" and "institutional emphasis on spending time on academics" items – was however, significant in the cross-sectional design, suggesting the possibility that other untested LAC items may also be significant. The cross-sectional design suggests the intervention had no measurable effect on perceptions toward teaching assistants, but does suggest a weak positive effect on student academic performance (see below).

A second round of bivariate regressions for the cross-sectional design was run specifically for the remaining LAC items; the results are shown in Table 6.

Table 6: Carleton University Regression Results Round 2					
	Cross-Sectional Design				
Dependent Variable (Remaining LAC Items)	R^2	B-Est	SE	t-score	p-value
Number of written papers 20+ pages	0.004	-0.109	0.040	-2.74	0.006
Coursework emphasis on analysis	0.002	0.062	0.034	1.87	0.062
Coursework emphasis on making judgments	0.003	0.094	0.039	2.39	0.017
Number of written papers 5 - 19 pages	0.006	0.146	0.042	3.50	0.001
Worked hard to meet expectations	0.000	0.019	0.037	0.48	0.629
Courework emphasis on synthesis	0.001	0.064	0.037	1.71	0.087
Number of assigned texts/readings	0.005	0.143	0.043	3.30	0.001
Number of written papers < 5 pages	0.000	0.023	0.048	0.47	0.639

Four of the eight LAC items modeled demonstrate a significant experimental effect (at the .05 level); one of these four – writing papers of 20+ pages – carries a negative coefficient (which cannot be considered intrinsically undesirable to the extent that the coefficient for medium-length papers increased).

Given the relatively weak regression results (low R^2 and few significant variables) and a preliminary analysis of intensity of exposure variation that yielded inconclusive results, a somewhat more qualitative approach to analyzing intensity of exposure was taken. Students registered in varying numbers of courses involving mentored TA's: 0 (the control group), one semester course equivalent, two semester course equivalents, etc., through to seven semester course equivalents (although insufficient numbers of students were registered in six or seven courses to permit analysis). The number of students at each exposure level permitted an examination of engagement responses for exposure intensity ranging from 0 to 5 semester course equivalents, as shown in Figures 6(a) and (b) for the cross-sectional design and for the original 15 dependent variables. (The significant item difference and benchmark difference shown in Table 5 are shown as dotted lines in Figures 6(a) and (b)). Virtually identical results were obtained for the successive cohort design. As measured, intensity of exposure to courses involving mentored TA's had no effect on either NSSE item responses or grade averages.

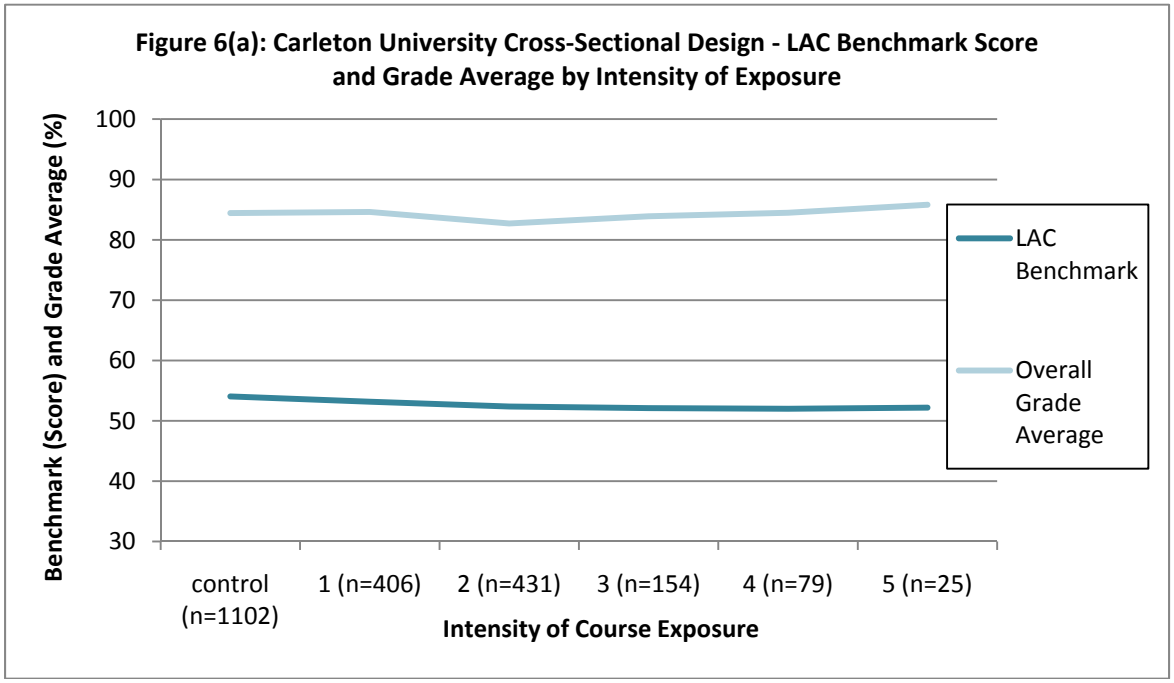
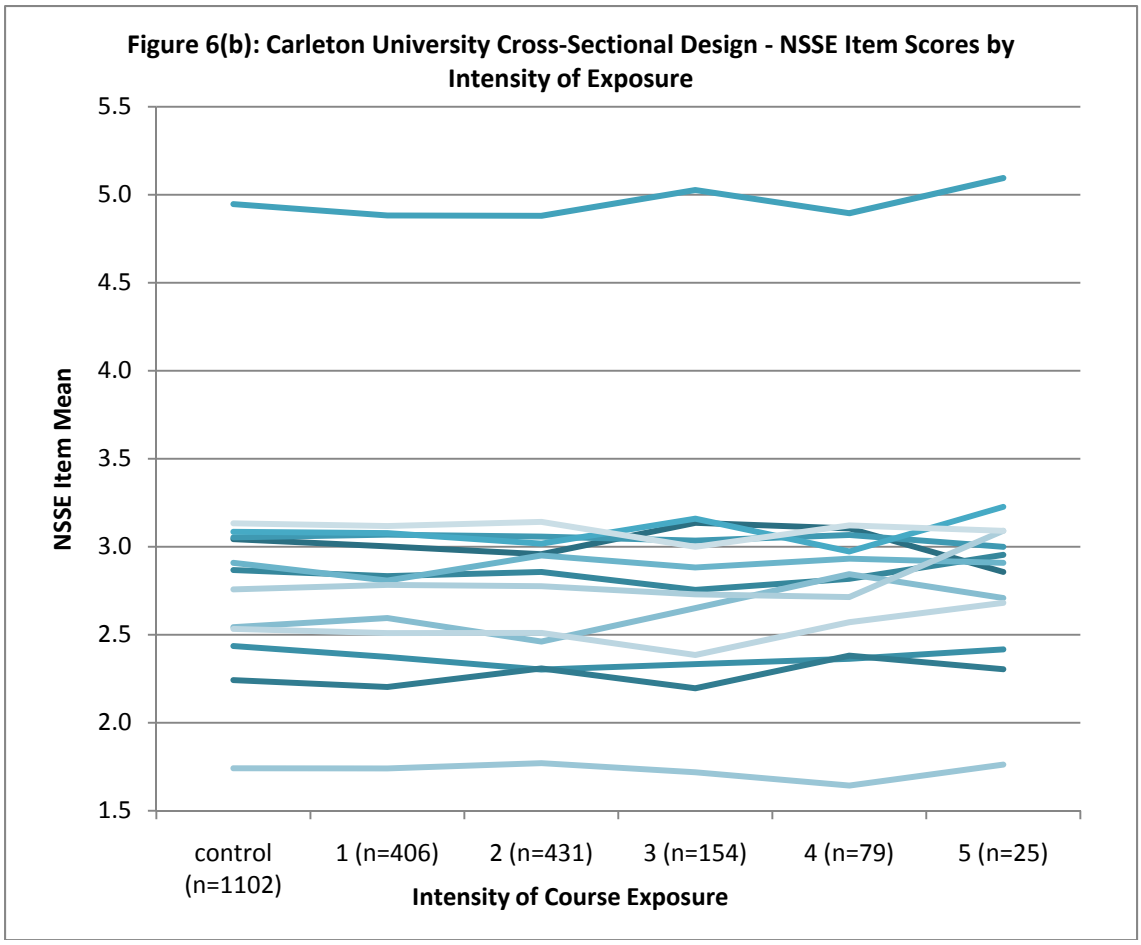


Figure 6(a): Carleton University Cross-Sectional Design - LAC Benchmark Score and Grade Average by Intensity of Exposure
Benchmark (Score) and Grade Average (%)
100
90
80
70
60
50
40
30
control (n=1102)
1 (n=406)
2 (n=431)
3 (n=154)
4 (n=79)
5 (n=25)
Intensity of Course Exposure
LAC Benchmark
Overall Grade Average



In addition to core NSSE item impacts, Carleton's project was also predicated on student perceptions toward TA's, defined for this assessment as the proportion of students identifying TA's as one of the classroom issues requiring attention (from the 2006 and 2008 Ontario NSSE consortium question set). As noted above, concern over TA's declined from 35 per cent in 2006 to 28 per cent in 2008 across the entire first-year NSSE respondent population. But do the data indicate that the decline was associated disproportionately with students who participated in the TA mentoring program? Table 7 suggests not: the decline was experienced by both participants and non-participants in the TA mentoring program. It appears that exogenous factors have contributed to the relative decline in the importance of TA's. Other issues may have become more important between 2006 and 2008, or non-experiment related events might have had effects on student perceptions.

Table 7: Carleton University Percentage of Students Identifying TA's as an Issue That Needs to be Addressed 2006 and 2008			
Year	**Gross NSSE Response**	**Control Group**	**Experimental Group**
2006	35%	33%	n/a
2008	28%	26%	26%

4.2.5 Summary

The propensity matching process was successful in generating matched pairs from two pools that differed significantly pre-match in both assessment designs. About 64 per cent of the candidate experimental pool could be utilized in the matched group regressions, leaving sufficiently large “n” for analysis. It is possible that the matched groups differ with respect to overall course registration profile, and that this difference is a contributor to weak regression results.

The majority of students in the experimental group were exposed to only one or two semester-equivalent courses, and less than 10 per cent to four or more courses. As a (multiple) course-based intervention, the Carleton project might benefit from a CLASSE-based assessment that could address possible dilution effects associated with the use of NSSE (as suggested by the very low explanatory power of the regressions). The use of CLASSE could also generate two other benefits: it could control for unknown differences in course registration behaviour between control and experimental groups, and it would permit development of alternate and more detailed intensity of involvement/exposure measures.

Input from Carleton project staff following the intervention suggests that lower-than-expected level of interaction between TA’s and mentors, and inconsistent promotion of the mentoring program across academic units, may have limited the overall impact of the program, and hence of the absence of NSSE-measured impact.

Across the two available NSSE data points, student perceptions toward TA’s show significant change, while core NSSE items show a high level of stability. Additional investigation may reveal the explanation for the sudden change in a multi-year history of consistent student perceptions. It has been suggested by Carleton staff that the recent institution-wide focus on improving student satisfaction with TA’s may have generated changes (e.g., faculty mentoring behaviour, communication with TA’s regarding their responsibilities) that were felt across the board – that is, in both control and experimental groups. At the present time, and given the very limited number of significant experimental effects, it is not possible to assess the relative merits of the cross-sectional design (which has limitations but which was absolutely necessary given significant differences pre-match) and the successive cohort design (which must accommodate “confusing” background noise with respect to student perceptions toward TA’s). Notwithstanding the

limited explanatory power of the regressions and possible limitations in the propensity matching process, the results of the cross-sectional analysis do indicate that a limited number of LAC benchmark items, and the benchmark score itself, differ between control and experimental groups. Absent a preponderance of evidence or more conclusive follow-up research, caution is warranted in attributing these differences to the intervention.

4.3 University of Guelph (Supported Learning Groups in High Risk Courses)

4.3.1 Intervention Description and Background

Supported Learning Groups (SLG's) at the University of Guelph provide enhanced academic support to students through additional voluntary study/assistance sessions. The initiative is based on the University of Missouri – Kansas City "Supplemental Instruction Model". Guelph's SLG program commenced in 1998 and has four objectives: to increase course and degree program retention; improve academic competencies; develop group and self-reliant learning skills; and foster a passion for learning and intellectual interchange. SLG's are run as weekly review sessions led by accomplished senior students who attend the course lectures and design the SLG sessions to allow students to compare notes, discuss course concepts, develop study strategies and self-test. SLG's are currently offered to Guelph students in 20 first-, second- and third-year high risk courses – generally those in which drops, failures and "D" grades constitute 30 per cent or more of initial course enrolments. The SLG-enhanced courses cover numerous disciplines, including Business, Chemistry, Economics, Physics, Psychology and Statistics. Previous evaluations of the program within the University have suggested that SLG's improve both course completion rates and course grade distributions. These positive results have contributed to the incorporation of SLG's into Guelph's Multi-Year Agreements with the Ministry of Training, Colleges and Universities.

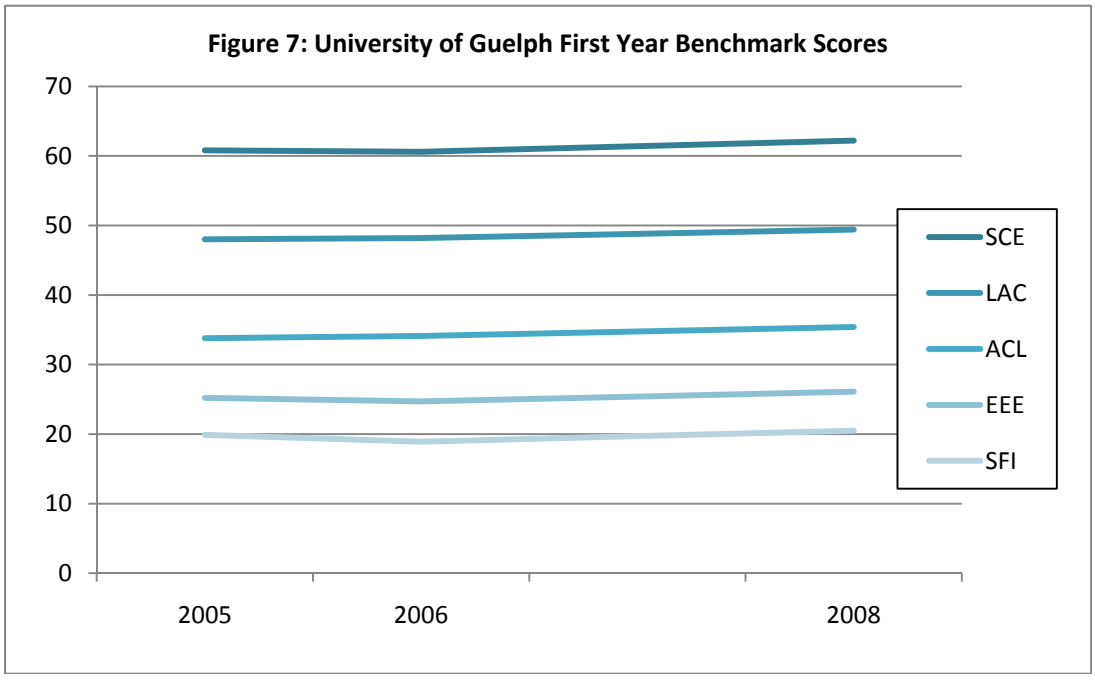
The University of Guelph administered NSSE in 2005, 2006 and 2008; BCSSE in 2006; and FSSE in 2007. In addition, it has conducted follow-up focus groups with NSSE respondents to assess cognitive response variation, it has constructed Pike scalets for Guelph's NSSE data, and it has utilized NSSE results in a review of the undergraduate science curriculum.

Guelph's objective in participating in this study was to determine whether the apparently positive pass rate and course grade impacts of SLG's translated into corresponding improvements in engagement measures at the course level. Six first-year courses with SLG's were selected for the analysis. Students in the SLG-supported courses were provided with information about the program, and self-selected for SLG participation.

4.3.2 Context Provided by NSSE Administrations

Guelph's three NSSE administrations (2005, 2006, and 2008) generated highly consistent benchmark scores over time. The LAC, ACL and SCE first-year benchmarks were

virtually identical in 2005 and 2006, and increased by less than five per cent in 2008. The EEE and SFI first-year benchmarks rose 5.7 per cent and 8.5 per cent respectively from 2006 to 2008. With respect to the 42 individual items contained within the benchmark measures, only a handful experienced 2006 – 2008 mean changes of more than five per cent (making a class presentation, tutoring other students, involvement in community-based projects, and discussing career plans with faculty). These results suggest a fairly stable engagement background for the SLG assessment, in which very small annual engagement changes have been experienced consistently across a wide range of items. Given its cross-sectional design, the Guelph project is not heavily dependent on such consistency over time; however, prior results suggest the NSSE survey administers reliably at Guelph and that there exists a stable environment for administration of the modified CLASSE instrument.



4.3.3 Assessment Design

To control for potential measurement distortion caused by self-selection bias, the experimental group members (SLG participants) were propensity matched to control group members using gender, secondary school average, international/domestic status and Faculty of student enrolment, on a course-by-course basis (not across courses). The absence of international students in some of the smaller courses prevented matching on immigration status, but immigration status was generally insignificant pre-match and in the overall sample post-match, and is unlikely to create distortion in the limited number of cases where it could not be included in the matching process. For all six courses, all

significant match variable differences that existed pre-match were eliminated post-match (see Table 8); candidate control group records were numerous enough to permit 1:1 matching with little, if any, reduction in the size of the experimental groups.

The primary tools for assessing the Guelph project were surveys administered to both SLG participants and non-participants, before and after the SLG's were implemented. The assessment design is quite sophisticated. Intervention impacts were assessed using a cross-sectional experimental design employing within-group pre- and post-testing. All students in the SLG-supported courses were administered a start-of-course (pre-intervention) survey containing modified NSSE/BCSSE/CLASSE questions dealing primarily with course and behavioural expectations; and an end-of-course (post-intervention) survey containing topic-matched questions dealing primarily with actual behaviours and course experiences. The survey responses of those students who self-identified were supplemented with data from the student records system (gender, domestic/international status, secondary school graduation average, program of study (Science or Social Sciences), course grade and current semester grade average). Students who completed, and self-identified on, both the pre- and post-surveys were categorized as non-SLG (control) or SLG (experimental) and included in the data set for analysis. Varying levels of SLG session attendance (self-reported on the end-of-course survey) and the tracking of the number of SLG's in which each student participated provide an opportunity to also measure the effect of the intensity of participation within the SLG experimental group at the course level and for multiple SLG participation.

The pre- and post- surveys dealt with expected (start-of-course) and actual (end-of-course) experiences such as asking questions in class, contributing to class discussions, preparing two or more drafts of a paper or assignment, integrating ideas or information from various sources and across multiple courses, coming to class unprepared, working with other students inside and outside of class, discussing grades and ideas with an instructor outside class, writing clearly and effectively, and thinking critically and analytically. Additional questions developed at Guelph dealt with student reactions to adversity (e.g., lack of initiative, academic challenges and poor grades), final course grade and level of interest in the course material.

The CLASSE data set permits analysis at two levels of aggregation. Each course registration in one of the six courses constitutes one record, so students involved in more than one of the six courses (and more than one of the SLGs) occupy multiple records. This data structure supports a course-by-course SLG assessment, since within a given course there is only one record per student. Thus, the first design involves course-specific SLG assessment, including varying levels of SLG participation within each course.

For the second design, two data set formats were created. In the first, the six course-level data files were simply concatenated. Thus, the control and experimental groups remain matched on a within-course basis. This data file has the same applications as the individual course files, but permits the inclusion of the smaller courses in the analysis. Because the records continue to reflect specific course affiliations, mixed model regressions were used to account for non-independence (i.e., within-course homogeneity). The second data format was created to explore the effects of participation in multiple SLG's which involve a much wider range of session involvement across courses. Multiple records per student (representing participation in multiple SLG's) were converted to a

single student record by summing total SLG session attendance across the one or more courses in which the student was involved, and averaging the values of other variables (e.g., each of the multiple pre- and post-survey responses provided by students for each of the courses they were involved in). Demographic and program variables were constant across each student's multiple participation records and did not require adjustment. Experimental group records (those students participating in one or more SLG's) were propensity matched to control group records and all significant pre-match differences were eliminated. Because the records of participants in multiple SLG's were merged into one record, such records no longer reflect a distinct course affiliation and mixed model regression was not performed.

Table 8: University of Guelph Propensity Matching Results									
		Pre-Match				Post-Match			
Course	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Sq	p-value
1	Gender	0.95	0.84	1.28	0.258	0.00	1.10	0.00	1.000
	Secondary school average	-0.07	0.06	1.62	0.202	0.00	0.07	0.00	0.980
	International status	not applicable				not applicable			
	Scie/SocSci program	11.57	227.31	0.00	0.959	not applicable			
		(n=12 experimental, n=44 control)				(n=12 experimental, n=12 control)			
2	Gender	0.16	0.20	0.60	0.440	0.00	0.22	0.00	1.000
	Secondary school average	-0.03	0.02	1.78	0.182	0.00	0.02	0.06	0.812
	International status	-0.72	0.74	0.95	0.330	0.00	1.01	0.00	1.000
	Scie/SocSci program	0.45	0.33	1.87	0.171	0.00	0.41	0.00	1.000
		(n=189 experimental, n=250 control)				(n=166 experimental, n=166 control)			
3	Gender	0.32	0.62	0.27	0.602	0.00	78.00	0.00	1.000
	Secondary school average	-0.03	0.04	0.49	0.486	0.00	0.06	0.00	0.991
	International status	12.47	890.15	0.00	0.989	not applicable			
	Scie/SocSci program	-0.93	0.64	2.10	0.147	0.00	0.78	0.00	1.000
		(n=29 experimental, n=102 control)				(n=22 experimental, n=22 control)			
4	Gender	0.59	0.81	0.53	0.467	0.00	1.48	0.00	1.000
	Secondary school average	-0.10	0.06	2.73	0.098	-0.02	0.08	0.04	0.849
	International status	not applicable				not applicable			
	Scie/SocSci program	-0.04	0.83	0.00	0.957	0.00	1.48	0.00	1.000
		(n=12 experimental, n=101 control)				(n=11 experimental, n=11 control)			
5	Gender	0.36	0.21	3.14	0.076	0.00	0.24	0.00	1.000
	Secondary school average	0.01	0.02	0.29	0.588	0.00	0.02	0.02	0.900
	International status	-2.18	1.08	4.05	0.044	0.00	1.42	0.00	1.000
	Scie/SocSci program	-0.29	0.42	0.46	0.497	0.00	0.59	0.00	1.000
		(n=253 experimental, n=476 control)				(n=199 experimental, n=199 control)			
6	Gender	0.78	0.21	13.25	0.000	0.00	0.28	0.00	1.000
	Secondary school average	-0.02	0.02	1.85	0.174	0.00	0.02	0.00	1.000
	International status	0.11	0.82	0.02	0.898	not applicable			
	Scie/SocSci program	0.47	0.23	4.17	0.041	0.00	0.27	0.00	1.000
		(n=222 experimental, n=704 control)				(n=166 experimental, n=166 control)			
pooled	Gender	0.34	0.11	9.62	0.002	0.00	0.13	0.00	1.000
	Secondary school average	-0.01	0.01	0.39	0.532	0.00	0.01	0.00	1.000
	International status	-0.84	0.41	4.12	0.042	0.00	0.64	0.00	1.000
	Scie/SocSci program	0.16	0.10	2.42	0.120	0.00	0.12	0.00	1.000
		(n=616 experimental, n=1190 control)				(n=588 experimental, n=588 control)			

An additional analysis exercise is also possible given the Guelph assessment design and the availability of critical data. The existence of pre-SLG CLASSE survey responses for both SLG participants and non-participants in combination with student record-level demographic and academic data presents an opportunity to perform an independent analysis of the success and impact of propensity matching – that is, to measure the extent to which propensity matching based on demographic and academic characteristics accomplishes the actual goal of correcting for the predisposition differences associated with participant self-selection.

Course-level (not record-level) NSSE response data had been previously assembled by the University of Guelph staff in 2006 and 2008 for the six courses examined here and are discussed briefly below to highlight the value of targeting in reducing measurement dilution.

4.3.4 Assessment Results

Note on Course-Level 2006 and 2008 NSSE Responses:

Table 9 (prepared by the University of Guelph staff for demonstration rather than assessment purposes) displays mean NSSE item scores for different groups of students. Columns (A) and (C) represent the average item scores for all students registered in Course #1 (of the six courses examined below) in 2006 and 2008. The course offered an SLG in both years, but the scores do not differentiate between those who participated and those who did not, so any SLG impact is in effect distributed over participants and non-participants. Columns (B) and (D) represent the average item scores for all first-year students who were not registered in Course #1 (and therefore not involved in its SLG) in 2006 and 2008 respectively. The qualitative comparison column indicates positive or negative differences between students in/not in the SLG course greater than five per cent in each of the two years.

Table 9: University of Guelph Course-Based NSSE Response Summary (Course #1)					
	2006		2008		
NSSE Item	Mean Item Score for Students Registered in the SLG Course (A)	Mean Item Score for Students Not Registered in the SLG Course (B)	Mean Item Score for Students Registered in the SLG Course (C)	Mean Item Score for Students Not Registered in the SLG Course (D)	Qualitative Comparison A:B and C:D (differences > 5% shown)
Course 1					
Asked questions … participated in discussions	2.274	1.974	2.103	1.973	+ +
Prepared 2+ drafts of paper	2.434	2.314	2.423	2.194	+ +
Project required integrating ideas	3.377	2.937	3.237	2.943	+ +
Came unprepared to class	2.321	2.316	2.289	2.274	
Worked with students during class	1.635	1.802	1.711	1.801	-
Worked with classmates outside class	2.283	2.765	2.289	2.707	- -
Integrated ideas from different courses	2.779	2.557	2.819	2.65	+ +
Tutored or taught other students	1.413	1.593	1.447	1.748	- -
Discussed grades with instructor	1.952	1.814	2.085	1.971	+ +
Discussed readings with others outside class	2.913	2.833	2.871	2.916	
Discussed ideas with faculty outside class	1.500	1.450	1.447	1.491	

The results highlight several issues. First, the absence of targeting (i.e., student record-level identification of SLG participants and non-participants) prevents any impacts of the SLG from being associated clearly with participants alone, or with varying levels of participation. Second, the existence of positive differences of 5 per cent or more for some items across both years might suggest that on balance, courses with SLG's score higher than those without. Without knowing what other courses students registered in or what their personal or academic characteristics are, it is impossible to associate the differences with the SLG. Third, the use of NSSE measures (which apply to an entire academic year experience) may dilute the SLG effect within a far broader pool of academic experiences. As a result, any analysis based on this data would be at least inconclusive, if not inappropriate.

Propensity Matching Assessment:

The data provide an opportunity to (a) assess the propensity matching methodology by comparing the pre-match engagement characteristics of the control and experimental groups (i.e., those factors measured in the pre-intervention survey instrument) for which

gender and other variables are intended to serve as surrogates, and (b) determine the extent to which engagement-based self-selection bias exists within the SLG program. This analysis was undertaken through a course-by-course and item-by-item analysis, and more specifically, through significance testing of control vs. experimental pre-intervention (September) engagement differences in both pre-match and post-match samples. Twenty-eight questions on the pre-intervention instrument were identified as being indicative of engagement expectations and/or predisposition; across the three courses having sufficient post-match enrolment counts (i.e., after eliminating the courses with 12+12, 22+22 and 11+11 registrations in the two groups), this suggests 84 points of comparison between the pre-match and post-match samples. The analysis identified four types of items:

- SS – The item showed a significant difference between the control vs. experimental groups both before and after the match (i.e., self-selection or predisposition bias existed and was not corrected by the matching algorithm);
- SI – The item showed a significant difference between the control and experimental groups pre-match but no significant difference post-match (i.e., the matching had the intended effect of eliminating self-selection bias through the use of instrumental demographic variables);
- IS – The item showed no significant difference between control and experimental groups pre-match but the difference became significant post-match (i.e., the matching process introduced self-selection bias that did not exist in the unmatched samples;
- II – The item differences between control and experimental groups were insignificant both before and after the match (i.e., these – and only these – items specific to certain courses did not require matching for proper assessment).

Significance tests were run at the relatively conservative .10 level and yielded the following results:

- 17 SS items (20 per cent of the data points) disproportionately within Course #2;
- 10 SI items (12 per cent of the data points) disproportionately within Courses #4 and #5;
- 1 IS item (1 per cent of the data points);
- 56 II items (67 per cent of the data points).

The analysis permits several observations and conclusions about the impact of propensity matching:

- In 67 per cent of the tests, engagement predisposition bias was not a factor: the control and experimental groups were similar with respect to predisposition. In an additional 12 per cent of the tests, bias was corrected. In other words, “reasonably matched” samples pre-match (67 per cent of the tests) became better matched post-match (79 per cent) as a result of the propensity matching process;
- In only one case did matching generate bias where the statistical test indicated a low probability of bias existing before the match;
- Therefore, in the 27 cases where bias correction could have occurred, the propensity matching method was successful in eliminating it in 10 cases;

- Post-measure analysis of control vs. experimental group differences is robust only when pre-measures are similar or where they can be statistically controlled. For all "SS" and "IS" cases, engagement predisposition differences need to be accounted for in the assessment;
- Across each of the courses, the component items in each of the SS, SI, IS and II groups differed, and some courses contributed disproportionately to the total.

The demographic and academic matching variables appear to be reasonable surrogates for the engagement behaviours they are intended to mirror: post-match groups are clearly an improvement over pre-match. For the University of Guelph intervention, self-selection or predisposition differences remaining post-match can be accommodated through the selective use of pre-disposition controls. The propensity matching results at Guelph are intended as a general guide only: the importance of various match variables and a different rate of predisposition correction may well occur in other circumstances.

Intervention Impact on CLASSE Responses:

As noted above, students in the six courses offered in 2008/09 were administered a modified CLASSE survey instrument both pre- and post-intervention. Summary SLG participation and survey response is shown in Table 10 below. Both SLG participants and non-participants were asked to complete surveys at the beginning and end of their courses. About 14 per cent of all students in the six courses – almost 1,000 in total – self-selected for SLG participation. Response rates to both pre-SLG and post-SLG surveys varied from 45 per cent to 77 per cent among SLG participants, and from 32 per cent to 77 per cent among non-participants; SLG participant and non-participant dual survey response rates were similar in five of the six courses (within five per cent) and differed in the sixth course by 13 per cent.

Table 10: University of Guelph SLG Participation Summary

		SLG Participants				SLG Non-Participants			
Course	**Total Enrolment**	**Number**	**As % of Course Enrolment**	**Number Pre + Post Survey Respondents**	**SLG Participant Dual Survey Response Rate**	**Number**	**As % of Course Enrolment**	**Number Pre + Post Survey Respondents**	**SLG Non-Participant Survey Response Rate**
Course 1	292	19	7%	12	63%	73	25%	44	60%
Course 2	1359	267	20%	191	72%	371	27%	253	68%
Course 3	1170	64	5%	29	45%	320	27%	102	32%
Course 4	577	31	5%	17	55%	213	37%	106	50%
Course 5	1555	321	21%	254	79%	639	41%	478	75%
Course 6	2039	294	14%	227	77%	916	45%	706	77%
Total/Avg	6992	996	14%	730	73%	2532	36%	1689	67%

Given the results of the propensity matching exercise described above, simple regression analysis expressing post-intervention engagement as a function of SLG participation is generally sufficient. In those instances (e.g., for particular items and in particular courses) where significant predisposition differences remained post-match, pre-intervention engagement was included as a covariate in the models (and is referred to below as the "September covariate".

The results of the analysis are presented in Tables 11 to 14. The analysis of predisposition bias discussed above (i.e., remaining post-match item differences, course by course) identified those regressions warranting a September predisposition covariate. In these cases both the bivariate (experiment-only) result and the experimental result controlling for predisposition bias are shown. Table 11 provides the results for basic participation (but not intensity of participation) in the Course #1 SLG. These results are typical of those for all three small SLG's (with post-match SLG participation of 12, 22 and 11 in course #1, #2 and #4 respectively). Only one or two items in each course were significant at the .05 level; two or three of the items in each course indicated predisposition bias requiring a September covariate (and these were different for each

Table 11: University of Guelph Regression Results for Course #1

Dependent Variable		SLG Participation				September Covariate			
	R^2	B-Est	SE	t-score	p-value	B-Est	SE	t-score	p-value
1. So far this semester, how often have you done each of the following in this class? a) Asked questions during your class	0.017	-0.167	0.271	-0.62	0.544				
b) Contributed to a class discussion that occurred during your class	0.109	-0.667	0.407	-0.16	0.115				
	0.205	-0.628	0.416	-1.51	0.147	0.335	0.231	1.45	0.163
c) Prepared two or more drafts of a paper or assignment before turning it in	0.000	0.000	0.346	0.00	1.000				

d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.031	-0.167	0.198	-0.84	0.409					
e) Came to your class without having completed readings or assignments	0.008	-0.167	0.392	-0.42	0.675					
f) Worked with other students on projects during your class	0.083	-0.333	0.235	-1.14	0.171					
g) Worked with classmates outside of your class to prepare class assignments	0.038	-0.250	0.269	-0.93	0.364					
h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.003	-0.083	0.332	-0.25	0.804					
i) Tutored or taught other students in your class	0.004	0.083	0.274	0.30	0.764					
j) Discussed grades or assignments with the instructor of your class	0.004	-0.083	0.287	-0.29	0.775					
	0.004	-0.098	0.336	-0.29	0.774	-0.025	0.281	-0.09	0.929	
k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.009	0.167	0.381	0.44	0.666					

l) Discussed ideas from readings or classes with instructor outside of class	0.000	0.000	0.201	0.00	1.000	
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.224	-1.170	0.463	-2.52	**0.020**	
b) Attended a peer-led group study session (Supported Learning Groups (SLGs))	0.460	1.833	0.423	4.33	**<.0001**	
c) Sought resources & supports to improve understanding and preparation for tests	0.029	-0.580	0.717	-0.81	0.425	
d) Participated regularly in course discussions, even when you didn't feel like it	0.002	-0.167	0.730	-0.23	0.822	
e) Asked instructors for help when you struggled with course assignments	0.030	0.500	0.606	0.82	0.418	
f) Finished something you started when you encountered challenges	0.000	0.000	0.572	0.00	1.000	
g) Stayed positive, even when you did poorly on a test or assignment	0.000	0.000	0.442	0.00	1.000	
3. How would you rate the contribution of this class to your ability a) Writing cleary and effectively	0.002	-0.083	0.457	-0.18	0.857	
b) Speaking clearly and effectively	0.010	0.250	0.541	0.46	0.648	

	0.020	0.180	0.570	0.32	0.755	0.139	0.299	0.47	0.646
c) Thinking critically and analytically	0.038	-0.500	0.535	-0.93	0.361				
d) Analyzing math or quantitative problems	0.041	-0.500	0.515	-0.97	0.342				
e) Using computing and information technology	0.009	-0.083	0.590	-0.14	0.889				
f) Working effectively with others	0.033	-0.500	0.576	-0.87	0.395				
g) Learning effectively on your own	0.110	-0.750	0.454	-1.65	0.113				
4. During your time at the University, do you expect to change your degree program or major?	0.094	-0.500	0.331	-1.51	0.146				
6. How often have you spent more than 3 hours a week preparing for this class?	0.114	-0.583	0.347	-1.68	0.107				
Final Course Grade	0.000	-0.250	3.980	-0.06	0.950				

course); and R^2 values generally fall within the .000 to .050 range. Whether due to small sample size (which is likely) or other factors, the CLASSE measures show no consistent differences between SLG participants and non-participants in the three small courses.

Problems associated with small sample size were easily overcome in the analysis of the SLG's in the three large courses (Course #2, #5 and #6), with post-match SLG participation of 166, 166 and 199 students (and an equal number of control group

observations). The results for SLG participation (but not intensity) are presented in Tables 12 to 14. The three large SLG's reflect three different disciplines, different personnel, and presumably different learning objectives, course "dynamics" and SLG formats, so it is not surprising that differences exist in terms of significance findings. Across the three courses, items related to class participation (asking questions and participating in class discussions) and out-of-class activity (discussions with/seeking help from instructor, choosing studying over competing activities, and amount of time spent in class preparation) showed significant differences with coefficients having the "desired" sign. Several items that were significant in bivariate models became insignificant after controlling for predisposition bias, and others remained significant even after controlling for predisposition bias: a clear demonstration of the benefits of the pre- and post-measure assessment design. Final course grade did not show any difference between SLG participants and non-participants.

Table 12: University of Guelph Regression Results for Course #2

Dependent Variable	R^2	SLG Participation				September Covariate			
		B-Est	SE	t-score	p-value	B-Est	SE	t-score	p-value
1. So far this semester, how often have you done each of the following in this class? a) Asked questions during your class	0.002	1.321	0.052	25.56	**<.0001**				
b) Contributed to a class discussion that occurred during your class	0.001	1.500	0.064	23.34	**<.0001**				
c) Prepared two or more drafts of a paper or assignment before turning it in	0.009	0.137	0.079	1.74	0.830				
d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.007	0.137	0.092	1.49	0.136				
e) Came to your class without having completed readings or assignments	0.002	0.075	0.122	0.62	0.536				
f) Worked with other students on projects during your class	0.003	0.114	0.120	0.95	0.345				
	0.017	0.105	0.122	0.86	0.390	0.147	0.071	2.08	**0.038**
g) Worked with classmates outside of your class to prepare class assignments	0.012	0.242	0.122	1.98	**0.048**				
	0.037	0.171	0.123	1.39	0.166	0.217	0.071	3.10	**0.002**
h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.003	-0.096	0.101	-0.95	0.344				
i) Tutored or taught other students in your class	0.001	0.072	0.104	0.69	0.492				
j) Discussed grades or assignments with the instructor of your class	0.012	0.124	0.061	2.02	**0.044**				
	0.035	0.106	0.061	1.73	0.085	0.110	0.040	2.79	**0.006**

k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.006	0.136	0.093	1.45	0.145				
l) Discussed ideas from your readings or classes with your instructor outside of class	0.013	0.156	0.073	2.11	**0.036**				
	0.042	0.137	0.073	1.86	0.063	0.141	0.046	3.01	**0.003**
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.012	0.298	0.146	2.04	**0.043**				
c) Sought resources and supports to improve your understanding and preparation for tests and assignments	0.048	0.619	0.151	4.10	**<.0001**				
	0.070	0.556	0.153	3.63	**0.000**	0.188	0.071	2.64	**0.009**
d) Participated regularly in course discussions, even when you didn't feel like it	0.008	0.226	0.140	1.61	0.108				
	0.119	0.087	0.135	0.65	0.519	0.371	0.057	6.46	**<.0001**
e) Asked instructors for help when you struggled with course assignments	0.023	0.369	0.131	2.81	**0.005**				
f) Finished something you started when you encountered challenges	0.001	-0.065	0.138	-0.48	0.635				
g) Stayed positive, even when you did poorly on a test or assignment	0.000	-0.023	0.138	-0.21	0.830				
3. How would you rate the contribution of this class to your ability in a) Writing cleary and effectively	0.009	0.280	0.156	1.79	0.075				
b) Speaking clearly and effectively	0.005	0.208	0.153	1.36	0.175				
c) Thinking critically and analytically	0.007	-0.189	0.124	-1.52	0.130				
d) Analyzing math or quantitative problems	0.005	-0.179	0.133	-1.34	0.180				
	0.042	-0.113	0.133	-0.85	0.396	0.181	0.051	3.55	**0.000**
e) Using computing and information technology	0.006	0.238	0.163	1.46	0.145				
f) Working effectively with others	0.010	0.317	0.170	1.87	0.062				
g) Learning effectively on your own	0.003	0.135	0.132	1.02	0.310				
4. During your time at the University, do you expect to change your degree program or major?	0.002	-0.064	0.077	-0.82	0.413				
6. How often have you spent more than 3 hours a week preparing for this class?	0.039	0.321	0.062	36.59	**<.0001**				
	0.127	0.257	0.085	3.00	**0.003**	0.342	0.061	5.60	**<.0001**
Final Course Grade	0.002	-0.786	1.081	-0.73	0.468				

With respect to SLG participation alone (i.e., without considering intensity of involvement) it is clear that the CLASSE instrument was able to measure selected differences between SLG participants and non participants in large courses.

Students reported on the intensity of their SLG involvement in their post-SLG (November) response as attending 1-3, 4-6, or 7+ (of the 12 total) SLG sessions. An initial analysis of the three large courses was undertaken; the results for Course #2 are shown in Figure 8. The significant items for basic participation in Table 12 above are indicated with a dotted line over varying levels of involvement in Figure 8. The number of students at each involvement level declines: this is captured in the regression models above but not in Figure 8. There appears to be a weak general tendency for students at the higher involvement levels to report higher engagement, but the results for some items are erratic. This involvement-engagement pattern was less pronounced in the other two large courses. The three large courses were each subject to regression analyses predicting item engagement scores as a function of intensity of SLG involvement (i.e., the three involvement levels noted above) and while there existed selected and slight engagement score increases at higher involvement levels, the results were largely inconclusive.

Table 13: University of Guelph Regression Results for Course #5									
Dependent Variable		**SLG Participation**				**September Covariate**			
	R^2	**B-Est**	**SE**	**t-score**	**p-value**	**B-Est**	**SE**	**t-score**	**p-value**
1. So far this semester, how often have you done each of the following in this class? a) Asked questions during your class	0.000	0.021	0.049	0.43	0.669				
b) Contributed to a class discussion that occurred during your class	0.000	-0.010	0.056	-0.16	0.859				
c) Prepared two or more drafts of a paper or assignment before turning it in	0.000	-0.002	0.073	-0.03	0.976				
d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.000 0.034	-0.004 -0.078	0.085 0.086	-0.05 -0.90	0.962 0.367	0.186	0.051	3.62	**0.000**
e) Came to your class without having completed readings or assignments	0.001	-0.084	0.111	-0.75	0.451				
f) Worked with other students on projects during your class	0.000	0.046	0.113	0.41	0.681				
g) Worked with classmates outside of your class to prepare class assignments	0.003	0.113	0.110	1.03	0.304				
h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.003	-0.093	0.092	-1.00	0.317				
i) Tutored or taught other students in your class	0.000	0.017	0.100	0.17	0.866				
j) Discussed grades or assignments with the instructor of your class	0.004 0.058	0.074 0.038	0.059 0.058	1.25 0.65	0.213 0.513	0.194	0.041	4.72	**<.0001**
k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.000 0.053	0.014 -0.012	0.097 0.094	0.15 -0.12	0.881 0.897	0.257	0.054	4.68	**<.0001**
l) Discussed ideas from your readings or classes with your instructor outside of class	0.005 0.044	0.086 0.055	0.060 0.060	1.43 0.92	0.154 0.360	0.154	0.039	3.95	**<.0001**
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.006	0.211	0.142	1.49	0.137				
c) Sought resources and supports to improve your understanding and preparation for tests and assignments	0.065	0.768	0.147	5.22	**<.0001**				
d) Participated regularly in course discussions, even when you didn't feel like it	0.009 0.045	0.189 0.133	0.099 0.098	1.91 1.36	0.056 0.174	0.161	0.042	3.84	**0.000**
e) Asked instructors for help when you struggled with course assignments	0.043	0.485	0.115	4.23	**<.0001**				
f) Finished something you started when you encountered challenges	0.001	0.069	0.129	0.54	0.593				
g) Stayed positive, even when you did poorly on a test or assignment	0.000	-0.035	0.138	-0.25	0.802				
3. How would you rate the contribution of this class to your ability in a) Writing cleary and effectively	0.003	0.166	0.156	1.06	0.288				
b) Speaking clearly and effectively	0.003	0.182	0.156	1.17	0.244				
c) Thinking critically and analytically	0.000	-0.044	0.123	-0.35	0.723				
d) Analyzing math or quantitative problems	0.006	-0.186	0.121	-1.53	0.126				

e) Using computing and information technology	0.000	-0.033	0.152	-0.22	0.828	
f) Working effectively with others	0.016	0.414	0.163	2.53	**0.012**	
g) Learning effectively on your own	0.008	-0.209	0.121	-1.73	0.084	
4. During your time at the University, do you expect to change your program or major?	0.001	-0.052	0.073	-0.70	0.482	
6. How often have you spent more than 3 hours a week preparing for this class?	0.022	0.261	0.087	2.99	**0.003**	
Final Course Grade	0.001	-1.230	1.680	-0.73	0.466	

The pooling of SLG participants described earlier in this section permits an analysis of the SLG program across all six of the courses; it provides an opportunity to incorporate the results from the SLG's in the three small courses; and it allows a potentially more powerful intensity of involvement measure (i.e., total SLG session participation across multiple SLG's per student). The analysis is presented in Tables 15 and 16. Table 15 presents the results of models that predict engagement outcomes as a function of within-course participation intensity (i.e., the number of SLG sessions attended) based on mixed model regression utilizing the first of the two pooled data structures noted above. Table 16 presents predicted engagement outcomes as a function of the total number of SLG sessions attended across all the six courses and relies on Ordinary Least Squares (OLS) regression and the second of the pooled data formats. (In Table 16, the grouped "sessions per course" measure required minor estimation when summed across multiple courses because it established a range (rather than a single value) of sessions attended in each course. The coefficients in Table 16 apply to the sum of the coded grouped session attendance totals not the actual totals.) Predisposition bias was assessed on an item-by-item basis and included as a control where appropriate.

Table 14: University of Guelph Regression Results for Course #6

Dependent Variable	R^2	SLG Participation				September Covariate			
		B-Est	SE	t-score	p-value	B-Est	SE	t-score	p-value
1. So far this semester, how often have you done each of the following in this class? a) Asked questions during your class	0.003	0.074	0.069	1.070	0.284				
b) Contributed to a class discussion that occurred during your class	0.012	0.153	0.076	2.010	**0.045**				
c) Prepared two or more drafts of a paper or assignment before turning it in	0.003	0.084	0.093	0.900	0.369				
d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.001	0.047	0.100	0.470	0.641				
e) Came to your class without having completed readings or assignments	0.004	-0.132	0.114	-1.160	0.245				
f) Worked with other students on projects during your class	0.000	0.000	0.134	0.000	0.999				
g) Worked with classmates outside of your class to prepare class assignments	0.003	0.122	0.114	1.060	0.290				

h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.014	0.223	0.106	2.110	**0.036**				
i) Tutored or taught other students in your class	0.000	0.030	0.114	0.270	0.790				
j) Discussed grades or assignments with the instructor of your class	0.000	-0.004	0.050	-0.060	0.955				
k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.001	0.046	0.111	0.410	0.679				
l) Discussed ideas from your readings or classes with your instructor outside of class	0.012	0.149	0.074	2.010	**0.046**				
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.023	0.424	0.153	2.770	**0.006**				
c) Sought resources and supports to improve your understanding and preparation for tests and assignments	0.021	0.414	0.157	2.630	**0.009**				
d) Participated regularly in course discussions, even when you didn't feel like it	0.009	0.228	0.131	1.750	0.082				
e) Asked instructors for help when you struggled with course assignments	0.026	0.431	0.147	2.930	**0.004**				
f) Finished something you started when you encountered challenges	0.001	-0.036	0.146	-0.240	0.807				
g) Stayed positive, even when you did poorly on a test or assignment	0.004 0.131	-0.166 0.033	0.153 0.146	-1.090 0.220	0.278 0.824	0.437	0.064	6.910	**<.0001**
3. How would you rate the contribution of this class to your ability a) Writing cleary and effectively	0.003	-0.163	0.178	-0.920	0.360				
b) Speaking clearly and effectively	0.000	0.004	0.163	0.030	0.979				
c) Thinking critically and analytically	0.000 0.023	-0.036 0.018	0.142 0.142	-0.260 0.130	0.798 0.898	0.220	0.080	2.760	**0.006**
d) Analyzing math or quantitative problems	0.003	-0.127	0.137	-0.930	0.354				
e) Using computing and information technology	0.000	-0.006	0.168	-0.040	0.971				
f) Working effectively with others	0.005	0.195	0.157	1.240	0.217				
g) Learning effectively on your own	0.003 0.159	0.132 0.349	0.144 0.136	0.910 2.560	0.361 **0.011**	0.574	0.074	7.760	**<.0001**
4. During your time at the University, do you expect to change your program or major?	0.006	0.112	0.079	1.420	0.157				
6. How often have you spent more than 3 hours a week preparing for this class?	0.030	0.313	0.097	3.210	**0.002**				
Final Course Grade	0.002	1.030	1.418	0.730	0.468				

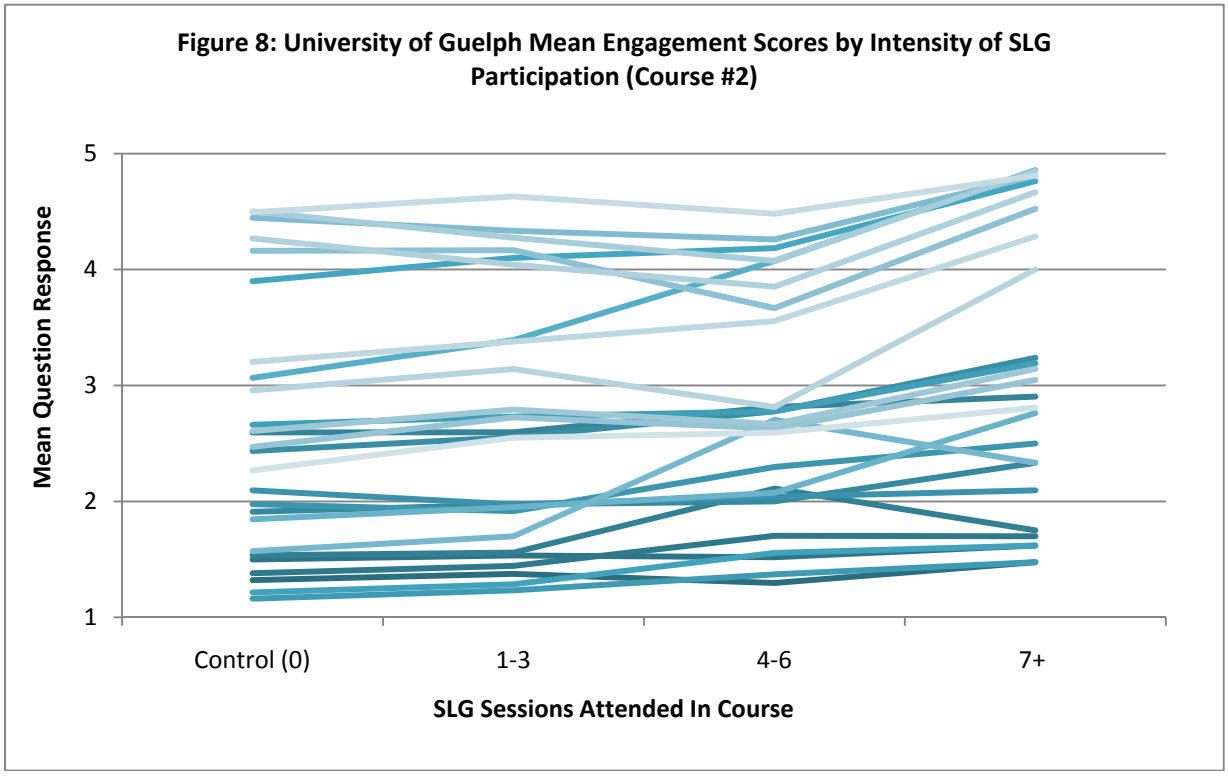


Table 15: University of Guelph Regression Results for All Courses Combined (Within-Course Intensity Measure)

Dependent Variable	Model		Participation Intensity			
	P-CCP	Intercept	B-Est	SE	t-score	p-value
1. So far this semester, how often have you done … in this class? a) Asked questions during your class	0.198	0.865	0.015	0.013	1.13	0.259
b) Contributed to a class discussion that occurred during your class	0.131	0.844	0.012	0.016	0.76	0.447
c) Prepared two or more drafts of a paper … before turning it in	0.082	1.354	0.032	0.018	1.77	0.078
d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.074	1.530	0.020	0.020	0.99	0.321
e) Came to your class without completing readings or assignments	0.161	1.970	-0.014	0.025	-0.56	0.572

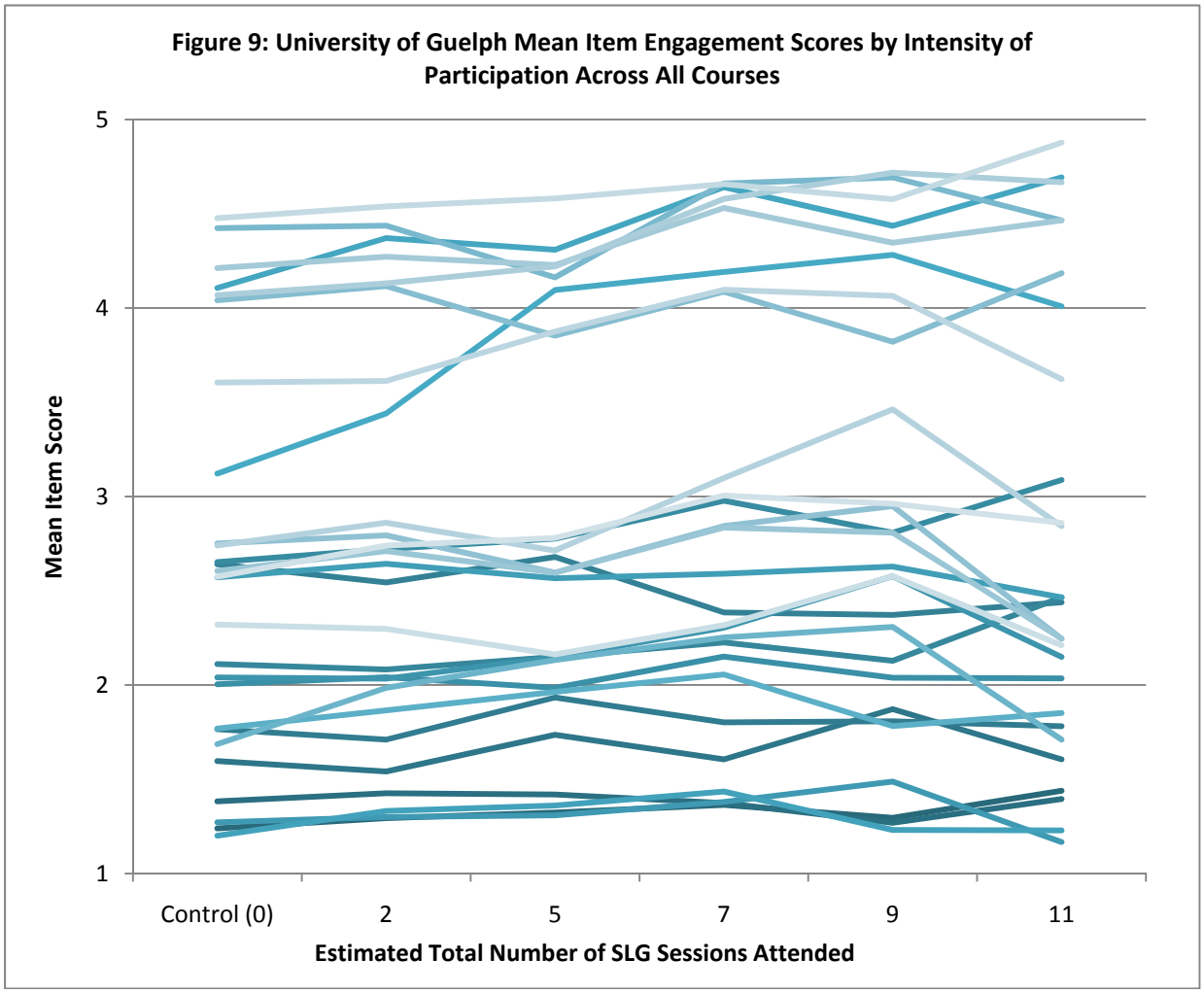
f) Worked with other students on projects during your class	0.129	1.291	0.030	0.026	1.14	0.255
g) Worked with classmates outside of your class on assignments	0.079	1.446	0.060	0.024	2.46	**0.014**
h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.163	1.532	0.012	0.022	0.57	0.566
i) Tutored or taught other students in your class	0.090	0.991	0.034	0.021	1.61	0.108
j) Discussed grades or assignments with the instructor of your class	0.216	0.873	0.020	0.014	1.45	0.149
k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.153	1.731	0.029	0.021	1.37	0.171
l) Discussed ideas … with your instructor outside of class	0.296	0.862	0.044	0.015	2.97	**0.003**
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.233	1.870	0.079	0.029	2.70	**0.007**
c) Sought resources and supports to improve your understanding of and preparation for tests and assignments	0.452	1.816	0.268	0.033	8.20	**<.0001**
d) Participated … in course discussions, even when you didn't feel like it	0.091	1.185	0.064	0.027	2.34	**0.019**
e) Asked instructors for help when you struggled with assignments	0.086	0.958	0.160	0.029	5.60	**<.0001**
f) Finished something you started when you encountered challenges	0.274	2.497	0.010	0.029	0.35	0.726
g) Stayed positive, even when you did poorly on a test or assignment	0.282	2.497	-0.007	0.029	-0.24	0.814
3. How would you rate the contribution of this class to your ability in a) Writing cleary and effectively	0.126	2.316	0.035	0.035	1.00	0.316
b) Speaking clearly and effectively	0.120	2.000	0.044	0.034	1.29	0.198
c) Thinking critically and analytically	0.444	3.129	0.003	0.028	0.12	0.907
d) Analyzing math or quantitative problems	0.060	2.527	-0.025	0.028	-0.88	0.381
e) Using computing and information technology	0.094	1.582	0.077	0.034	2.23	**0.026**
f) Working effectively with others	0.106	1.514	0.120	0.034	3.48	**0.001**
g) Learning effectively on your own	--	1.405	0.020	0.028	0.74	0.462
4. During your time at the University, do you expect to change your degree program or major?	0.317	0.945	-0.019	0.014	-1.30	0.194
6. How often have you spent more than 3 hours a week preparing for this class?	0.073	1.337	0.096	0.019	5.00	**<.0001**
P-CCP = prob-value on class (course) covariance parameter						

Tables 15 and 16 show that engagement outcomes for the pooled data analysis share some similarities with the individual large courses (because of the weight of those courses within the pooled file) but also display differences because of the distinct character of each of the three large courses being "averaged" in the pooled analysis. Engagement outcomes observed in both the individual files (for basic participation) and the pooled file (for intensity of involvement) include commitment to studying (amount of study and studying despite distractions), seeking resources/support and asking for help when needed. Intensity of involvement effects were more numerous across courses than within courses, suggesting that multiple SLG involvement offers additional benefit to students. Engagement scores across multiple SLG participation levels are also shown graphically in Figure 9 (in which dotted lines signify significant engagement outcomes corresponding to the Table 16 analysis).

Table 16: University of Guelph Regression Results for All Courses Combined (Across-Course Intensity Measure)

Dependent Variable	R2	SLG Participation Intensity				September Covariate			
		B-Est	SE	t-score	p-value	B-Est	SE	t-score	p-value
1. So far this semester, how often have you done each of the following in this class? a) Asked questions during your class	0.007	0.028	0.010	2.80	**0.005**				
b) Contributed to a class discussion that occurred during your class	0.001	0.011	0.012	0.92	0.357				
c) Prepared two or more drafts of a paper or assignment before turning it in	0.001 0.035	0.013 0.009	0.013 0.013	1.01 0.65	0.315 0.513	 0.160	 0.025	 6.41	 **<.0001**
d) Worked on a paper or a project in your class that required integrating ideas or information from various sources	0.001	0.014	0.015	0.99	0.323				
e) Came to your class without having completed readings or assignments	0.006	-0.049	0.018	-2.64	**0.008**				
f) Worked with other students on projects during your class	0.002	0.029	0.018	1.51	0.131				
g) Worked with classmates outside of your class to prepare class assignments	0.009 0.085	0.059 0.042	0.018 0.018	3.20 2.33	**0.001** **0.020**	 0.353	 0.035	 9.87	 **<.0001**
h) Put together ideas or concepts from different courses when completing assigments or during class discussions in your class	0.002	0.023	0.015	1.48	0.139				
i) Tutored or taught other students in your class	0.009 0.147	0.054 0.044	0.017 0.016	3.23 2.85	**0.001** **0.004**	 0.433	 0.032	 13.72	 **<.0001**
j) Discussed grades or assignments with the instructor of your class	0.002	0.017	0.010	1.70	0.089				
k) Discussed ideas from your class with others outside of class (students, family members, coworkers, etc.)	0.000	0.010	0.016	0.63	0.528				
l) Discussed ideas from your readings or classes with your instructor outside of class	0.011 0.052	0.039 0.034	0.011 0.010	3.70 3.28	**0.000** **0.001**	 0.155	 0.022	 6.93	 **<.0001**
2. So far this semester, how often have you done the following? a) Studied when there were other interesting things to do	0.018	0.112	0.024	4.63	**<.0001**				

c) Sought resources and supports to improve your understanding and preparation for tests and assignments	0.070 0.137	0.225 0.196	0.024 0.023	9.35 8.39	**<.0001** **<.0001**				
d) Participated regularly in course discussions, even when you didn't feel like it	0.009 0.101	0.067 0.044	0.021 0.019	3.21 2.19	**0.001** **0.029**	0.298	0.027	10.94	**<.0001**
e) Asked instructors for help when you struggled with course assignments	0.025 0.061	0.117 0.106	0.021 0.021	5.51 5.08	**<.0001** **<.0001**	0.199	0.029	6.65	**<.0001**
f) Finished something you started when you encountered challenges	0.001	0.021	0.022	0.94	0.345				
g) Stayed positive, even when you did poorly on a test or assignment	0.000	0.001	0.023	0.06	0.952				
3. How would you rate the contribution of this class to your ability a) Writing clearly and effectively	0.000	0.008	0.026	0.29	0.770				
b) Speaking clearly and effectively	0.002	0.034	0.025	1.35	0.178				
c) Thinking critically and analytically	0.004 0.061	0.043 0.050	0.021 0.020	2.06 2.44	**0.040** **0.015**	0.309	0.037	8.44	**<.0001**
d) Analyzing math or quantitative problems	0.009 0.129	0.083 0.089	0.025 0.023	3.32 3.79	**0.001** **0.000**	0.396	0.031	12.65	**<.0001**
e) Using computing and information technology	0.004 0.054	0.054 0.063	0.026 0.025	2.08 2.49	**0.038** **0.013**	0.280	0.036	7.88	**<.0001**
f) Working effectively with others	0.006	0.067	0.026	2.58	**0.010**				
g) Learning effectively on your own	0.003 0.079	0.038 0.051	0.021 0.020	1.84 2.56	**0.066** **0.011**	0.336	0.034	9.82	**<.0001**
4. During your time at the University, do you expect to change your degree program or major?	0.000	-0.006	0.012	-0.46	0.644				
6. How often have you spent more than 3 hours a week preparing for this class?	0.022 0.145	0.077 0.059	0.015 0.014	5.08 4.12	**<.0001** **<.0001**	0.439	0.034	12.98	**<.0001**



4.3.5 Summary

Participant and non-participant samples were reasonably similar pre-match, and did not display the very large predisposition differences that might be expected of self-selected groups. The propensity matching process resulted in a closer match between samples; and while it did not correct for predisposition bias in all cases, the available data permitted any remaining bias to be accommodated in the analysis. As such, the pre-/post-measure design in combination with propensity matching appears to be a very useful assessment approach.

CLASSE was clearly a suitable measurement tool for the interventions: in the pooled analysis, nearly one-half of the measures were significant after controlling for predisposition bias where required, and coefficients had the expected/desired signs.

While basic participation effects were also detected, the addition of the intensity of involvement measure across all courses permits even firmer conclusions. Staff at the University of Guelph will have the insight and knowledge to associate different course-by-course results (at least for the three large courses) with actual SLG experiences in those courses. The three small courses were not amenable to this kind of analysis. Even in courses where the experimental effect was significant, R^2 values remain fairly low, confirming the existence of numerous non-experimental factors in course-based engagement.

The University of Guelph design permitted a depth of analysis not possible in several other intervention projects. Because of large control groups, the post-match utilization rate for experimental records was high, ranging from 75 per cent to 100 per cent at the course level and 95 per cent at the pooled level. Ideally, pre-experiment surveying would be done before student self-selection in order to ensure against "expectation bias". The pre-/post-measure design permitted an additional layer of predisposition bias control. By following an identical surveying approach across all courses, the design allowed for the pooling of data and an analysis of the SLG effort across all six courses (not just individual program efforts) and the construction of an overall program intensity of involvement measure that appears more powerful than the within-course intensity measure alone.

4.4 University of Ottawa (FSS+: Faculty of Social Sciences Integration Program)

4.4.1 Intervention Description and Background

FSS+ is a program developed at the University of Ottawa to assist in the integration of new students into the Faculty of Social Sciences. The university's enrolment increased 37 per cent from 23,000 in 1997 to over 35,000 in 2007; within the Faculty of Social Sciences specifically, enrolment grew 70 per cent over the same period. Ottawa's objective in establishing FSS+ was to support students in establishing relationships with peers, professors and university staff of the sort that might be expected to characterize a smaller university and/or Faculty. The program has its foundation in the University's "Vision 2010" strategic plan, and in one of its specific goals, "to strengthen our competitive edge by offering our students an excellent education". Ottawa links this goal with providing an education that goes well beyond academe, ensuring that the University is welcoming and pleasant, and promoting diversified means of learning. The Faculty of Social Sciences was selected for the program because although its graduation rate is comparable to the Ontario average, it is lower than that in other programs at the University.

The FSS+ program consists of a number of components:

- A flexible registration process (in person, on-line or by phone) with a dedicated academic advisor;
- A timetable that ensures at least three courses in common with other FSS+ participants;
- Weekly study groups supervised by specially-trained senior student mentors;

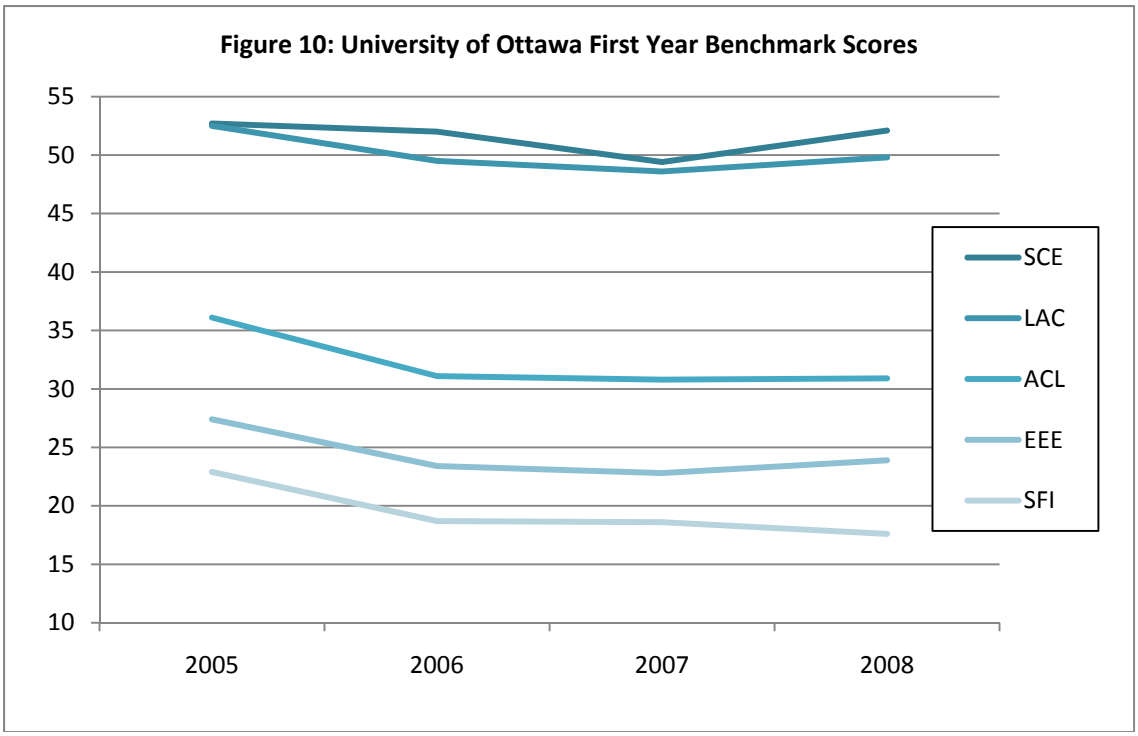
- A dedicated room for FSS+ student use for teamwork, studying, reading, etc.;
- Service orientation in the form of visits to the Library, Career Services, Writing Help Centre, etc.;
- Cultural events and outings for FSS+ participants, and invitations to academic events;
- A dedicated Student Experience Officer to coordinate the program and assist FSS+ students;
- A specialized website, Facebook network, and podcasts;
- An introductory course focusing on stimulating student interest and providing study skills in the Social Sciences.

Following the introduction of its Student Academic Success Service and the enrolment growth resulting from the "double cohort", a number of initiatives have been implemented over the past several years that directly or indirectly reflect Ottawa's NSSE results, including the early identification of at-risk students, improved classroom technology, a virtual mentorship program in engineering programs, a community service learning program, and opportunities for international study. The University has administered NSSE each year since 2005; FSSE has also previously been administered, and an analysis of NSSE/FSSE results was undertaken.

Students in the Faculty of Social Sciences were provided information about FSS+ at the start of the academic year and were invited to participate; participation was by self-selection.

4.4.2 Context Provided by NSSE Administrations

Ottawa's overall first-year NSSE results are presented in Figure 10. Four of the five benchmark scores declined from 2005 to 2006, by amounts ranging from six per cent to 22 per cent; only the SCE benchmark remained stable. Benchmark scores in 2006, 2007 and 2008 were much more stable: three increased – all by less than five per cent; one decreased by six per cent and one remained constant. These benchmark changes correspond to increases and decreases in several item scores, particularly the relatively volatile scores for participation in such EEE items as internships, learning communities and study abroad (these items are measured using a 0/1 scale and typically generate first-year means of .01 - .03 which are subject to considerable change upon the participation of only a few more or few less students).



Results for the individual items of interest to the Ottawa intervention are shown in Table 17. Over the past two administrations, the item means also show considerable rates of change – some in excess of 20 per cent - with the majority trending in a positive direction.

Benchmark and item score changes over time for the University overall provide a somewhat erratic background for the NSSE-based assessment of the 2007/08 FSS+ program in the spring of 2008. While there is no strong evidence the survey results are unreliable in the statistical sense, it does appear that the campus environment is experiencing change in excess of that observed at several other Ontario universities. The cross-sectional design employed (see below) controls for such change over time, but not for any point-in-time volatility that may exist.

Table 17: University of Ottawa First Year NSSE Item (Dependent Variable) Means			
NSSE Item	**2007**	**2008**	**% Change**
Asked questions in class/contributed to class discussions	2.16	2.09	-3.24%
Worked with classmates outside of class to prepare assignments	2.33	2.40	3.00%
Discussed ideas with others outside of class	2.74	2.80	2.19%
Participated in learning community	0.05	0.07	40.00%
Quality of relationships with other students	5.25	5.28	0.57%
Quality of relationships with faculty members	4.62	4.52	-2.16%
Quality of relationships with administrative staff	4.09	4.10	0.24%
Institutional emphasis on providing support for academic success	2.64	2.85	7.95%
Institutional emphasis on encouraging contact among different students	2.13	2.31	8.45%
Institutional emphasis on providing support to thrive socially	2.02	2.19	8.42%
Institutional emphasis on attending campus events and activities	2.45	2.63	7.35%
Institutional contribution to using computing and information technology	2.47	2.54	2.83%
Institutional contribution to working effectively with others	2.37	2.45	3.38%
Institutional contribution to encouraging voting in elections	1.86	2.29	23.12%

4.4.3 Assessment Design

The 2007/08 FSS+ program was assessed using a cross-sectional post-measure experimental design. Those students self-selecting for FSS+ participation were initially identified and to a limited degree, tracked during their involvement(s) with one or more components of the program. Because of the number of program components and the difficulty of monitoring usage across all components, an intensity of involvement measure was not constructed. Thus, the program design permits one definition of the experimental group (rather than several, each scaled to its intensity of involvement), and one of the control group (all first-year Social Science students who did not participate in the FSS+ program). The primary assessment tool was NSSE 2008, which was administered to a 100 per cent sample of first-year Social Science students. NSSE results were linked to FSS+ participation data and a number of student records system items including secondary school average and subsequent retention/attrition. The design considers two separate experimental outcomes: a series of engagement items (see Table 17 above), and two academic performance measures (student grade point average and subsequent attrition/retention status).

Propensity matching of control and experimental group records was performed in order to control for self-selection bias as shown in Table 18. For the analysis of engagement outcomes (for which a NSSE response was a prerequisite), the match was performed using age, gender and application type; admission average was missing for a sufficiently large number of experimental records that its inclusion would have reduced sample sizes below desirable levels. The large control group permitted multiple (3:1) control record matches per experimental group record.

Table 18: University of Ottawa Propensity Matching Results									
		Pre-Match				Post-Match			
Design	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
Engagement Outcomes	Age	0.52	0.24	4.70	0.030	0.00	0.33	0.00	1.000
	Gender	-0.29	0.36	0.65	0.419	0.00	0.40	0.00	1.000
	Application Type	-0.55	0.46	1.46	0.227	0.07	0.52	0.02	0.896
	Admission Average					-0.06	0.01	2.48	0.116
		(n=39 experimental, n=454 control)				(n=39 experimental, n=117 control)			
Academic Outcomes	Age	0.52	0.14	13.44	0.000	0.00	0.21	0.00	1.000
	Gender	0.04	0.23	0.04	0.850	0.00	0.26	0.00	1.000
	Application Type	-0.83	0.33	6.51	0.011	0.00	0.37	0.00	1.000
	Admission Average					0.00	0.00	0.05	0.824
		(n=93 experimental, n=1569 control)				(n=93 experimental, n=279 control)			

A post-match analysis of admission average indicated no significant control and experimental group differences in any of the matching variables or on admission average (checked post-match). (The result of the post-match test on admission average – p-value .11 – was of sufficient concern to warrant performing the analysis on engagement outcomes both with and without an admission average control). For the analysis of academic outcomes, NSSE responses were not required. Initial analysis indicated significant differences between NSSE respondents and non-respondents, so a separate matching process was undertaken. Matching was again performed using age, gender and application type (but not admission average due to the rate of missing data); three control records were matched to each experimental record without exceeding the distance matrix criteria. The matching removed all significant differences between control and experimental groups; a post-match test on admission average also confirmed no experimental/control difference.

4.4.4 Assessment Results

Table 19 indicates that NSSE engagement measures did not detect FSS+ participation: only the “asked questions in class” item was significant (with and without a control for admission average). Admission average proved significant in only one other test – “campus events and activities” – where the engagement measure itself was insignificant. Possible explanations for insignificant experimental effects include the dilution effect of NSSE-based measurement, generally low or varying levels of student involvement in FSS+ activities and services, and the relatively small number of FSS+ participants for whom NSSE responses were available.

The models predicting academic outcomes indicated significant differences between the control and experimental group with respect to student grade point average. (The precautionary inclusion of an admission average control (which showed no significant post-match difference) resulted in a substantial decline in the experimental group “n” and a disproportionate decline in the experimental group size relative to the control group, possibly resulting in mismatched groups and unreliable results. No other potential predisposition measures (i.e., matching variables) were significant in combination with student grade point average. Retention status (modeled using logistic regression) showed no significant experimental effect.

4.4.5 Summary

The existence of a single significant engagement effect and two significant but conflicting academic outcome effects indicates that the NSSE measures provide no evidence of intervention impacts. It cannot be determined whether this is the result of the intervention itself (low impact and/or low intensity of involvement), the dilution effect of NSSE in relation to the scope of the intervention, the relatively small experimental group for which NSSE results were available, or other factors. The large control group allowed multiple matches per experimental record while utilizing 100 per cent of the control group; however missing data on admission average prevented its utilization in both the propensity matching exercise and as a control in the regression models. A larger experimental group, more complete data for propensity matching and modeling, and an intensity of involvement measures would have supported a more detailed analysis.

Table 19: University of Ottawa Regression Results for Engagement and Academic Outcomes									
Dependent Variable		**FSS+ Participation**				**Admission Average Control**			
	R^2	**B-Est**	**SE**	**t-score**	**p-value**	**B-Est**	**SE**	**t-score**	**p-value**
Engagement Measures									
Asked questions in class/contributed to class discussions	0.039	0.368	0.146	2.51	**0.013**				
	0.040	0.334	0.161	2.06	**0.041**	-0.003	0.003	-1.32	0.190
Worked with classmates outside of class to prepare assignments	0.002	0.086	0.153	0.56	0.578				
Discussed ideas with others outside of class	0.020	0.282	0.167	1.69	0.093				
Participated in learning community	0.017	0.270	0.174	1.55	0.123				
Quality of relationships with other students	0.012	0.346	0.267	1.30	0.197				
Quality of relationships with faculty members	0.008	-0.256	0.246	-1.04	0.301				
Quality of relationships with administrative staff	0.000	0.031	0.264	0.12	0.905				
Institutional emphasis on providing support for academic success	0.002	-0.078	0.169	-0.47	0.640				
Institutional emphasis on encouraging contact among different students	0.003	0.133	0.207	0.64	0.521				
Institutional emphasis on providing support to thrive socially	0.003	-0.110	0.182	-0.60	0.546				
Institutional emphasis on attending campus events and activities	0.002	0.101	0.197	0.51	0.609				
	0.063	0.000	0.216	0.00	0.999	0.009	0.003	2.74	**0.007**
Institutional contribution to using computing and information technology	0.016	-0.273	0.183	-1.49	0.139				
Institutional contribution to working effectively with others	0.008	0.200	0.195	1.02	0.309				
Institutional contribution to encouraging voting in elections	0.001	0.060	0.204	0.30	0.768				
Academic Outcomes									
Student grade point average	0.014	0.712	0.316	2.25	**0.025**				
	0.027	0.734	0.339	2.17	**0.031**	0.010	0.005	2.02	**0.044**
Attrition/Retention status (* Wald chi-square)	n/a	0.221	0.339	.423*	0.516				

4.5 Queen's University (Enrichment Component in a Large Introductory Course)

4.5.1 Intervention Description and Background

The core first-year Psychology course at Queen's University is delivered in three large lecture sections of 300 – 400 students each, and provides limited opportunity for in-class student-faculty interaction or discussion. The course covers both Fall and Winter terms. In order to enhance faculty-student interaction, provide opportunities for discussion and

expose students to research and professional practice in the discipline, the "Discovery Project" pilot was developed. It involved the delivery of six intensive small group (20 students) sessions, laboratory tours, experiments and demonstrations highlighting both the natural science and social science dimensions of professional practice and research (perceptual phenomena, pediatric cochlear implants, neuroimaging and brain injury, laboratory and self-reported health measures, quantitative sensory testing and Doppler imaging of blood flow). Prior to the start of the course, all course registrants were sent an information package and were invited to self-select for project participation. All students in the course earn a portion of their grade through subject pool participation; Discovery Project participants received similar consideration if they attended a specified minimum number of sessions.

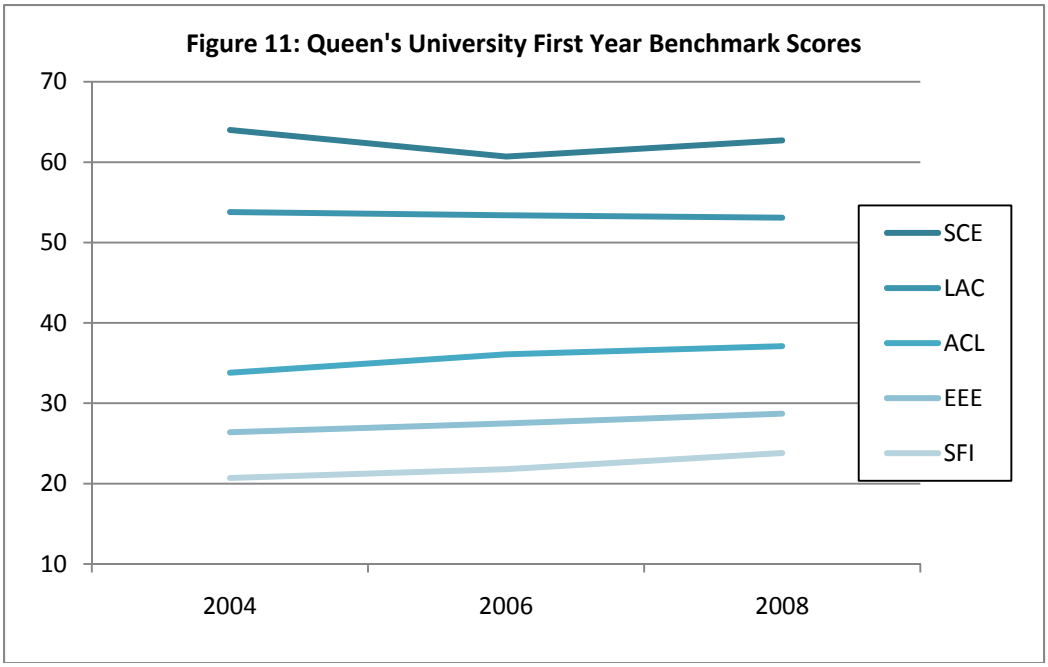
The Discovery Project is a direct response to concerns by the course professors that the exclusively large lecture format limited the learning experience. Less directly, Queen's considers its core strength to be the linked social and academic experience offered to students (including faculty-student and student-student interaction and enrichment opportunities); and the University's strategic plan focuses on both internal and external engagement of students, faculty and the university overall.

Queen's administered NSSE in 2004, 2006 and 2008 (the latter two with 100 per cent samples) and has constructed drilldown reports for Faculties, selected departments, course clusters and student subgroups. It also participated in BCSSE in 2005 and constructed an integrated BCSSE/NSSE analysis from the results.

4.5.2 Context Provided by NSSE Administrations

Multi-year benchmark results for Queen's show a gradual improvement with each administration in the ACL, SFI and EEE benchmark scores (Figure 11). LAC scores remain virtually identical over time, while the SCE benchmark is more erratic (-5.2 per cent and 3.3 per cent).

Individual NSSE items that are the focus of the intervention are presented in Table 20 for the Faculty of Arts and Science overall. (Queen's students in the Faculty of Arts and Science do not generally declare a concentration until second year, so a program-specific drilldown is not possible.) Similar results were obtained for each of a series of first-year course clusters involving the Introductory Psychology Course. The majority of the items show slight to moderate year-over-year increases.



The general though minor increases in NSSE items between 2006 and 2008 need to be accommodated in the successive cohort design involving NSSE (see below).

Table 20: Queen's University First-Year Faculty of Arts and Science NSSE Item (Dependent Variable) Means

NSSE Item	2006	2008	% Change
Talked about career plans with faculty member/advisor	1.50	1.61	7.33%
Discussed ideas from readings/classes with faculty outside class	1.60	1.65	3.12%
Discussed grades or assignments with instructor	2.01	2.11	4.98%
Worked on research project with faculty outside class/program	0.02	0.01	-50.00%
Participation in practicum/internship/field experience/clinical assignment	0.06	0.05	-16.67%
Quality of relationships with other students	5.67	5.69	0.35%
Quality of relationships with faculty members	4.91	4.97	1.22%
Institutional environment provides support you need to succeed academically	3.03	3.06	0.99%
Worked harder to meet instructor expectations	2.34	2.46	5.13%
Coursework emphasis on analyzing elements of idea/experience/theory	3.10	3.18	2.58%
Coursework emphasis on synthesizing/organizing ideas and information	2.84	2.86	0.70%
Coursework emphasis on making judgments about value of info/arguments	2.72	2.80	2.94%
Coursework emphasis on applying theories to practical problems	2.91	3.03	4.12%
Asked questions or contributed to class discussions	2.15	2.11	-1.86%
Worked with classmates outside class to prepare assignments	2.48	2.61	5.24%
Discussed ideas from readings/classes with others outside class	2.90	2.93	1.03%

4.5.3 Assessment Design

Both a successive cohort post-measure design (for NSSE) and a cross-sectional post-measure design (for CLASSE) were employed. In addition, a custom survey was administered to Discovery participants at the half-way point and at the completion of the project. The three surveys provide a "continuum" of measures dealing with the Discovery Project itself (the custom survey), the Psychology course overall incorporating the Discovery Project for participants (CLASSE) and the year-long experience of which the Psychology course was a single component (NSSE).

NSSE was administered to 100 per cent of students registered in the Introductory Psychology course in both March 2008 (prior to the intervention) and March 2009 (to both Discovery participants and non-participants). The general upward trend in NSSE scores (at least within the Faculty overall) suggested the need to measure intervention impacts net of this background noise. 2009 Discovery participants (the experimental group) were propensity matched to 2008 students to construct a "mirror experimental" group – that is, a prior group having similar characteristics to current participants. Similarly, 2009 Discovery non-participants (the control group) were matched to remaining 2008 students to construct a "mirror control" group. In both cases, matching was performed using gender and admission average (attached to the NSSE response record along with other demographic and academic data) as shown in Table 21. The experimental and mirror experimental samples showed a marginally significant pre-match difference on admission average that was eliminated post-match. A 2:1 match ratio was possible given the large number of 2008 NSSE respondents. Regression analysis was performed to provide a baseline measure of engagement change (background noise) between the mirror control (2008) and control (2009) groups. A parallel set of regression models was run to measure engagement change between the mirror experimental (2008) and experimental (2009) groups. The regression coefficients generated by each of these parallel sets of models were compared; where differences exist and where the experimental group coefficients themselves are significant, it is reasonable to assume that the differences are associated with the intervention itself, net of background changes between 2008 and 2009.

Table 21: Queen's University Propensity Matching Results for NSSE and CLASSE Designs									
		Pre-Match				Post-Match			
Design	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
NSSE Experimental	Gender	-0.04	0.29	0.01	0.904	0.00	0.34	0.00	1.000
	Admission Average	-0.05	0.03	2.88	0.090	0.00	0.03	0.00	0.995
		(n=75 experimental, n=612 mirror)				(n=75 experimental, n=150 mirror)			
NSSE Control	Gender	0.12	0.22	0.31	0.580	0.00	0.22	0.00	1.000
	Admission Average	0.01	0.02	0.19	0.666	0.00	0.02	0.00	0.986
		(n=186 control, n=462 mirror (records remaining after experimental matching)				(n=186 control, n=372 mirror)			
CLASSE	Gender	-0.71	0.46	2.34	0.126	-0.21	0.51	0.17	0.677
	Admission Average	-0.05	0.04	1.54	0.215	-0.02	0.04	0.13	0.724
		(n=33 experimental, n=122 control)				(n=31 experimental, n=93 control)			

The cross-sectional assessment design utilized a slightly modified CLASSE survey instrument, which was administered in-class in all three sections of the course with optional respondent self-identification (and subsequently via survey mail-out/mail back to Discovery Project participants who did not self-identify in the classroom administration). Those students who self-identified were assigned to either the experimental (project participant) or control group. As with NSSE response records, CLASSE responses were linked to student demographic and academic information from the student records system (gender, admission average, Psychology course grade, full-/part-time status). Propensity matching was performed using gender and admission average and resulted in no significant post-match differences on the matching variables (see Figure 21) with a 3:1 control-experimental match ratio. CLASSE responses were also linked to project session attendance data provided by the course coordinator that were intended to provide an intensity of participation analysis. However, low attrition and consistently high attendance (with virtually all students attending five or more of the six sessions) made it unnecessary to incorporate intensity of participation.

The custom survey administered to Discovery participants during the final session in each term dealt with the topics students found (un)interesting, interactions with professors and fellow students, participation in experiments, and (on the second term instrument only)

changes in student interest in psychology. The two surveys achieved response rates of about 80 per cent and 60 per cent respectively.

Though not included in this report because of time limitations, the declared academic program majors of project participants and non-participants will be tracked over time and subject to future analysis to assess whether the Discovery Project played a role in subsequent program choice by students.

Another assessment approach was planned but subsequently abandoned: the linking of CLASSE and NSSE responses at the student record-level to more directly assess the consistency and "power" of the two sets of measures with respect to the comparable items on both instruments. The experimental group was too small and the number of students responding to both NSSE and CLASSE surveys too low to generate sufficient group sizes for an integrated analysis.

4.5.4 Assessment Results

Custom Survey Results:

Though designed primarily as a qualitative assessment tool, the custom survey provided some quantitative data that map, at least indirectly, to questions on both NSSE and CLASSE. Summary results are presented in Table 22. Though obviously comparisons cannot be made against Discovery non-participants, the results suggest students perceived several program impacts:

- An appreciation for the hands-on aspects of the program (high satisfaction with opportunities to participate in demonstrations), but a clear desire for a more hands-on and less lecture/instructional approach (unprompted mentions of "uninteresting" program components);
- Acknowledgement of a greater level of interaction with faculty (both unprompted mention of greater interaction, and a high satisfaction rating for faculty contact opportunities) and with student peers;
- A sustained high level of participation (due at least in part to the awarding of a Discovery participation grade in lieu of subject pool participation);

High satisfaction with "opportunities to talk with professor" and "opportunities to discuss ideas with peers" might be expected to roughly mirror similarly-intentioned questions on the CLASSE survey (e.g., asking questions in class, contributing to discussions, using email to communicate with instructor, level of comfort communicating with TA or instructor). Similarly, somewhat weaker responses to "opportunities to participate in demonstrations" and "growth of interest in psychology" might also be reflected in the CLASSE questions dealing with exposure to psychological research and research tools, and level of interest in course material. As a stand-alone assessment tool (and tools like this are common in academic assessment), the custom survey suggests that the Discovery Project achieved several of its academic objectives.

Table 22: Summary of Queen's University Custom Survey Results		
Interesting Elements of Discovery Program (Open-Ended Mult-Response, n=259)		
Content-Related		59%
General	22%	
Applications/examples	12%	
Specific content examples	10%	
Complemented lectures/depth of topic coverage	16%	
Delivery-Related		36%
General	6%	
Experiments/hands-on	7%	
Small groups	6%	
Increased interactions	17%	
Helped in studying/coursework		5%
Uninteresting Elements of Discovery Program (Open-Ended Mult-Response, n=81)		
Less lecturing/more discussion/more experiments		54%
Specific content examples		25%
Other		21%
Satisfaction/Interest Measures (means for 1-5 Likert scale)		
Opportunity to talk with professor		4.47
Opportunity to discuss ideas with peers		4.06
Opportunity to participate in demonstrations		4.00
Growth of interest in psychology (end-of-course only)		3.94
Number of Sessions Attended (mean out of 6 possible)		5.40

CLASSE Results:

The base CLASSE instrument was modified only slightly and three questions specifically related to research exposure and understanding (the vehicle used by the Discovery Project to enrich learning and enhance interaction) were added on-site. The results are presented in Table 23 in the order they appeared on the instrument. The first test of the sensitivity of the CLASSE instrument to the existence of the Discovery Project is identification of those items that should not show an experimental effect. Participation did not involve any extra assignments or tests and did not affect lectures in any way. It is encouraging, then, that numerous items showed no significant differences including:

- Papers and assignments: submitting 2+ drafts of a paper, working on a paper that required integrating ideas from various sources, discussing grades or assignments with instructor, promptness of feedback, papers written greater than 5 pages in length, number of homework assignments taking more than an hour to complete;
- Lectures: making a class presentation, note-taking in class, reviewing notes prior to class, ease of following lectures, number of class absences;
- Other: tutoring or teaching other students.

Several items did show significant differences or differences bordering on standard significance; most of these have a direct connection to the purpose and practice of the Discovery Project (i.e., encouraging oral session participation, providing a professional and research perspective, increasing peer interaction, improving student-faculty interaction and encouraging academic effort:

- Asking questions in class and contributing to class discussions;
- Including diverse perspectives in discussions or writing assignments;
- Participation in a study partnership;
- Emailing instructor and level of comfort talking with TA, instructor or researcher;
- Spending 3 or more hours per week in class preparation and participating in a community-based project. (The latter is ambiguous: it is unclear whether participants considered visits to community research facilities as a community activity or whether their participation spurred other activity.)

A third category of CLASSE items showed no statistical significance but might be considered desired or expected project outcomes in that they directly or indirectly capture academic effort or peer interaction level, or reflect indirect or potential higher-order learning objectives of the project including:

- Research: the three research questions (amount learned about research, understanding of subject helped by participation in research, or opportunities to discuss tools used in studies);
- Peer interaction: worked with students during class or outside class, discussed course ideas with others;
- Academic effort: course emphasis on analysis/synthesis/judgment /application of theory, frequency of coming to class unprepared, working harder to meet expectations.

Table 23: Queen's University Regression Results for CLASSE Engagement and Academic Outcome Measures					
Dependent Variable	**Experimental Effects**				
	R^2	**B-Est**	**SE**	**t-score**	**p-value**
Engagement Measures (questions dealing with activities during the course ask students to consider lectures, tutorials, research studies and Discovery sessions in the Psychology course)					
Asked questions during course activities	0.108	0.815	0.215	3.79	**0.000**
Contributed to discussion	0.196	1.044	0.194	5.36	**<.0001**
Prepared two or more drafts of a paper or assignment	0.006	-0.057	0.069	-0.83	0.411
Worked on a paper that required integrating ideas or info from various sources	0.002	-0.040	0.093	-0.43	0.670
Included diverse perspectives … in discussions or writing assignments	0.045	0.486	0.207	2.35	**0.020**
Came to class without completing readings or assignments	0.000	-0.040	0.189	-0.21	0.834
Worked with other students on projects during class	0.001	0.057	0.180	0.32	0.752
Worked with classmates outside of class to prepare assignments	0.001	0.066	0.204	0.32	0.747
Put together ideas/concepts from different courses when doing assignments	0.009	0.230	0.220	1.05	0.296
Tutored or taught other students in the course	0.002	0.089	0.191	0.46	0.645
Used an electronic medium to discuss or complete an assignment	0.001	-0.052	0.203	-0.25	0.801
Used email to communicate with instructor, TA or coordinator	0.025	0.326	0.186	1.75	**0.083**
Discussed grades or assignments with instructor, TA or coordinator	0.019	0.257	0.173	1.49	0.140
Discussed course ideas with others outside of class	0.007	-0.174	0.187	-0.93	0.354
Made a presentation in class	0.007	-0.057	0.063	-0.90	0.371
Participated in a community-based project as part of the course	0.029	-0.250	0.135	-1.86	**0.066**
Discussed ideas re: course with instructor outside class	0.001	-0.069	0.187	-0.37	0.711
Received prompt feedback on your academic performance in the course	0.014	-0.146	0.114	-1.28	0.205
Worked harder to meet instructor's standards or expectations	0.000	-0.007	0.206	-0.03	0.974
Course emphasis on memorization	0.023	0.244	0.147	1.65	0.101
Course emphasis on analysis	0.000	-0.025	0.188	-0.13	0.895
Course emphasis on synthesis	0.000	-0.015	0.196	-0.07	0.941
Course emphasis on making judgments	0.001	-0.064	0.207	-0.31	0.758
Course emphasis on application of theory	0.020	0.314	0.205	1.53	0.129
Number of required papers more than 5 pages long	0.001	0.019	0.072	0.26	0.797
Extent to which exams challenge you to do your best work	0.001	-0.076	0.187	-0.40	0.687
Numer of homework assignments per week taking more than one hour	0.001	0.052	0.181	0.29	0.775
Frequency of 3+ hours per week preparing for course	0.023	-0.358	0.215	-1.67	**0.098**
Number of absences in course throughout year	0.014	0.258	0.199	1.30	0.197
Frequency of note-taking in course	0.001	0.161	0.157	1.02	0.309
Frequency of reviewing notes prior to next class	0.013	-0.258	0.211	-1.22	0.224
Frequency of participation in a study partnership with classmate	0.026	0.421	0.239	1.76	**0.081**
Frequency of attending review/help session	0.000	0.055	0.233	0.24	0.813
Level of interest in learning course material	0.019	0.245	0.163	1.50	0.136
Level of comfort talking with TA, instructor or researchers	0.033	0.380	0.190	1.99	**0.049**

Level of enjoyment of group work with classmates	0.000	0.031	0.194	0.16	0.873
Difficulty of course material	0.007	-0.181	0.194	-0.93	0.353
Ease of following lectures	0.001	0.056	0.143	0.39	0.696
Amount learned about psychological research	0.008	0.162	0.164	0.99	0.325
Participation in psych research helped understanidng of methods/theories	0.009	-0.155	0.155	-1.00	0.320
Opportunity to discuss tools used by psychologists in studies	0.001	0.074	0.195	0.38	0.706
Academic Outcome Measure					
Course Grade	0.001	-0.849	2.360	-0.36	0.720

NSSE Results:

As noted above, two parallel regressions series were run: one on the control group and its prior matched cohort to establish a baseline, and one on the experimental group and its prior matched cohort, to establish experimental effects in comparison to the baseline. A large number of insignificant coefficients for the control vs. mirror control regressions is to be expected, to the extent that the members did not participate in the intervention and because for most items, background changes between 2008 and 2009 would be expected to be relatively minor (though significant positive or negative coefficients could indeed occur). Figure 24 presents the results of the analysis.

Table 24

The table identifies three items that are close to the standard significance level:

- Coursework emphasis on analysis: carries a negative coefficient (p=.087) in the control groups regression and a positive coefficient (p=.099) in the experimental groups regression;
- Coursework emphasis on synthesis: is insignificant for the control groups and carries a positive coefficient (p=.059) in the experimental groups regression;
- Discussed ideas with others outside class: is insignificant for the control groups and carries a positive coefficient (p=.087) in the experimental groups regression.

On their own, such results might suggest a cautious finding that the Discovery Project was the cause of the differences. However, the CLASSE analysis above reveals all three items to be insignificant at the course level. Similarly, while the NSSE results indicate no difference on the “asked questions or contributed to class discussions” item, the CLASSE results indicate a strong relationship on both the “questions” and “discussions” items. This constitutes compelling evidence, for this situation at least, that NSSE item scores measure variation in the student experience that cannot be associated with a particular subset of that experience (i.e., the Discovery Project).

4.5.5 Summary

Two of the three sample groups (the NSSE-experimental successive cohorts and the CLASSE cross-sectional cohorts) displayed sufficiently large differences pre-match to cause concern; all significant differences were eliminated post-match. The control groups were sufficiently large to achieve 93 per cent - 100 per cent utilization of experimental records in the three matching exercises. Neither the NSSE successive cohort design nor the CLASSE cross-sectional design presented any analytical or interpretive difficulties. However, because the matching uses demographic characteristics as surrogates for self-selection and/or predisposition bias, the cross-sectional CLASSE results should be interpreted with some degree of caution (as suggested by the University of Guelph analysis above).

Because this project was undertaken at the author's own university, issues related to the ethical approval and field administration of the CLASSE instrument came to light that are worth reporting here. In order to accommodate potential self-selection bias and permit propensity matching, CLASSE requires respondent self-identification: anonymous response is only useful when high response rates are achieved in very high participation rate experiments. Queen's ethics board requirements for the structure and content of the CLASSE information/consent form resulted in a relatively low rate of self-identification that made a mailout follow-up administration necessary. A requirement to administer the survey at the end, rather than at the beginning of the classes resulted in a far lower response rate than that achieved in administrations elsewhere. A cooperative effort among institutions using the instrument would likely result in more consistent classroom administration and higher response. And finally, the ethics board expressed concern (despite voluntary Discovery Project participation and the existence of alternative means for students to satisfy the course's subject pool participation requirement) that project participation could result in an improved grade that would warrant compensatory treatment for non-participants. It is therefore noteworthy that Discovery participation showed no statistical effect on course grades.

The results of the three surveys suggest a progressive dilution of the measured experimental effects. While the findings from the custom survey carry over reasonably well to the CLASSE findings, a complete break occurred in the transition from CLASSE to NSSE, where statistical results became either insignificant or contradictory. The unfulfilled objective to link NSSE and CLASSE responses at the record-level might have helped address this issue.

Despite the difficulties noted above, CLASSE provided the most promising findings with respect to the survey's ability to detect experimental effects in oral session participation, professional and research perspective, peer and student-faculty interactions and (perhaps more ambiguously) encouraging academic effort. It did not detect any effects with respect to the research questions (amount learned about research, understanding of subject helped by participation in research, or opportunities to discuss tools used in studies). Whether this is the result of the intervention itself or the measurement tool cannot be determined.

4.6 University of Western Ontario (Biology Science Literacy Initiative)

4.6.1 Intervention Description and Background

The objective of the Biology Science Literacy Initiative (BSLI) is to fully integrate the development of science literacy skills into the first year of the undergraduate Biology curriculum at the University of Western Ontario. Science literacy is not currently part of any introductory science courses at Western except for two courses with a non-core science literacy on-line component. The BSLI was implemented in two large full-year (2 semester long) introductory Biology courses in the 2008/09 academic year with a combined enrolment of about 1,800 almost exclusively first-year students enrolled in the Faculties of Science and Health Sciences. BIOL 1222 (offered in two sections) is intended for students who are fully prepared in terms of secondary school Biology coursework and consists of 2 hours of lectures plus 3 hours of lab/tutorial; BIOL 1223 (offered in one section) is intended for those students without recent experience or sufficiently high performance in secondary school Biology and consists of 3 hours of lectures and 3 hours of lab/tutorial.

The BSLI was implemented primarily via the tutorial component of the courses. The tutorials were completely revised to centre on science literacy skills development within a Biology context including:

- Scientific information retrieval and evaluation;
- Integration of scientific information;
- Scientific writing;
- Critical analysis of science writing;
- Connecting lecture content and laboratory assignments/activities;
- Demonstrating connections in content among Biology, Chemistry, Physics and Health Sciences.

In addition to the labs and tutorials, students had access to various on-line information and skills development resources. Within the tutorials, they evaluated their own scientific writing assignments, and they were required to work collaboratively in completing various assignments.

Western administered NSSE in 2004, 2006 and 2008. It has since incorporated engagement into its performance measurement and accountability reporting, and into its University and Faculty of Science strategic plans, whose goals are to increase undergraduate engagement through course and curricular initiatives.

4.6.2 Context Provided by NSSE Administrations

Western's first-year benchmark scores show a slight upward trend over the three administrations: the largest increase is in the SFI benchmark (up 16 per cent from 2006 to 2008) while the other four benchmarks increased between 2.5 per cent and five per cent

during the same period (see Figure 12). The individual items selected by Western as the focus for the assessment follow the same pattern (see Table 25) with most increasing between zero per cent and five per cent. These generally increasing engagement scores require an assessment design that at least qualitatively controls for the prospect of non-experimental change occurring during the intervention period.

Figure 12: University of Western Ontario First Year Benchmark Scores

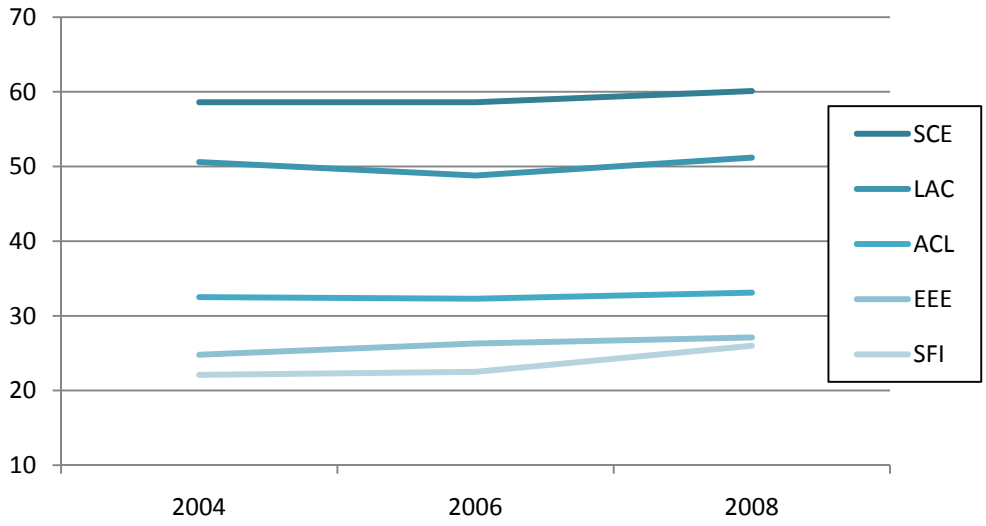


Table 25: University of Western Ontario First-Year Selected NSSE Item (Dependent Variable) Means			
NSSE Item	**2006**	**2008**	**% Change**
Paper/project required integrating ideas/info from various sources	2.71	2.93	8.12%
Worked with other students on projects during class	1.78	1.70	-4.49%
Worked with classmates outside class to prepare assignments	2.46	2.46	0.00%
Put together ideas from different courses when completing assignments	2.50	2.69	7.60%
Discussed ideas from readings/classes with others outside class	2.76	2.88	4.35%
Coursework emphasis on memorization	2.87	2.89	0.70%
Coursework emphasis on analyzing elements of idea/experience/theory	3.09	3.16	2.27%
Coursework emphasis on synthesizing/organizing ideas and information	2.78	2.94	5.76%
Coursework emphasis on making judgments about value of info/arguments	2.68	2.88	7.46%
Coursework emphasis on applying theories to practical problems	3.11	3.15	1.29%
Number of papers fewer than 5 pages	2.12	2.24	5.66%
Examination challenged you to do your best work	5.60	5.61	0.18%
Hours per week spent preparing for class	4.26	4.30	0.94%
Institutional contribution to writing clearly and effectively	2.56	2.64	3.13%
Institutional contribution to thinking critically and analytically	3.23	3.24	0.31%
Institutional contribution to working effectively with others	2.68	2.71	1.12%

4.6.3 Assessment Design

The primary focus of the assessment involves a successive cohort design utilizing a 2008 control group and a 2009 experimental group. Inclusion of a cross-sectional design either alone or in combination with the successive cohort design (i.e., limiting the project to one section of the course while leaving others untouched) would have required random student assignment to course sections and the effective denial of potential benefits to non-participants (an ethical issue mentioned earlier). The intervention as designed, and the range of data items assembled, permit a multi-faceted assessment and analysis approach that consists of seven different components.

The first involves NSSE as the intervention assessment tool. NSSE was administered to a 100 per cent sample of students in the two Biology courses in both 2008 (control) and 2009 (experimental) as shown in Figure 26. Analysis can be performed for the two courses combined and at the individual course level (since the preparation level of students may influence BSLI impact). Western identified a number of NSSE response items for the analysis that it considered consistent with the objectives and potential outcomes of the intervention. The second analysis component uses a CLASSE-type course-specific experience and engagement instrument that was modified from the base CLASSE tool and substantially enhanced with additional questions by the project

participants at Western. The CLASSE survey was also administered to a 100 per cent sample of students under identical conditions in all three sections of the courses in both 2008 and 2009, and like NSSE, also permits both course-specific and combined-course analysis. Almost all of the 40+ questions on the instrument were included in this analysis. The modified CLASSE instrument asks students to record the number of lecture, lab and tutorial sessions they missed throughout the course, and provides a third assessment dimension: a rough indicator of the intensity of student involvement.

Table 26: University of Western Ontario Survey Response and Test Capture Rates

Course	Year	NSSE		CLASSE		Literacy Assessment	
		Completions	Response %	Completions	Response %	Completions	Capture %
BIOL 1222	2008	309	24%	892	68%	63	5%
	2009	386	28%	769	56%	59	4%
BIOL 1223	2008	74	19%	251	52%	11	2%
	2009	74	19%	154	39%	4	1%
Combined	2008	383	23%	1143	64%	74	4%
	2009	460	26%	923	52%	63	4%

The fourth component consists of an on-line science literacy assessment test consisting of about a half-dozen questions. The test was developed by Western and administered under identical conditions in both 2008 and 2009 to a relatively small subset of students from both courses who self-selected to take the test outside of class hours and who were offered a small financial incentive to do so. Test scores can be compared across the 2008 and 2009 cohorts.

The fifth component is made possible by the linkage of all NSSE, CLASSE and literacy test respondents within a single data file to facilitate comparison of the sensitivity and explanatory power of the three tools, each of which offers a different focus: the specific science literacy outcome of BSLI participation, the course experience including BSLI, and the overall student experience of which the Biology courses were a component.

A key sixth dimension of the assessment allows an exploration of the necessity of propensity matching. Large, close-in-time successive cohorts within the same course in the absence of self-selection bias could be assumed to show little change year over year. Some analyses were undertaken both with and without propensity matching in order to

assess the validity of the large successive cohort assumption. Propensity matching was performed for both the NSSE response set and the CLASSE response set, for BIOL 1222 and BIOL 1223 separately. Matching was performed using basis of admission (direct/indirect from secondary school), Faculty of student registration, gender) in all cases, and using Grade 12 Biology grade and entering (secondary school) average in the case of BIOL 1222 only as shown in Table 27. (Many BIOL 1223 registrants were admitted on a basis other than their secondary school average and the missing data rate for secondary school information was too high to match without significant loss of sample size.)

Finally, Western was able to provide demographic and academic data for all students registered in both courses over both years. While the NSSE, CLASSE and literacy test analyses are limited to the respondents of each of those instruments, data for all course registrants permits an analysis of non-response patterns for single-survey (NSSE or CLASSE), multiple survey (NSSE plus CLASSE) and test administrations, thus providing some insight into the applicability of survey-based assessment results to the entire course population.

Table 27: University of Western Ontario Propensity Matching Results for NSSE and CLASSE Designs

Design	Variable	Pre-Match				Post-Match			
		Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
NSSE Design (BIOL 1222)	Basis of admission	-0.54	0.30	3.27	0.071	0.00	n/a	n/a	n/a
	Faculty	-1.75	0.76	5.33	0.021	0.00	1.42	0.00	1.000
	Gender	-0.29	0.16	3.14	0.076	0.00	0.18	0.00	1.000
	G12 Biology grade	-0.02	0.02	2.00	0.157	-0.01	0.02	0.25	0.618
	Entering average	0.01	0.02	0.08	0.778	0.01	0.02	0.15	0.699
		(n=386 experimental, n=309 control)				(n=279 experimental, n=279 control)			
NSSE Design (BIOL 1223)	Basis of admission	0.38	0.40	0.97	0.325	0.00	0.45	0.00	1.000
	Faculty	-1.84	0.78	6.14	0.013	0.00	1.02	0.00	1.000
	Gender	0.07	0.37	0.03	0.854	0.00	0.40	0.00	1.000
	G12 Biology grade	not used due to missing data rate				not used due to missing data rate			
	Entering average	not used due to missing data rate				not used due to missing data rate			
		(n=74 experimental, n=74 control)				(n=64 experimental, n=64 control)			
CLASSE Design (BIOL 1222)	Basis of admission	0.50	0.19	6.85	0.009	0.00	n/a	n/a	n/a
	Faculty	0.24	0.29	0.67	0.414	0.00	0.40	0.00	1.000
	Gender	0.16	0.10	2.38	0.123	0.00	0.11	0.00	1.000
	G12 Biology grade	-0.01	0.01	0.73	0.393	0.00	0.01	0.03	0.856
	Entering average	-0.03	0.01	5.35	0.021	0.00	0.01	0.04	0.841
		(n=769 experimental, n=892 control)				(n=691 experimental, n=691 control)			
CLASSE Design (BIOL 1223)	Basis of admission	0.37	0.25	2.18	0.139	0.00	0.28	0.00	1.000
	Faculty	0.60	0.32	3.45	0.063	0.00	0.36	0.00	1.000
	Gender	0.24	0.21	1.31	0.252	0.00	0.24	0.00	1.000
	G12 Biology grade	not used due to missing data rate				not used due to missing data rate			
	Entering average	not used due to missing data rate				not used due to missing data rate			
		(n=154 experimental, n=251 control)				(n=150 experimental, n=150 control)			

4.6.4 Assessment Results

NSSE Results:

Table 28 presents the NSSE results for the pooled responses of students in both Biology courses. Because the Western project uses a successive cohort design (i.e., without a cross-sectional control) significant coefficients must be at least qualitatively benchmarked against the overall time-series trend in NSSE results. The table identifies several items that show significant year over year change in either the matched groups or both the matched and unmatched groups. As with the Queen's project, one might wish to conclude that the change is the result of the BSLI project. However, the successive cohort design requires a comparison of the item changes against the background environment. University-wide NSSE results presented in Table 25 above are based on sample sizes producing approximately two per cent maximum likely error (at the 95 per cent level) and the comparisons suggest:

- Working with other students during class (significantly positive for the matched and unmatched groups) is inconsistent with the university-wide pattern (significantly negative); Working with classmates outside of class (weak negative for both groups) is inconsistent with the university wide trend (significantly positive);
- Coursework emphasis on the application of theory (weak negative for the matched groups) and coursework emphasis on memorization (weak positive for both groups) are inconsistent with an insignificant difference in the university-wide results for both items;
- Coursework emphasis on synthesis, number of papers less than 5 pages in length, and institutional contribution to writing (all significantly positive for at least the matched group) mirror significantly positive results university-wide.

Table 28: University of Western Ontario Regression Results for NSSE Design (BIOL 1222 + BIOL 1223)										
Dependent Variable	**Unmatched Groups**					**Matched Groups**				
	R^2	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Paper/project required integrating ideas/info from various sources	0.000	0.009	0.061	0.15	0.880	0.000	-0.021	0.067	-0.32	0.750
Worked with other students on projects during class	0.032	0.296	0.056	5.26	**<.0001**	0.023	0.249	0.062	4.03	**<.0001**
Worked with classmates outside class to prepare assignments	0.004	-0.123	0.069	-1.78	**0.075**	0.005	-0.138	0.076	-1.81	**0.071**
Put together ideas from different courses when completing assignments	0.000	-0.034	0.058	-0.59	0.553	0.002	-0.069	0.065	-1.07	0.287
Discussed ideas from readings/classes with others outside class	0.000	0.023	0.062	0.37	0.712	0.000	-0.018	0.069	-0.26	0.797
Coursework emphasis on memorization	0.005	0.123	0.063	1.94	**0.053**	0.006	0.129	0.071	1.82	**0.069**
Coursework emphasis on analyzing elements of idea/experience/theory	0.000	0.011	0.054	0.20	0.843	0.000	-0.018	0.060	-0.29	0.769
Coursework emphasis on synthesizing/organizing ideas and information	0.001	-0.051	0.062	-0.82	0.412	0.006	-0.138	0.068	-2.04	**0.042**
Coursework emphasis on making judgments about value of info/arguments	0.001	-0.063	0.068	-0.92	0.356	0.002	-0.780	0.075	-1.03	0.304
Coursework emphasis on applying theories to practical problems	0.001	-0.055	0.059	-0.94	0.349	0.004	-0.111	0.065	-1.70	**0.090**
Number of papers fewer than 5 pages	0.012	0.208	0.068	3.07	**0.002**	0.010	0.191	0.074	2.57	**0.011**
Examination challenged you to do your best work	0.001	-0.058	0.080	-0.72	0.470	0.000	-0.034	0.086	-0.39	0.695
Hours per week spent preparing for class	0.002	0.181	0.132	1.36	0.174	0.000	0.024	0.146	0.16	0.870
Institutional contribution to writing clearly and effectively	0.007	0.148	0.065	2.28	**0.023**	0.008	0.160	0.072	2.23	**0.026**
Institutional contribution to thinking critically and analytically	0.000	0.011	0.055	0.21	0.836	0.000	0.031	0.060	0.51	0.607
Institutional contribution to working effectively with others	0.000	-0.015	0.067	-0.23	0.816	0.001	-0.049	0.073	-0.66	0.508

The comparisons indicate that course-specific engagement effects do not appear to carry over to the more broadly-focused NSSE instrument.

Table 28 also indicates that regression results for the matched and unmatched groups are quite similar, despite the evidence in Table 27 that there exist several significant or near-significant pre-match differences in both BIOL 1222 (basis of admission, Faculty and gender) and BIOL 1223 (Faculty). This finding casts doubt on the assumption – at least at the course-level – that large successive cohorts are sufficiently similar to permit

comparison. But even given the similarity of the results, matched samples are still appropriate: in this particular case, basis of admission, Faculty and gender were not sufficiently strong predictors of engagement to substantially alter the regression results. Propensity matching provides assurance that these factors – minor in this case, but possibly major in others – have been accounted for in those cases where sample similarity cannot be assured.

CLASSE Results:

Table 29 summarizes the results of the successive CLASSE administrations for BIOL 1222 and BIOL 1223 separately to determine whether impacts differ with respect to student academic preparation.

Table 29: University of Western Ontario Regression Results for CLASSE Design - BIOL 1222 + BIOL 1223

	Matched Groups - BIOL 1222					Matched Groups - BIOL 1223				
Engagement and Academic Outcomes	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Engagement, Experience and Perception Measures										
Asked questions of instructor or TA	0.013	-0.240	0.058	-4.14	**<.0001**	0.013	0.260	0.131	1.97	**0.050**
Contributed to class discussions	0.002	0.068	0.046	1.47	0.143	0.018	0.267	0.114	2.34	**0.020**
Worked on assignment requiring integration of ideas from various sources	0.142	0.687	0.046	15.10	**<.0001**	0.116	0.636	0.102	6.25	**<.0001**
Included diverse perspectives in discussions or assignments	0.010	0.157	0.043	3.64	**0.001**	0.000	0.004	0.085	0.05	0.961
Came to class without completing readings or assignments	0.011	-0.233	0.060	-3.65	**0.000**	0.000	0.047	0.132	0.36	0.721
Worked with other students on assignments during course	0.013	0.264	0.061	4.30	**<.0001**	0.016	0.293	0.131	2.24	**0.026**
Worked with classmates outside class on assignments	0.000	-0.046	0.058	-0.79	0.429	0.001	0.060	0.121	0.50	0.619
Tutored or taught other students	0.008	-0.187	0.056	-3.36	**0.001**	0.002	0.092	0.109	0.84	0.399
Communicated with instructor or TA via email or WebCT	0.001	0.061	0.049	1.25	0.211	0.056	0.407	0.097	4.21	**<.0001**

Discussed grades with instructor or TA	0.006	0.098	0.035	2.79	**0.005**	0.043	0.293	0.080	3.65	**0.000**
Discussed course/reading ideas with others outside class	0.034	-0.394	0.057	-6.96	**<.0001**	0.009	-0.200	0.121	-1.65	**0.099**
Discussed course/reading ideas with instructor or TA outside class	0.003	-0.078	0.038	-2.07	**0.038**	0.015	0.191	0.089	2.15	**0.032**
Received prompt feedback on performance from instructor or TA	0.024	0.280	0.049	5.73	**<.0001**	0.035	0.313	0.095	3.30	**0.001**
Worked harder to meet instructor standards or expectations	0.016	-0.251	0.053	-0.479	**<.0001**	0.002	-0.082	0.109	-0.75	0.455
Made connection between BIOL 1222 and secondary school biology	0.005	-0.120	0.046	-2.60	**0.010**	0.004	-0.133	0.121	-1.11	0.269
Made connection between BIOL 1222 lectures and labs	0.009	0.174	0.049	3.58	**0.000**	0.006	0.133	0.103	1.30	0.195
Made connection between BIOL 1222 and other courses	0.000	-0.035	0.049	-0.71	0.479	0.007	0.160	0.108	1.48	0.140
Made connection between BIOL 1222 and issues in daily life	0.005	-0.130	0.049	-2.64	**0.009**	0.000	0.002	0.108	0.01	0.988
Made connection between BIOL 1222 and news/media coverage	0.008	-0.157	0.048	-3.30	**0.001**	0.002	0.073	0.104	0.71	0.480
Coursework emphasized memorizing facts, ideas, methods	0.002	0.077	0.052	1.48	0.138	0.008	-0.153	0.098	-1.57	0.119
Coursework emphasized analyzing elements of an idea or theory	0.003	-0.093	0.044	-2.09	**0.037**	0.002	0.080	0.103	0.78	0.437
Coursework emphasized synthesizing and organizing information	0.005	-0.126	0.047	-2.66	**0.008**	0.000	0.040	0.106	0.38	0.705
Coursework emphasized making judgments about value of information or arguments	0.000	-0.002	0.051	-0.04	0.968	0.005	0.133	0.106	1.25	0.212
Coursework emphasized applying theories to practical problems	0.005	-0.135	0.052	-2.59	**0.010**	0.001	-0.076	0.115	-0.67	0.506
Exams in course challenge you to do your best work	0.051	-0.363	0.042	-8.63	**<.0001**	0.027	-0.282	0.098	-2.89	**0.004**
Number of homework assignments per week taking 1+ hours	0.003	-0.078	0.041	-1.93	**0.054**	0.020	0.240	0.097	2.46	**0.014**
Frequency of spending 3+ hours per week on course outside class	0.010	-0.192	0.052	-3.68	**0.000**	0.001	0.054	0.110	0.49	0.627
Number of lecture session absences	0.006	0.152	0.053	2.89	**0.004**	0.003	0.114	0.115	0.99	0.321
Frequency of writing notes during lectures	0.005	-0.084	0.031	-2.74	**0.006**	0.002	-0.099	0.115	-0.86	0.392

Frequency of reviewing notes prior to lecture	0.007	-0.152	0.048	-3.16	**0.002**	0.000	-0.036	0.096	-0.38	0.708
Frequency of study partnership to study for tests	0.021	-0.306	0.057	-5.40	**<.0001**	0.000	-0.029	0.139	-0.21	0.835
Frequency of attending review or help session	0.002	-0.104	0.058	-1.79	**0.073**	0.047	-0.512	0.134	-3.82	**0.000**
Level of interest in course material	0.001	-0.057	0.041	-1.37	0.170	0.005	-0.114	0.097	-1.17	0.241
Level of comfort talking with instructor or TA	0.000	0.037	0.050	0.74	0.462	0.002	0.075	0.101	0.73	0.463
Level of enjoyment working with classmates in course	0.003	-0.100	0.046	-2.17	**0.030**	0.001	0.065	0.104	0.63	0.530
Level of difficulty of course material	0.036	-0.304	0.042	-7.21	**<.0001**	0.000	0.010	0.087	0.12	0.906
Ease of following material in course lectures	0.005	-0.101	0.040	-2.54	**0.011**	0.000	0.024	0.092	0.26	0.793
Extent to which course contributed to writing effectively	0.003	0.071	0.035	2.00	**0.046**	0.004	0.081	0.078	1.04	0.300
Number of tutorial session absences	0.236	-0.918	0.046	-20.19	**<.0001**	0.308	-1.094	0.097	-11.29	**<.0001**
Effectiveness of tutorials in helping you learn course material	0.166	-0.877	0.054	-16.21	**<.0001**	0.307	-1.530	0.138	-11.12	**<.0001**
Number of laboratory session absences	0.012	0.103	0.026	4.04	**<.0001**	0.003	0.061	0.063	0.97	0.333
Effectiveness of labs in helping you learn course material	0.004	0.116	0.050	2.34	**0.020**	0.034	0.409	0.127	3.21	**0.002**
Self-rated ability searching for and finding authoritative resources	0.025	0.292	0.050	5.87	**<.0001**	0.036	0.365	0.109	3.33	**0.001**
Self-rated ability evaluating validity of resources	0.023	0.259	0.045	5.68	**<.0001**	0.042	0.364	0.101	3.60	**0.000**
Self-rated ability using found resources in appropriate way	0.011	0.193	0.050	3.90	**0.000**	0.026	0.304	0.107	2.83	**0.005**
Self-rated ability paraphrasing passage while avoiding plagiarism	0.012	0.203	0.050	4.03	**<.0001**	0.014	0.213	0.106	2.01	**0.045**
Self-rated ability summarizing article while avoiding plagiarism	0.011	0.204	0.052	3.93	**<.0001**	0.005	0.133	0.109	1.22	0.224
Academic Outcome										
Final course grade	0.007	-1.640	0.514	-3.19	**0.002**	0.006	-1.726	1.251	-1.38	0.169

(Additional analysis dealing with intensity of tutorial involvement follows.) The most striking aspects of the table are the number of items that show significant or near-significant differences in the experimental groups, the apparently greater number of significant differences in BIOL 1222, and the mix of positive and negative coefficients for the significant items (which must be interpreted within the context of project objectives and item wording). About half of the questions are contained in the base CLASSE instrument; the others were developed for the project itself. Table 30 summarizes the large amount of information by major theme.

Table 30: University of Western Ontario Summary of CLASSE Regression Results by Theme

Theme	BIOL 1222 Participants …	BIOL 1223 Participants …
Involvement with Peers	report lower levels of tutoring/teaching other students, working with/discussing ideas with classmates and others outside class	display peer involvement patterns generally similar to those of non-participants
In-Class Engagement	show no consistent pattern of engagement relative to non-participants: higher level of working with classmates in class; no difference in class discussion participation; lower level of asking questions in class	are more active than non-participants with respect to asking questions, participating in class discussions and working with classmates
Faculty/TA Interactions	report higher interaction (discussing grades and receiving prompt feedback); lower interaction (discussing ideas with faculty); and show no differences with respect to email/WebCT use and level of comfort in communications	report generally higher interaction levels than non-participants (email/WebCT, grades discussions, discussion of ideas, receiving prompt feedback) but show no difference with respect to level of comfort in communications
Academic Effort	display generally lower level of effort with respect to working hard to meet expectations, exam challenge, number of homework assignments and time spent on homework, notetaking in class and review of notes outside class, attendance at help sessions, and number of missed lectures and labs; but higher effort in terms of class preparedness, and fewer tutorial absences	are similar to non-participants for the majority of items (course preparedness, working hard to meet expectations, time spent on homework, lecture and lab attendance, notetaking in class and review of notes outside of class; show higher effort in terms of number of homework assignments and tutorial attendance; and display lower effort regarding exam challenge and attendance at help sessions
Integration and Application of Knowledge	report lower levels of connection to secondary school biology, daily life and news events; higher levels of connection between lectures and labs	are statistically indistinguishable from non-participants on all items
Self-Assessed Learning and Learning Outcomes	generally report higher levels of idea integration, consideration of diverse perspectives, and self-assessed science literacy; earned grades in course significantly but only slightly lower; and found course easier	generally report higher levels of idea integration, consideration of diverse perspectives, and self-assessed science literacy; earned grades in course significantly but only slightly lower; and found course as difficult as non-participants

Course-Specific	indicate lower coursework emphasis on analysis, synthesis and application of theory; and comparable coursework emphasis on memorization, making judgment and interest level	are statistically indistinguishable from non-participants on all items
Base CLASSE Items *(shown in italics in Table 29)*	display lower item scores on 12 of the 21 items several of which appear to be associated with students finding the course easier (which was one of the project's objectives): asking questions, working with classmates outside of class, discussing ideas with instructor and others, working hard to meet expectations, number of homework assignments and time spent on homework, and course difficulty rating	report similar scores to non-participants on 12 of the 21 items, but significantly higher scores on 8 items: asking questions and participating in class discussions, integration of ideas through assignments, working with classmates in class, discussing ideas with instructor, receiving prompt feedback, number of homework assignments and time spent on homework; found course as difficult as non-participants

To the extent that BSLI was implemented primarily through the tutorial component of the courses, tutorial attendance provides at least a rough measure of the effects of intensity of project participation, as shown in Table 31. In order to account for possible differences in student motivation for attendance between the two courses, only BIOL 1222 registrants are included in the analysis; in order to isolate any participation- dependent effect of BSLI, only 2008/09 (experimental group) students were examined. Given the question wording, negative coefficients generally mean that higher tutorial attendance is associated with higher item scores. A cautionary comment is warranted: while control and experimental groups were propensity matched, no matching was performed to equalize characteristics across participation levels within the experimental group. As a result, interpretation of the regressions should take account of possible self-selection and/or predisposition bias for higher attendance. Nonetheless, the survey instrument was able to detect experimental differences on a number of items; if possible self-selection implies tautology in the results because, say, admission average has not been accounted for, this can be addressed in future interventions through a second round of matching (within-group).

Table 31: University of Western Ontario Regression Results for CLASSE Design (BIOL 1222 + BIOL 1223) - Intensity of Involvement Effects (Experimental Group Only)

	Matched Groups				
Dependent Variable	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Engagement, Experience and Perception Measures					
Asked questions of instructor or TA	0.006	-0.156	0.069	-2.24	**0.025**
Contributed to class discussions	0.011	-0.167	0.056	-2.94	**0.003**
Worked on assignment requiring integration of ideas from various sources	0.010	-0.157	0.054	-2.89	**0.004**
Included diverse perspectives in discussions or assignments	0.000	-0.004	0.053	-0.08	0.936
Came to class without completing readings or assignments	0.001	0.080	0.074	1.07	0.285
Worked with other students on assignments during course	0.005	-0.140	0.067	-2.10	**0.036**
Worked with classmates outside class on assignments	0.000	-0.019	0.065	-0.30	0.766
Tutored or taught other students	0.000	0.033	0.062	0.52	0.604

Communicated with instructor or TA via email or WebCT	0.001	-0.046	0.056	-0.83	0.409
Discussed grades with instructor or TA	0.001	-0.044	0.045	-0.97	0.331
Discussed course/reading ideas with others outside class	0.001	-0.049	0.066	-0.74	0.462
Discussed course/reading ideas with instructor or TA outside class	0.003	-0.065	0.045	-1.47	0.143
Received prompt feedback on performance from instructor or TA	0.004	-0.102	0.059	-1.71	**0.087**
Worked harder to meet instructor standards or expectations	0.013	-0.201	0.061	-3.30	**0.001**
Made connection between BIOL 1222 and secondary school biology	0.010	-0.166	0.058	-2.86	**0.004**
Made connection between BIOL 1222 lectures and labs	0.001	-0.051	0.057	-0.88	0.379
Made connection between BIOL 1222 and other courses	0.001	-0.057	0.058	-0.97	0.331
Made connection between BIOL 1222 and issues in daily life	0.001	-0.057	0.058	-0.99	0.324
Made connection between BIOL 1222 and news/media coverage	0.002	-0.070	0.056	-1.26	0.207
Coursework emphasized memorizing facts, ideas, methods	0.000	0.024	0.059	0.40	0.688
Coursework emphasized analyzing elements of an idea or theory	0.000	-0.001	0.054	-0.02	0.982
Coursework emphasized synthesizing and organizing information	0.009	-0.152	0.057	-2.70	**0.007**
Coursework emphasized making judgments about value of information or arguments	0.001	-0.039	0.061	-0.65	0.516
Coursework emphasized applying theories to practical problems	0.001	-0.057	0.063	-0.91	0.364
Exams in course challenge you to do your best work	0.000	-0.006	0.057	-0.10	0.918
Number of homework assignments per week taking 1+ hours	0.003	0.071	0.048	1.47	0.142
Frequency of spending 3+ hours per week on course outside class	0.004	-0.108	0.063	-1.71	**0.089**
Number of lecture session absences	0.072	0.537	0.068	7.84	**<.0001**
Frequency of writing notes during lectures	0.009	-0.123	0.047	-2.64	**0.009**
Frequency of reviewing notes prior to lecture	0.001	-0.172	0.057	-2.99	**0.003**
Frequency of study partnership to study for tests	0.000	0.034	0.073	0.46	0.643
Frequency of attending review or help session	0.000	-0.011	0.069	-0.16	0.870
Level of interest in course material	0.001	-0.034	0.051	-0.67	0.505
Level of comfort talking with instructor or TA	0.002	-0.730	0.060	-1.22	0.223
Level of enjoyment working with classmates in course	0.001	-0.052	0.057	-0.91	0.363
Level of difficulty of course material	0.000	-0.019	0.054	-0.36	0.722
Ease of following material in course lectures	0.000	0.020	0.051	0.38	0.701
Extent to which course contributed to writing effectively	0.000	-0.010	0.042	-0.23	0.815
Effectiveness of tutorials in helping you learn course material	0.001	-0.045	0.053	-0.85	0.397
Effectiveness of labs in helping you learn course material	0.000	0.003	0.063	-0.05	0.957
Self-rated ability searching for and finding authoritative resources	0.000	-0.044	0.060	-0.72	0.473
Self-rated ability evaluating validity of resources	0.000	-0.009	0.055	-0.17	0.869
Self-rated ability using found resources in appropriate way	0.004	-0.109	0.058	-1.87	**0.062**
Self-rated ability paraphrasing passage while avoiding plagiarism	0.004	-0.110	0.059	-1.87	**0.062**
Self-rated ability summarizing article while avoiding plagiarism	0.002	-0.069	-0.062	-1.11	0.266
Academic Outcome					
Final course grade	0.035	-3.573	0.664	-5.33	**<.0001**

Self-Assessed Science Literacy and Literacy Test Results:

The purpose of BSLI – to enhance overall science literacy – can be measured subjectively through student self-assessment of science literacy on the CLASSE instrument, and objectively using scores on the science literacy quiz. Table 32 presents mean scores for

BIOL 1222 student self-assessments of the five literacy items on the CLASSE survey. In all cases, self-assessed literacy levels were significantly higher in the experimental group. If students were informed of the revised focus and enhanced content of the courses in 2008/09 (and the author is not aware if they were), the results would clearly reflect this fact and would have to be heavily discounted.

Table 32: University of Western Ontario CLASSE-Based Student Self-Assessment of Science Literacy – BIOL 1222

Item (5-point scales from 1 (poor) to 5 (excellent))	2008 Mean	2009 Mean	p-value
Searching for and finding authoritative Biology resources	2.86	3.14	<.0001
Evaluating the validity of Biology resources	2.83	3.11	<.0001
Using the information obtained from Biology resources in an appropriate way	3.09	3.31	<.0001
Paraphrasing a Biological passage in writing while avoiding plagiarism	3.13	3.37	<.0001
Summarizing a Biological article in writing while avoiding plagiarism	3.12	3.32	<.0001
	(n=887)	(n=763)	

A more valid approach to measuring literacy development involves a comparison of science literacy test scores for the students who volunteered to take the test in 2008 and 2009. Because a relatively small number of students took the test (about 60 in each year), and because the characteristics of those self-selecting to do it varied, it was necessary to explore test group differences before comparing test grades. Table 33 indicates that while the two test groups differed significantly only on their Grade 12 Biology grades, they are sufficiently dissimilar on other characteristics to justify examining test scores for each subgroup in addition to scores for the group overall. Table 34 below presents test scores for each year for strata within the overall test groups. While the experimental group showed consistently higher test scores, the differences are significant only with respect to first-year students. Multiple regression analysis and/or propensity matching along with much larger sample sizes and a more detailed test to provide greater score differentiation would provide greater certainty for what can currently be described as a potentially promising but weak relationship. Further work in this area might provide a clearer link between engagement/experience and learning outcomes.

Table 33: University of Western Ontario Literacy Test Group Comparability 2008/2009			
Basis of Comparison	**2008**	**2009**	**Significance**
Mean Grade 12 Biology grade	87.7	88.9	0.125
Year of study (percentage of all students in first year)	96.8	91.5	0.104
Mean student age	18.1	18.5	0.179
Immigration status (percentage of all students with domestic origins)	96.0	88.2	0.063
Gender (percentage of students who are female)	61.9	66.1	0.316
Basis of admission (percentage of students admitted directly from secondary school)	90.5	93.2	0.295
Academic load (percentage of students studying full-time)	100.0	100.0	n/a
Mean BIOL 1222 grade	77.0	80.1	0.041
	(n=63)	(n=59)	

Table 34: University of Western Ontario Science Literacy Test Scores by Student Group			
	Mean Score		
Group	**2008**	**2009**	**Significance**
Registered in first year	4.15	4.63	0.035
Domestic origin	4.31	4.48	0.330
Male	4.25	4.60	0.218
Female	4.15	4.46	0.176
Admitted direct from secondary school	4.23	4.53	0.138
Overall	4.19	4.51	0.113
	(n=26 to 63)	(n=20 to 59)	

Survey Non-Response Bias:

Survey non-response associated with consistent and relevant respondent/non-respondent differences limits the extent to which survey samples can be used to generalize over the entire population. And while assessment is enhanced by the availability of multiple survey tools having common respondents, it is possible that non-response bias is increased for dual-survey response relative to single-survey response. The Western project provides an opportunity to examine the characteristics of respondents for different survey response behaviours in relation to the characteristics of the overall class. Table 35 presents the

characteristics for two key survey response groups: those who responded to neither NSSE nor CLASSE, and those who responded to both.

Table 35: University of Western Ontario Survey Non-Response Patterns

Group	% Responding to Neither Survey	% Responding to Both Surveys
Registered in first year	32.7	18.0
Registered in second year	55.3	3.6
Domestic origin *	31.3	27.3
Male	42.2	11.2
Female	29.5	20.4
Admitted direct from secondary school	33.3	17.1
Admitted indirect from secondary school	45.1	13.0
Full-time	34.3	16.7
Part-time	61.0	10.1
Faculty (Science or Health Sciences)	33.7	17.1
Faculty (Other)	51.2	8.0
Overall	34.8	16.6

Characteristic	Average for Those Responding to Neither Survey	Average for Those Responding to Both Surveys
Overall average at Western	68.2	77.0
Overall average in Grade 12	85.5	88.5
Average grade in Grade 12 Biology	83.3	86.8
Average Age	18.2	18.4
Average grade in BIOL 1222 or 1223	68.2	76.2
* high (50%) missing data rate		

The NSSE and CLASSE analysis for Western excludes, of course, students who responded to neither survey, and hence, somewhat over-represents first-year students, females, direct entries from secondary school, full-time students, those within the Science and Health Sciences Faculties, and students with higher grades in secondary school and at Western. (Though not shown in the table, students responding to just one of the surveys displayed the attributes of dual-responders though to a lesser extent.) While these survey response patterns do not seriously limit the applicability of the assessment analyses, it is important that they be taken into account in survey administration planning (i.e., by recognizing that, in this instance, analysis requiring dual-response reduces sample size by about 50 per cent) and that analysis of intervention impact based on the surveys acknowledges that the impact may or may not apply to non-respondents.

4.6.5 Summary

The quantity of data available for the Western project, and the integrated structure of those data permitted several additional analyses that provided considerable input into the

appropriate use and interpretation of the various measurement tools. Despite the large successive cohorts of survey respondents, pre-match differences existed that were corrected through the use of propensity matching; this avoided the necessity of multivariate modeling in which group differences would have been included as covariates. However, because the control and experimental group populations were about the same size, matching did result in some reduction in experimental group counts (between two per cent and 28 per cent) that could easily be accommodated given the large sample sizes.

Analysis of NSSE results indicates that BSLI project participation is not reliably captured in NSSE item scores. The upward trend in NSSE scores overall carries over to higher experimental group scores on some items, suggesting a background rather than experimental effect; and at the same time, BSLI participants display year-over-year changes on other items that are inconsistent with the corresponding changes in NSSE over time.

The CLASSE-based basic participation analysis indicates that experimental effects were detected, particularly given that age, admission average and other confounding factors were controlled through propensity matching (in the basic participation case). Even in the intensity of participation analysis, the consistent coefficient signs suggest BSLI impacts were successfully detected, though further analysis will be required to differentiate between self-selection/predisposition and experimental factors. The potentially "troubling" finding that BSLI students participated in fewer peer and faculty interactions and expended lower levels of academic effort may indicate that the project provided the support that might otherwise have been achieved through a higher level of interaction and effort; in any case, this apparently unintended effect warrants further study. The survey instrument appears to have been sensitive enough to detect engagement and experience differences even between the two Biology courses, with somewhat stronger effects in BIOL 1222, the course containing students with higher preparedness. While CLASSE-measured experimental effects cannot be formally placed within the generally rising (2004 to 2008) engagement background, the results appear robust for at least two reasons. The first is the inconsistency between successive NSSE and CLASSE results with respect to item significance and direction. The second is the preponderance of evidence in the CLASSE results in terms of the number of significant individual items, consistency of item significance within general themes, and the significance of BSLI-specific items for which there is little support in the engagement background.

Literacy test score results provide some initial evidence that the experimental group achieved a higher level of science literacy, but the results do not appear to have translated into final course grades (which show a slight decline in BIOL 1222 and no difference in BIOL 1223). A more detailed literacy test (perhaps as a component of the final exam), much larger test groups, and controls for dissimilar test groups over time could be considered in future. The implications of significantly higher student self-assessments of literacy in the experimental groups cannot be determined here.

The likelihood that better prepared, better performing students in first-year will be somewhat over-represented in survey samples argues for some caution when inferring to the entire class, and may warrant a minor "scaling down" of projected BSLI effects. However, to the extent that survey non-respondents are disproportionately located in

strata that constitute only a small proportion of total class registrants, the results remain viable.

4.7 Ryerson University (Improving Writing Skills in Selected Academic Programs)

4.7.1 Intervention Description and Background

Ryerson University is located in downtown Toronto in the centre of Canada's most culturally and ethnically diverse city. Ryerson's student body contains a significant proportion (20 per cent) of first-generation students, and many of these are recent Canadians or the children of recent Canadians. About two-thirds of Ryerson's NSSE respondents in 2006 indicated that the University had contributed "quite a bit" or "very much" to the development of their writing skills; in another survey, 28 per cent reported that the University's contribution to these skills was "excellent". Ryerson was concerned that the University's perceived contribution was not higher, particularly given the views of several faculty members concerning the English skills of new students. The primary objective of Ryerson's "Writing for Success" initiative is to improve the writing skills of first-year students in the Faculty of Community Services (which contains eight professional undergraduate programs). The initiative was a joint effort of the Learning and Teaching Office, the Writing Centre, the Experiential Learning Centre and the eight schools within the Faculty. The pilot initiative is based on a model developed by Cadwallader and Scarboro in which writing is taught as an integral part of courses, and the design was built on an effective writing course previously developed at Ryerson. The initiative has several components:

- A curricular framework was developed that integrated writing skills development into one core/ required course within each of the eight academic programs;
- The courses were redesigned to achieve the dual objective of mastery of content and effective writing skills development;
- Writing support was provided by specially trained teaching assistants in weekly small group tutorials in which students presented and discussed their own written work and received timely feedback on their writing skills;
- Sessions were designed so as to develop writing skills across all levels of writing competency, and not just for those with deficient writing skills.

The initiative is grounded in Ryerson's academic plan for 2008 - 2013, *Shaping our Future*. One of the plan's five main priorities – student engagement and success – includes the specific objective of improving English language skills and writing.

4.7.2 Context Provided by NSSE Administrations

Ryerson administered NSSE in 2005, 2006 and 2008; across all three administrations, first-year benchmark scores university-wide remained virtually constant, with less than three per cent movement on all benchmarks (Figure 13). The university previously

identified 15 core NSSE questions through a broad campus consultation process for which the University wished to improve its performance, and these items have been incorporated into academic planning and budgeting (Table 36). Three of these 15 items – questions and discussions in class, the provision of support for academic success and the development of writing skills – are at the heart of "Writing for Success"; in addition, 10 additional NSSE items selected by Ryerson provided both the rationale for, and testable outcomes of the initiative.

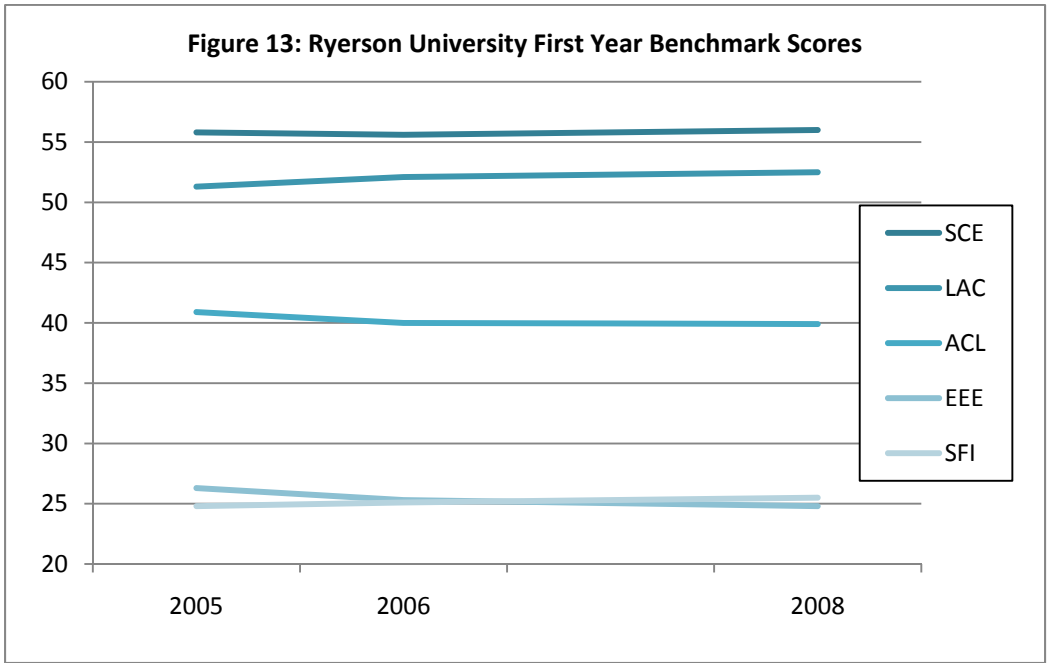


Table 36: Ryerson University First-Year Selected NSSE Item (Dependent Variable) Values			
NSSE Item	**2006**	**2008**	**% Change**
	means		
From the University's Set of 15 Core Questions			
Asked questions in class/contributed to class discussions	2.44	2.40	-1.64%
University provides the support you need to succeed academically	2.83	2.88	1.77%
University's contribution to development of writing skills	2.76	2.74	-0.72%
Relevant to the Assessment of "Writing for Success"			
Prepared 2+ drafts of assignment before turning it in	2.36	2.33	-1.27%
Worked on paper/project requiring integrating ideas/info from various sources	3.16	3.12	-1.27%
Worked with other students on projects during class	2.35	2.34	-0.43%
Discussed ideas from readings/classes with others outside of class	2.74	2.71	-1.09%
Course emphasis on synthesizing and organizing ideas	2.77	2.80	1.08%
Course emphasis on making judgments about the value of information	2.75	2.79	1.45%
Number of written reports 5 - 19 pages	2.58	2.58	0.00%
Number of written reports < 5 pages	2.53	2.47	-2.37%
Hours per week spent preparing for class (categorized: not actual hours)	4.27	4.31	0.94%
	percentages		
Identified "expanding or improving quality of academic support services" (Ontario consortium question: select 2 from list of 10)	24%	22%	-8.33%

4.7.3 Assessment Design

Ryerson employed a successive cohort design using 2007/08 students (control group) and 2008/09 students (experimental group) in the eight courses. A 100 per cent first-year NSSE sample was administered throughout the Faculty in both years. While tutorial participation attendance and involvement are likely to vary, no intensity of involvement measure was developed. All registrants in the eight courses were subject to the intervention, so there was no opportunity for self-selection bias. Experimental group subjects were propensity matched to control group subjects using full-/part-time status, gender, basis of admission (direct or indirect from secondary school) and domestic/ international status (see Table 37). The missing data rates for student age and secondary school grade average were too high (40 per cent+) for these variables to be included in the matching process; however, both were measured before and after matching and showed no significant differences in either case. The relatively small number of pre-match differences was eliminated post-

Table 37: Ryerson University Propensity Matching Results									
		Pre-Match				Post-Match			
Course/School	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
Early Childhood Education	Full-/Part-Time	-0.09	0.52	0.03	0.857	0.00	0.64	0.00	1.000
	Gender	0.08	1.24	0.00	0.952	exact match			
	Basis of Admission	0.38	0.48	0.61	0.433	0.00	0.60	0.00	1.000
	Domestic/Int'l	-1.21	0.80	2.30	0.130	0.00	1.44	0.00	1.000
	Age *	*-0.15*	*0.11*	*2.07*	*0.150*	*-0.20*	*0.17*	*1.31*	*0.253*
	Secondary School Average *	*0.14*	*0.17*	*0.66*	*0.416*	*0.12*	*0.10*	*1.38*	*0.241*
		(n=35 experimental, n=65 control)				(n=32 experimental, n=32 control)			
Child and Youth Care	Full-/Part-Time	0.69	0.95	0.53	0.465	exact match			
	Gender	-13.69	304.24	0.00	0.964	exact match			
	Basis of Admission	0.90	0.51	3.09	0.079	0.00	0.62	0.00	1.000
	Domestic/Int'l	-0.06	0.81	0.00	0.944	0.00	1.05	0.00	1.000
	Age *	*-0.01*	*0.07*	*0.04*	*0.833*	*-0.01*	*0.11*	*0.01*	*0.921*
	Secondary School Average *	*0.98*	*1.08*	*0.82*	*0.366*	*0.56*	*1.32*	*0.18*	*0.670*
		(n=30 experimental, n=38 control)				(n=21 experimental, n=21 control)			
Occupational and Public Health	Full-/Part-Time	-1.04	1.15	0.82	0.365	exact match			
	Gender	0.30	0.57	0.28	0.600	0.00	0.68	0.00	1.000
	Basis of Admission	-0.22	0.56	0.16	0.689	0.00	0.63	0.00	1.000
	Domestic/Int'l	-13.03	310.10	0.00	0.966	exact match			
	Age *	*-0.10*	*0.05*	*3.63*	*0.057*	*-0.09*	*0.05*	*2.71*	*0.100*
	Secondary School Average *	*0.67*	*1.29*	*0.27*	*0.604*	*0.67*	*1.29*	*0.27*	*0.604*
		(n=23 experimental, n=35 control)				(n=22 experimental, n=22 control)			
Nutrition and Food	Full-/Part-Time	-1.35	0.68	3.95	0.047	0.00	0.86	0.00	1.000
	Gender	1.03	1.14	0.82	0.366	0.00	1.44	0.00	1.000
	Basis of Admission	0.30	0.43	0.48	0.490	0.00	0.50	0.00	1.000
	Domestic/Int'l	-1.48	1.10	1.79	0.181	0.00	1.44	0.00	1.000
	Age *	*-0.03*	*0.06*	*0.37*	*0.541*	*-0.05*	*0.06*	*0.71*	*0.400*
	Secondary School Average *	*2.79*	*1.58*	*3.13*	*0.077*	*2.57*	*1.57*	*2.67*	*0.103*
		(n=36 experimental, n=54 control)				(n=32 experimental, n=32 control)			
Midwifery	Full-/Part-Time	-0.92	1.64	0.31	0.577	0.00	1.73	0.00	1.000
	Gender	exact match				exact match			
	Basis of Admission	exact match				exact match			
	Domestic/Int'l	-10.85	293.23	0.00	0.970	exact match			

	*Age **	*0.60*	*0.49*	*1.49*	*0.223*	*0.51*	*0.42*	*1.47*	*0.225*
	*Secondary School Average **	*-0.92*	*1.64*	*0.31*	*0.577*	*0.00*	*1.73*	*0.00*	*1.000*
		(n=3 experimental, n=6 control)				(n=3 experimental, n=3 control)			
Nursing	Full-/Part-Time	15.01	153.41	0.01	0.922	exact match			
	Gender	-0.41	0.58	0.50	0.478	0.00	1.44	0.00	1.000
	Basis of Admission	0.17	0.38	0.21	0.649	0.00	0.57	0.00	1.000
	Domestic/Int'l	-1.14	0.49	5.40	0.020	0.00	0.77	0.00	1.000
	*Age **	*0.03*	*0.05*	*0.39*	*0.534*	*0.06*	*0.07*	*0.87*	*0.652*
	*Secondary School Average **	*0.13*	*0.13*	*1.06*	*0.304*	*0.12*	*0.21*	*0.33*	*0.565*
		(n=57 experimental, n=97 control)				(n=25 experimental, n=25 control)			
Urban and Regional Planning	Full-/Part-Time	0.88	0.87	1.01	0.316	0.00	1.46	0.00	1.000
	Gender	0.02	0.58	0.00	0.975	0.00	0.73	0.00	1.000
	Basis of Admission	-0.32	0.67	0.23	0.634	0.00	0.82	0.00	1.000
	Domestic/Int'l	-12.79	331.15	0.00	0.969	exact match			
	*Age **	*0.25*	*0.20*	*1.67*	*0.196*	*0.00*	*0.26*	*0.00*	*1.000*
	*Secondary School Average **	*0.11*	*0.09*	*1.40*	*0.236*	*0.11*	*0.12*	*0.81*	*0.368*
		(n=18 experimental, n=39 control)				(n=16 experimental, n=16 control)			
Social Work	Full-/Part-Time	0.27	1.02	0.07	0.793	0.00	1.03	0.00	1.000
	Gender	0.48	0.74	0.42	0.516	0.00	0.85	0.00	1.000
	Basis of Admission	0.83	0.43	3.78	0.052	0.00	0.48	0.00	1.000
	Domestic/Int'l	0.32	0.53	0.37	0.545	0.00	0.68	0.00	1.000
	*Age **	*0.01*	*0.03*	*0.11*	*0.743*	*0.06*	*0.04*	*2.44*	*0.119*
	*Secondary School Average **	*0.12*	*0.11*	*1.06*	*0.304*	*0.10*	*0.08*	*1.50*	*0.220*
		(n=41 experimental, n=53 control)				(n=35 experimental, n=35 control)			
* Age and Secondary School Average were compared pre- and post-match but were not used in matching due to the frequency of missing data									

match; however secondary school average (Nutrition and Food program) and age (Social Work program) remained near-significant after matching. Propensity matching was performed, and all analysis was undertaken, on a course-by-course basis: the academic programs themselves, student characteristics, and the engagement background (i.e., pre-experiment NSSE scores) were sufficiently different across programs that a pooled analysis for the entire Faculty was considered but abandoned. This resulted in the elimination of one course from the analysis because of very low sample size, and usable but relatively small sample sizes (n = 32 – 70) for the remaining courses.

Ryerson also administered surveys at the beginning and end of each course to measure student-self assessment of writing skills and their expectations and perceptions toward their university experience that will be used on-site for Ryerson's own qualitative assessment activity.

4.7.4 Assessment Results

The results of the course-by-course analysis are presented in Table 38. Findings from the previously reported projects suggest that course-specific engagement changes appear not to be captured in, and may in fact be inconsistent with overall engagement scores. The inconsistency of results across the seven Ryerson programs and the very limited number of significant coefficients suggest the same conclusion applies here.

Table 38: Ryerson University Regression Results for Engagement and Academic Outcomes										
	Early Childhood Education					Child and Youth Care				
Dependent Variable	R^2	B-Est	SE	t-score	p-value	R^2	B-Est	SE	t-score	p-value
Engagement Measures										
Asked questions in class/contributed to class discussions	0.038	-0.357	0.231	-1.54	0.128	0.011	-0.171	0.264	-0.65	0.521
University provides the support you need to succeed academically	0.008	0.156	0.224	0.70	0.488	0.039	-0.291	0.234	-1.24	0.222
University's contribution to development of writing skills	0.001	-0.035	0.190	-0.18	0.857	0.016	0.200	0.251	0.80	0.431
Prepared 2+ drafts of assignment before turning it in	0.001	-0.063	0.241	-0.26	0.796	0.027	-0.319	0.305	-1.05	0.302
Worked on paper/project requiring integrating ideas/info from various sources	0.061	0.313	0.156	2.00	**0.050**	0.126	-0.500	0.214	-2.34	**0.025**
Worked with other students on projects during class	0.035	0.313	0.208	1.50	0.138	0.031	0.352	0.315	1.12	0.271
Discussed ideas from readings/classes with others outside of class	0.042	0.344	0.208	1.65	0.104	0.087	-0.500	0.260	-1.92	**0.062**
Course emphasis on synthesizing and organizing ideas	0.001	0.063	0.232	0.27	0.789	0.006	-0.131	0.277	-0.47	0.639
Course emphasis on making judgments about the value of information	0.003	0.094	0.235	0.40	0.692	0.001	0.048	0.240	0.20	0.844
Number of written reports 5 - 19 pages	0.004	-0.094	0.180	-0.52	0.605	0.005	-0.930	0.209	-0.44	0.660
Number of written reports < 5 pages	0.007	-0.188	0.281	-0.67	0.507	0.017	0.240	0.293	0.82	0.416
Hours per week spent preparing for class (categorized: not actual hours)	0.027	-0.569	0.433	-1.31	0.194	0.009	0.255	0.420	0.61	0.547
Identified "expanding or improving quality of academic support services" (Ontario consortium question: select 2 from list of 10) *		1.300	0.735	3.14	**0.076**		2.459	1.120	4.82	**0.028**
LAC Benchmark	0.004	-1.712	3.333	-0.51	0.609	0.009	-2.140	3.650	-0.59	0.561
ACL Benchmark	0.019	-4.217	3.832	-1.10	0.275	0.026	-4.920	4.848	-1.01	0.317
SCE Benchmark	0.031	7.765	5.560	1.40	0.168	0.000	0.334	7.357	0.05	0.964
Academic Outcomes										
Grade in course	0.032	-0.625	0.436	-1.43	0.156	0.042	-1.190	0.896	-1.33	0.191
Overall academic standing (good standing/not) *	insufficient "n" in bad standing category						0.470	0.691	0.46	0.496
Fall GPA	0.005	0.093	0.165	0.57	0.574	0.007	-0.162	0.299	-0.54	0.590
Winter GPA	0.013	-0.226	0.245	-0.92	0.361	0.000	0.021	0.278	0.08	0.939
* modeled using logistic regression and Wald chi-square (not t-score)										

table continued next page …

Two issues highlighted by the table are worth noting, however:

- Only four of the NSSE engagement items or grades outcomes are significant in more than two of the seven programs: working on a paper that required integration of ideas (four programs), number of papers of fewer than 5 pages (three programs), identification of academic support service (one of the Ontario consortium questions – three programs), and course grade (three programs);

Table 38 (continued): Ryerson University Regression Results for Engagement and Academic Outcomes

	Occupational and Public Health					Nutrition and Food				
Dependent Variable	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Engagement Measures										
Asked questions in class/contributed to class discussions	0.001	0.045	0.242	0.19	0.852	0.047	0.355	0.207	1.71	**0.092**
University provides the support you need to succeed academically	0.018	-0.227	0.256	-0.89	-0.379	0.000	0.031	0.199	0.16	0.875
University's contribution to development of writing skills	0.011	0.205	0.306	0.67	0.508	0.058	0.419	0.219	1.19	**0.060**
Prepared 2+ drafts of assignment before turning it in	0.021	0.288	0.307	0.94	0.355	0.001	0.057	0.232	0.25	0.805
Worked on paper/project requiring integrating ideas/info from various sources	0.271	0.909	0.230	3.95	**0.000**	0.006	0.117	0.187	0.63	0.533
Worked with other students on projects during class	0.031	-0.273	0.233	-1.17	0.249	0.021	-0.220	0.193	-1.14	0.260
Discussed ideas from readings/classes with others outside of class	0.001	0.045	0.254	0.18	0.859	0.007	-0.125	0.187	-0.67	0.506
Course emphasis on synthesizing and organizing ideas	0.015	0.227	0.289	0.79	0.436	0.003	-0.094	0.206	-0.46	0.651
Course emphasis on making judgments about the value of information	0.026	0.305	0.292	1.04	0.302	0.005	-0.103	0.192	-0.54	0.594
Number of written reports 5 - 19 pages	0.103	0.318	0.145	2.20	**0.034**	0.007	0.125	0.192	0.65	0.519
Number of written reports < 5 pages	0.027	0.227	0.210	1.08	0.286	0.177	0.531	0.145	3.65	**0.001**
Hours per week spent preparing for class (categorized: not actual hours)	0.000	-0.091	0.527	-0.17	0.864	0.004	0.227	0.481	0.47	0.638
Identified "expanding or improving quality of academic support services" (Ontario consortium question: select 2 from list of 10) *		0.238	0.668	0.13	0.721		0.519	0.689	0.57	0.452
LAC Benchmark	0.045	5.637	4.010	1.41	0.167	0.018	3.309	3.129	1.06	0.294
ACL Benchmark	0.000	0.450	4.207	0.11	0.914	0.001	0.713	3.000	0.24	0.813
SCE Benchmark	0.002	-1.540	5.850	-0.26	0.794	0.036	7.151	4.689	1.52	0.132
Academic Outcomes										
Grade in course	0.001	0.136	0.937	0.15	0.885	0.108	1.094	0.400	2.73	**0.008**
Overall academic standing (good standing/not) *		0.000	0.879	0.00	1.000		0.000	0.858	0.00	1.000

Fall GPA	0.014	0.220	0.287	0.77	0.447	0.061	0.507	0.255	1.99	**0.052**
Winter GPA	0.008	0.118	0.203	0.58	0.565	0.057	-0.421	0.218	-1.92	**0.059**
	Nursing					**Urban and Regional Planning**				
Dependent Variable	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Engagement Measures										
Asked questions in class/contributed to class discussions	0.008	0.160	0.254	0.63	0.531	0.046	0.250	0.208	1.20	0.239
University provides the support you need to succeed academically	0.009	0.152	0.246	0.62	0.541	0.122	0.592	0.295	2.01	**0.054**
University's contribution to development of writing skills	0.000	0.000	0.242	0.00	1.000	0.000	0.000	0.342	0.00	1.000
Prepared 2+ drafts of assignment before turning it in	0.021	-0.308	0.309	-1.00	0.324	0.133	0.750	0.351	2.14	**0.041**
Worked on paper/project requiring integrating ideas/info from various sources	0.005	-0.080	0.170	-0.47	0.641	0.051	0.250	0.198	1.26	0.216
Worked with other students on projects during class	0.006	-0.153	0.291	-0.53	0.601	0.005	-0.125	0.315	-0.40	0.694
Discussed ideas from readings/classes with others outside of class	0.014	0.205	0.252	0.81	0.420	0.007	-0.134	0.292	-0.46	0.650
Course emphasis on synthesizing and organizing ideas	0.025	-0.251	0.236	-1.07	0.292	0.008	0.129	0.266	0.49	0.631
Course emphasis on making judgments about the value of information	0.045	-0.337	0.231	-1.46	0.153	0.039	0.313	0.285	1.10	0.281
Number of written reports 5 - 19 pages	0.071	-0.413	0.217	-1.90	**0.063**	0.019	0.188	0.243	0.77	0.447
Number of written reports < 5 pages	0.121	0.533	0.210	2.54	**0.014**	0.100	-0.500	0.273	-1.83	**0.078**
Hours per week spent preparing for class (categorized: not actual hours)	0.007	-0.306	0.548	-0.56	0.579	0.067	0.937	0.640	1.47	0.153
Identified "expanding or improving quality of academic support services" (Ontario consortium question: select 3 from list of 15) *		0.163	0.713	0.05	0.820		0.480	0.991	0.23	0.628
LAC Benchmark	0.002	-1.188	4.139	-0.29	0.775	0.020	3.032	0.387	0.78	0.439
ACL Benchmark	0.003	-1.301	3.913	-0.33	0.741	0.014	-2.698	4.114	-0.66	0.517
SCE Benchmark	0.020	5.529	5.847	0.95	0.348	0.082	11.028	6.750	1.63	0.113
Academic Outcomes										
Grade in course	0.069	-0.920	0.487	-1.89	**0.065**	0.212	2.375	0.836	2.84	**0.008**
Overall academic standing (good standing/not) *		-10.746	211.100	0.00	0.959		-1.157	0.929	1.55	0.213
Fall GPA	0.192	1.226	0.363	3.38	**0.001**	0.065	0.379	0.262	1.44	0.159
Winter GPA	0.001	0.038	0.168	0.23	0.820	0.096	0.492	0.275	1.79	**0.084**
* modeled using logistic regression and Wald chi-square (not t-score)										

table continued next page …

however, coefficient signs are generally a mix of positive and negative for each of these items. The one exception is the consortium question, scores for which are consistently higher in the three programs where there was a difference;

- The items that might be expected to show a difference – institutional contribution to writing skill, preparing 2+ drafts of a paper, working with other students in class and the number of papers of less than 5 pages and/or 5–19 pages – did not show any consistent pattern across programs. No two of the programs show similar results across more than one or two items.

Table 38 (continued): Ryerson University Regression Results for Engagement and Academic Outcomes

Dependent Variable	Social Work				
	R^2	B-Est	SE	t-score	p-value
Engagement Measures					
Asked questions in class/contributed to class discussions	0.028	-0.303	0.219	-1.38	0.172
University provides the support you need to succeed academically	0.002	0.071	0.212	0.34	0.737
University's contribution to development of writing skills	0.001	-0.048	0.220	-0.22	0.830
Prepared 2+ drafts of assignment before turning it in	0.003	0.118	0.267	0.44	0.661
Worked on paper/project requiring integrating ideas/info from various sources	0.059	-0.252	0.123	-2.05	**0.045**
Worked with other students on projects during class	0.006	0.109	0.167	0.66	0.514
Discussed ideas from readings/classes with others outside of class	0.000	0.029	0.190	0.15	0.877
Course emphasis on synthesizing and organizing ideas	0.007	0.156	0.227	0.69	0.495
Course emphasis on making judgments about the value of information	0.005	0.126	0.233	0.54	0.591
Number of written reports 5 - 19 pages	0.022	0.247	0.207	1.20	0.236
Number of written reports < 5 pages	0.002	-0.070	0.191	-0.37	0.715
Hours per week spent preparing for class (categorized: not actual hours)	0.000	-0.032	0.391	-0.08	0.935
Identified "expanding or improving quality of academic support services" (Ontario consortium question: select 2 from list of 10) *		1.674	0.717	5.45	**0.020**
LAC Benchmark	0.030	4.703	3.360	1.40	0.166
ACL Benchmark	0.000	-0.212	3.468	-0.06	0.951
SCE Benchmark	0.013	4.577	5.052	0.91	0.368

Academic Outcomes					
Grade in course	0.001	-0.176	0.761	-0.23	0.817
Overall academic standing (good standing/not) *		0.216	0.659	0.11	0.743
Fall GPA	0.027	0.424	0.308	1.38	0.173
Winter GPA	0.009	0.189	0.246	0.77	0.444
* modeled using logistic regression and Wald chi-square (not t-score)					

4.7.5 Summary

Program-level differences in student characteristics and engagement backgrounds made program-specific analysis rather than pooled analysis appropriate. Even if higher NSSE response rates can be achieved, fairly small program populations are a reality at Ryerson that present assessment challenges. With one exception (Nursing), propensity matching was accomplished without substantial reductions in the experimental group sizes (14 – 30 per cent).

The Ryerson analysis provides further evidence that modest (i.e., single course-based) engagement interventions require an assessment instrument that is more sensitive than NSSE. The situation is somewhat more confusing in this instance because of the program-level inconsistencies in NSSE score changes that may reveal something about program behaviour but that might just as easily be random effects. It is likely – based on the Guelph, Queen's and Western Ontario experience – that a CLASSE-type instrument would have provided more precision in this case, and it is fortunate that Ryerson administered its own post-project participant survey to support qualitative assessment. A CLASSE-based assessment would also justify development of an intensity of involvement/participation measure. If undertaken again, Ryerson's approach might also benefit from formal literacy assessment testing in order to provide a formal outcome measure.

4.8 Wilfrid Laurier University (Peer Learning Program for Information Literacy, Research and Writing Skills)

4.8.1 Intervention Description and Background

In the face of an increasingly diverse student population, faculty and academic administrators at Wilfrid Laurier University (WLU) had become increasingly concerned about an apparent decline in fundamental academic skills in first-year entrants over a several year period, and believed these deficiencies were associated with student attrition and a lower level of academic performance. Prior to this project, the Study Skills and Supplementary Instruction Centre and the Writing Centre had developed several peer mentoring programs for students to assist in the transition to university. The peer learning

program analyzed here was developed for students in two writing-intensive first-year courses with a combined enrolment of about 500 in order to improve information literacy and research and writing skills. Graduate and senior undergraduate TA's selected on the basis of their academic performance and leadership skills, led the program sessions after completing an intensive 4-day training session; they also participated in ongoing debriefing during the term. This and other peer learning programs have been developed in response to Wilfrid Laurier's *Century Plan 2005 – 2011*, one objective of which is to improve services to support and enhance the student learning experience.

Session content was both general and specific to the courses (based on faculty input) and focused on

- academic writing;
- writing thesis statements and literature reviews;
- citation practices and plagiarism;
- research and critical thinking strategies and improving research strategies;
- oral presentation skills;
- examining and evaluating information;
- formulating search strategies and techniques.

4.8.2 Context Provided by NSSE Administrations

WLU achieved highly consistent NSSE benchmark scores in its 2006 and 2008 NSSE administrations with the exception of the SFI benchmark, which increased about 10 per cent (see Figure 14). The various individual items the University identified as useful measures of the project's impacts also showed considerable consistency over the two years, with differences of less than +/-five per cent (Table 39).

4.8.3 Assessment Design

The project utilized a successive cohort design based on a 100 per cent sample of students in the two courses in 2008 (control group) and 2009 (experimental group). As with the other interventions, NSSE response records were merged with information from the student records system to facilitate propensity matching. Matching was performed using gender, entering average and basis of admission for both courses (Table 40). The relatively equal sizes of the experimental groups and candidate control groups in 2008 and 2009 resulted in insufficient matches being found for all experimental records using just the 2008 control pool; as a result, the 2006 records of students in each of the two courses were added to the control pool. All experimental records were then matched with a 1:1 match ratio. One near-significant pre-match difference was successfully eliminated post-match.

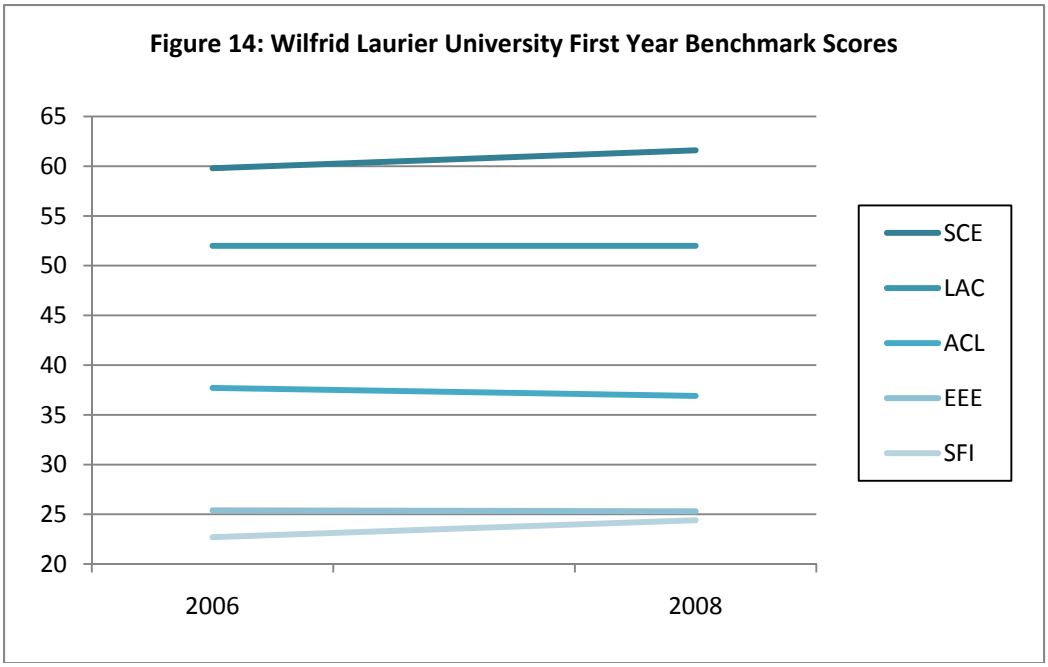


Table 39: Wilfrid Laurier University First-Year Selected NSSE Item (Dependent Variable) Means			
NSSE Item	**2006**	**2008**	**% Change**
Prepared 2+ drafts of paper before handing it in	2.28	2.21	-3.07%
Worked on paper requiring integration of ideas from various sources	3.11	3.19	2.57%
Worked with other students on projects during class	1.97	1.91	-3.05%
Worked with classmates outside of class to prepare assignments	2.73	2.69	-1.47%
Put together ideas from different courses when preparing assignments	2.67	2.71	1.50%
Used electronic medium to discuss or complete an assignment	2.89	2.82	-2.42%
Worked harder to meet instructor's expectations	2.41	2.53	4.98%
Discussed ideas with others outside of class	2.79	2.81	0.72%
Coursework emphasis on analysis	3.06	3.16	3.27%
Coursework emphasis on synthesis	2.72	2.74	0.74%
Coursework emphasis on applying theories to practical problems	3.06	3.03	-0.98%
Learned something that changed the way you understand an issue	2.89	2.89	0.00%
Hours per week spent preparing for class	4.09	4.10	0.24%
Extent to which institution provides support for academic success	3.01	3.14	4.32%
Institutional contribution to writing clearly and effectively	2.88	2.95	2.43%
Institutional contribution to thinking critically and analytically	3.23	3.30	2.17%
Institutional contribution to working effectively with others	2.90	2.95	1.72%
Institutional contribution to working effectively on your own	3.01	3.10	2.99%

Table 40: Wilfrid Laurier University Propensity Matching Results									
		Pre-Match				Post-Match			
Course	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
Course #1	Gender	-0.11	0.27	0.17	0.680	0.00	0.33	0.00	1.000
	Entering average	0.00	0.00	0.00	0.960	0.00	0.01	0.21	0.650
	Basis of admission	-0.68	0.40	2.88	0.090	0.00	0.46	0.00	1.000
		(n=73 experimental, n=243 control)				(n=73 experimental, n=73 control)			
Course #2	Gender	12.49	877.70	0.00	0.990	not used (insufficient # of one gender)			
	Entering average	0.00	0.01	0.02	0.900	0.00	0.01	0.02	0.900
	Basis of admission	0.00	0.86	0.00	1.000	0.00	0.86	0.00	1.000
		(n=34 experimental, n=100 control)				(n=34 experimental, n=34 control)			

An additional survey was also administered to participating students at the end of term in Course #2. The results indicated overall satisfaction with most aspects of the skills development sessions and in particular with the multiple drafts requirement for preparing/submitting papers, and a strong preference for individual rather than group consultations on writing assignments.

4.8.4 Assessment Results

The uncertainty of the university environment mentioned elsewhere in this report has relevance in the WLU project as well. It was necessary for project personnel to find a replacement for Course #1 fairly early on in the planning process because of staffing changes; implementation of the intervention in the replacement course did not go as smoothly as anticipated. The intervention was successfully completed in Course #2 according to design, however, and the results are presented in Table 41. Course #1 year-over-year engagement results are also presented to informally observe the stability of NSSE scores in a course that was largely unchanged, and to explore changes in a stable course environment in relation to changes in the university-wide engagement background.

Table 41: Wilfrid Laurier University Regression Results for Engagement and Academic Outcomes										
	Course #1					Course #2				
Dependent Variable	R^2	B-Est	SE	t-score	p-value	R^2	B-Est	SE	t-score	p-value
Engagement Measures										
Prepared 2+ drafts of paper before handing it in	0.016	-0.159	0.105	-1.52	0.131	0.012	-0.144	0.162	-0.89	0.377

Worked on paper requiring integration of ideas from various sources	0.003	0.040	0.065	0.61	0.544	0.074	0.191	0.085	2.26	**0.028**
Worked with other students on projects during class	0.011	-0.113	0.090	-1.26	0.211	0.001	0.043	0.155	0.28	0.782
Worked with classmates outside of class to prepare assignments	0.000	-0.006	0.093	-0.06	0.949	0.030	-0.204	0.145	-1.41	0.164
Put together ideas from different courses when preparing assignments	0.009	-0.086	0.078	-1.11	0.271	0.003	-0.064	0.141	-0.45	0.654
Used electronic medium to discuss or complete an assignment	0.014	-0.137	0.099	-1.39	0.167	0.105	-0.443	0.159	-2.78	**0.007**
Worked harder to meet instructor's expectations	0.002	-0.050	0.087	-0.58	0.563	0.003	-0.060	0.128	-0.47	0.638
Discussed ideas with others outside of class	0.010	-0.105	0.088	-1.19	0.236	0.000	-0.017	0.140	-0.12	0.902
Coursework emphasis on analysis	0.003	-0.043	0.071	-0.60	0.548	0.008	-0.101	0.145	-0.70	0.488
Coursework emphasis on synthesis	0.000	-0.018	0.084	-0.21	0.833	0.011	-0.133	0.157	-0.85	0.398
Coursework emphasis on applying theories to practical problems	0.005	-0.073	0.087	-0.85	0.398	0.028	-0.192	0.143	-1.34	0.185
Learned something that changed the way you understand an issue	0.000	0.009	0.081	0.11	0.916	0.006	-0.079	0.132	-0.60	0.553
Hours per week spent preparing for class	0.000	0.016	0.149	0.11	0.915	0.001	0.060	0.294	0.20	0.839
Extent to which institution provides support for academic success	0.008	0.082	0.082	1.00	0.319	0.006	0.071	0.114	0.62	0.536
Institutional contribution to writing clearly and effectively	0.000	-0.003	0.083	-0.04	0.967	0.002	0.036	0.121	0.30	0.766
Institutional contribution to thinking critically and analytically	0.019	-0.119	0.076	-1.56	0.121	0.000	0.010	0.128	0.08	0.938
Institutional contribution to working effectively with others	0.000	0.015	0.096	0.15	0.879	0.003	-0.065	0.154	-0.42	0.675
Institutional contribution to working effectively on your own	0.000	-0.009	0.088	-0.10	0.922	0.008	0.080	0.118	0.67	0.503
LAC benchmark	0.000	-0.018	1.204	-0.020	0.988	0.001	-0.413	2.000	-0.210	0.837
SCE benchmark	0.012	2.780	2.240	1.240	0.218	0.048	5.382	3.047	1.770	**0.082**
Academic Outcomes										
GPA at year end	0.013	-0.390	0.285	-1.37	0.173	0.013	-0.413	0.452	-0.91	0.364
Final grade in course	0.056	-0.685	0.234	-2.92	**0.004**	0.024	0.523	0.408	1.28	0.204

4.8.5 Summary

Both the primary experimental course (Course #1) and Course #2 in which the intervention had limited implementation success show essentially no engagement changes over the prior year. As with other NSSE-based assessments discussed above, at least part of the explanation likely results from the dilution of experimental effects across the students' entire year experiences and the intensity of the intervention itself. The absence of significant changes on key questions in Course #2 provides informal support for the existence of stability of (in this case matched) successive cohorts; the results also correspond with the relatively stable engagement background at WLU overall.

4.9 University of Waterloo (Teaching Excellence Academy)

4.9.1 Intervention Description and Background

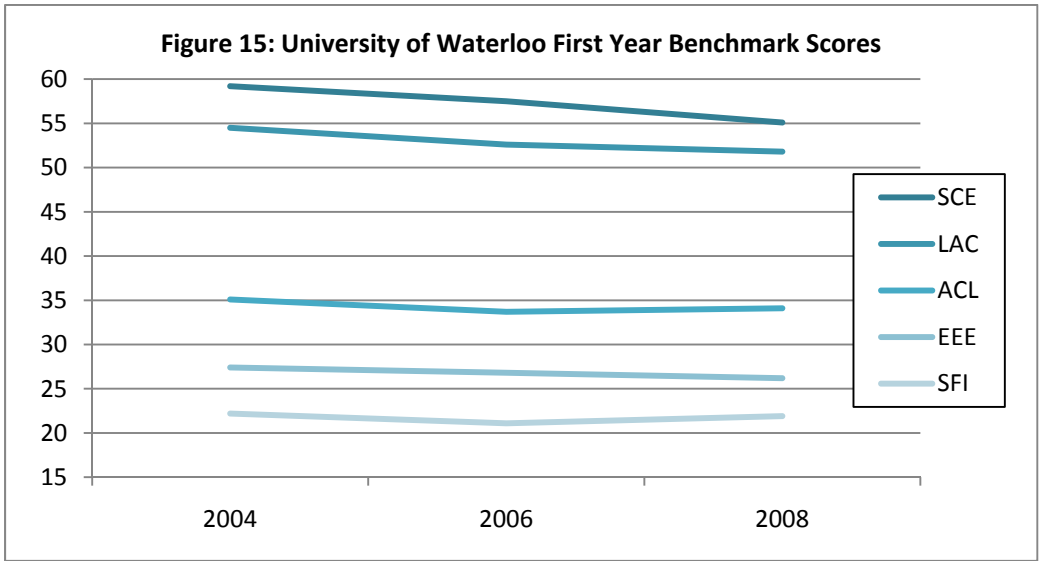
For several years, the University of Waterloo (UW) has offered a Teaching Excellence Academy (TEA), an intensive teaching improvement workshop offered to mid-career faculty members nominated by their Chair or Dean. The TEA is based on a model developed at McGill University and Simon Fraser University. It is delivered as a series of sessions over four days led by teaching and learning professionals at UW. Faculty participants redesign their courses consistent with best teaching and learning practices to align the course learning objectives, teaching techniques and student assessment methods. The outcome of the sessions is a substantially revised course as reflected in revised course outlines, reading lists, course content, delivery methods and tests/assignments/ exams. A more broad-based impact of the TEA is the establishment of a UW-based community of practice in teaching and learning and greater communication/support among faculty members.

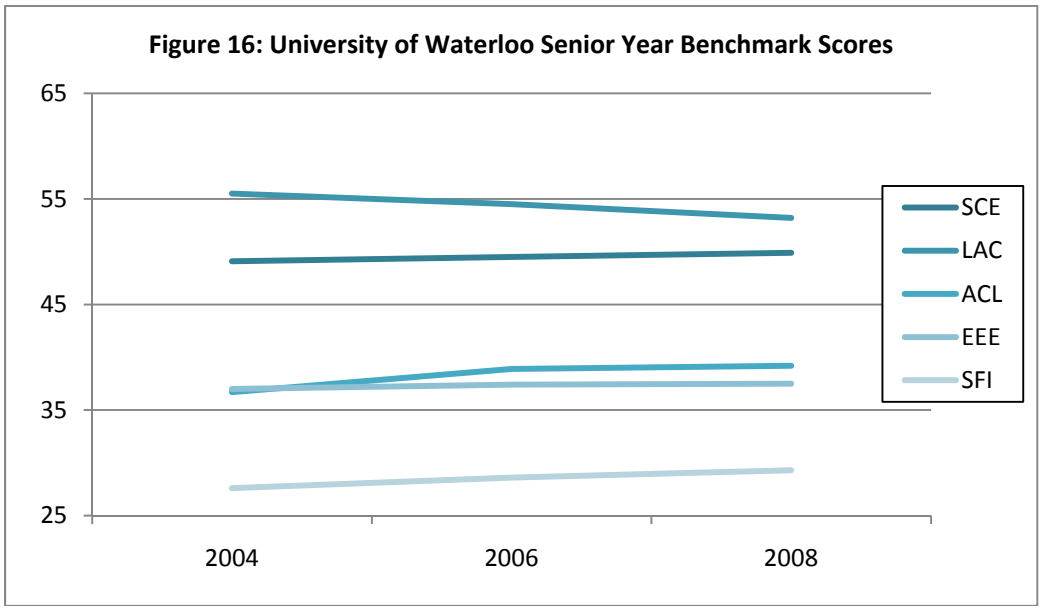
TEA activity preceded this interventions project, and a number of primarily qualitative assessment activities were already in place: pre- and post-course interviews with faculty to assess the expected impact of the TEA, summarize changes planned by the faculty member, and discuss the impacts of those changes; focus groups with students; and an in-class questionnaire administered without respondent self-identification. The University's participation in the interventions project was focused on determining whether the impacts of TEA-based course revisions – which were positive against various qualitative and satisfaction criteria – could be detected in NSSE results. UW selected three courses taught by faculty members participating in TEA – two in first-year and one in fourth-year – to contribute to this analysis.

UW administered NSSE in 2004, 2006 and 2008. It has incorporated NSSE results into its MYA reporting and performance indicator reports, and the results are used by the Undergraduate Student Relations Committee to assist in identifying areas for the improvement of teaching and learning throughout the University.

4.9.2 Context Provided by NSSE Administrations

Waterloo's first- and fourth-year overall engagement scores are reasonably consistent over time, with changes over the three administrations of less than five per cent in all cases. The pattern in first-year is one of stability in some scores and slight declines in others; in fourth-year most scores are stable or increasing (Figures 15 and 16).





UW identified a number of specific NSSE items for analysis that are consistent with the objectives of TEA; the 2006 and 2008 results for these items are presented in Table 42.

Table 42: University of Waterloo First Year and Senior Year Selected NSSE Item (Dependent Variable) Means

NSSE Item	First Year			Senior Year		
	2006	2008	% Change	2006	2008	% Change
Submitted 2+ drafts of paper before handing it in	2.15	2.06	-4.19%	2.11	2.12	0.47%
Worked on paper requiring integration of ideas from various sources	2.75	2.66	-3.27%	3.13	3.11	-0.64%
Received prompt feedback from faculty on academic performance	2.11	2.12	0.47%	2.28	2.27	-0.44%
Worked harder to meet instructor expectations	2.34	2.37	1.28%	2.38	2.40	0.84%
Coursework emphasis on memorization	2.73	2.73	0.00%	2.65	2.74	3.40%
Coursework emphasis on analysis	3.11	3.09	-0.64%	3.17	3.12	-1.58%
Coursework emphasis on synthesis	2.78	2.75	-1.08%	2.91	2.85	-2.06%
Coursework emphasis on making judgments	2.64	2.63	-0.38%	2.75	2.75	0.00%
Extent to which exams challenged you to do your best work	5.57	5.52	-0.90%	5.13	5.15	0.39%
Examined strengths and weaknesses of your own views	2.30	2.29	-0.43%	2.49	2.42	-2.81%
Extent to which institution supports academic success	3.02	2.93	-2.98%	2.59	2.62	1.16%
Institutional contribution to thinking analytically	3.27	3.19	-2.45%	3.35	3.35	0.00%
Institutional contribution to analyzing quantitative problems	3.12	3.12	0.00%	3.16	3.18	0.63%
Institutional contribution to solving complex real world problems	2.71	2.64	-2.58%	2.68	2.72	1.49%

4.9.3 Assessment Design

Both a successive cohort and cross-sectional design were originally envisaged, with each 2008/09 experimental course being compared with both its 2007/08 equivalent and with a "comparable" concurrent course. 100 per cent of the students in the three courses and the selected control group courses were sampled for NSSE in both 2008 and 2009. However, the cross-sectional design was abandoned when the instructor assigned to teach one of the courses changed, and when the delivery of a second course was moved to another semester. This left each of the three courses paired with a prior year course taught by the same instructor. (The loss of the cross-sectional design was not significant – if control/experimental differences had been observed it would have been difficult to interpret them without prior year data for the control anyway.) One of the courses was delivered in both classroom and distance mode, and those students registered in distance delivery were eliminated from that course's sample. Students in the experimental courses were propensity matched (within-course) to the prior year control using age, gender, co-op status and entering average. Registrant characteristics in the 2009 offerings of each course differed enough from the 2008 controls that the matching process resulted in a

higher-than-average loss of experimental sample (18 per cent - 48 per cent) although it did eliminate all significant differences post-match as shown in Table 43.

The three deep learning subscales and the aggregate deep learning scale developed by NSSE were calculated for the Waterloo data subsequent to the project getting underway, and were included in the analysis because of their relevance to TEA objectives. The learning scales are constructed in the same fashion as benchmarks, and assume values between 0 and 100.

Table 43: University of Waterloo Propensity Matching Results

		Pre-Match				Post-Match			
Course	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
Course 1 (First Year)	Age	-0.33	0.07	23.73	0.000	0.00	0.12	0.00	1.000
	Gender	-0.16	0.23	0.46	0.498	0.00	0.30	0.00	1.000
	Co-op Status	0.64	0.23	7.59	0.006	0.00	0.33	0.00	1.000
	Entering Average	0.02	0.02	0.66	0.415	-0.01	0.03	0.13	0.714
		(n=95 experimental, n=535 control)				(n=78 experimental, n=78 control)			
Course 2 (Fourth Year)	Age	3.24	1.05	9.59	0.002	0.00	1.46	0.00	1.000
	Gender	-0.42	0.51	0.68	0.408	0.00	0.76	0.00	1.000
	Co-op Status	all records had same co-op status				all records had same co-op status			
	Entering Average	0.02	0.09	0.03	0.870	-0.04	0.14	0.08	0.783
		(n=29 experimental, n=38 control)				(n=16 experimental, n=16 control)			
Course 3 (First Year)	Age	-0.09	0.15	0.35	0.554	0.00	0.33	0.00	1.000
	Gender	0.36	0.42	0.71	0.401	0.00	0.59	0.00	1.000
	Co-op Status	0.26	0.40	0.43	0.512	0.00	0.55	0.00	1.000
	Entering Average	-0.01	0.04	0.03	0.853	-0.03	0.07	0.17	0.683
		(n=54 experimental, n=48 control)				(n=28 experimental, n=28 control)			

4.9.4 Assessment Results

The results of the UW project (Table 44) are generally consistent with those of the other NSSE-only interventions. While some items show significant differences between control and experimental groups, there is little consistency in those items, and the coefficient signs are a mix of positive and negative.

Table 44: University of Waterloo Regression Results for Engagement Effects															
	Course 1 (First Year)					Course 2 (Fourth Year)					Course 3 (First Year)				
Dependent Variable	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**	**R^2**	**B-Est**	**SE**	**t-score**	**p-value**
Submitted 2+ drafts of paper before handing it in	0.002	0.087	0.166	0.52	0.601	0.075	-0.625	0.402	-1.56	0.130	0.014	0.246	0.280	0.88	0.384
Worked on paper requiring integration of ideas from various sources	0.007	0.143	0.136	1.05	0.295	0.080	-0.437	0.271	-1.61	0.118	0.013	0.228	0.272	0.84	0.406
Received prompt feedback from faculty on academic performance	0.002	0.077	0.135	0.57	0.569	0.131	-0.563	0.264	-2.13	**0.041**	0.068	0.464	0.232	2.00	**0.051**
Worked harder to meet instructor expectations	0.002	0.083	0.135	0.62	0.539	0.061	-0.438	0.314	-1.39	0.174	0.089	0.608	0.267	2.28	**0.027**
Coursework emphasis on memorization	0.031	0.295	0.134	2.20	**0.029**	0.107	0.563	0.297	1.89	**0.068**	0.014	0.214	0.241	0.89	0.378
Coursework emphasis on analysis	0.002	0.064	0.121	0.53	0.597	0.034	-0.313	0.306	-1.02	0.315	0.003	0.071	0.187	0.38	0.704
Coursework emphasis on synthesis	0.003	-0.087	0.132	-0.65	0.514	0.020	-0.250	0.323	-0.77	0.445	0.020	0.250	0.241	1.04	0.305
Coursework emphasis on making judgments	0.022	0.270	0.145	1.86	**0.065**	0.006	-0.125	0.301	-0.42	0.681	0.115	0.607	0.229	2.65	**0.011**
Higher order thinking subscale	0.002	1.348	2.816	0.48	0.633	0.008	2.902	5.942	0.49	0.629	0.007	2.694	4.516	0.60	0.553
Integrative learning subscale	0.002	1.431	2.501	0.57	0.568	0.104	-9.039	4.925	-1.84	**0.077**	0.083	10.607	5.120	2.07	**0.044**
Reflective learning subscale	0.013	4.516	3.252	1.39	0.167	0.055	9.130	8.074	1.13	0.270	0.006	3.189	5.702	0.56	0.579
Deep learning scale	0.005	1.911	2.222	0.86	0.391	0.004	-1.337	3.852	-0.35	0.731	0.063	6.834	3.665	1.86	**0.068**
Extent to which exams challenged you to do your best	0.003	-0.109	0.169	-0.64	0.520	0.007	0.250	0.528	0.47	0.640	0.010	0.286	0.378	0.76	0.453

work															
Examined strengths and weaknesses of your own views	0.004	0.115	0.142	0.81	0.419	0.005	0.125	0.338	0.37	0.714	0.046	0.393	0.240	1.63	0.109
Extent to which institution supports academic success	0.007	-0.128	0.125	-1.02	0.308	0.111	0.533	0.285	1.87	**0.072**	0.017	0.214	0.224	0.95	0.344
Institutional contribution to thinking analytically	0.022	0.223	0.121	1.84	**0.068**	0.022	-0.223	0.279	-0.80	0.431	0.089	0.464	0.202	2.29	**0.026**
Institutional contribution to analyzing quantitative problems	0.001	0.062	0.151	0.41	0.684	0.010	-0.170	0.318	-0.53	0.598	0.046	0.393	0.243	1.62	0.112
Institutional contribution to solving complex real world problems	0.017	-0.239	0.148	-1.61	0.109	0.003	0.119	0.415	0.29	0.777	0.020	0.298	0.289	1.03	0.308
LAC benchmark	0.000	0.502	2.222	0.23	0.821	0.021	-3.570	4.479	-0.80	0.432	0.017	3.624	3.729	0.97	0.336
ACL benchmark	0.000	-0.597	2.323	-0.26	0.797	0.037	-5.656	5.232	-1.08	0.288	0.024	4.371	3.845	1.14	0.261
SFI benchmark	0.001	1.106	2.521	0.44	0.662	0.018	-3.441	4.577	-0.75	0.458	0.024	5.430	4.746	1.14	0.258

4.9.5 Summary

While it is true that small samples in two of the three courses might have limited the opportunity for significant results to emerge, the larger course, as with large sample sizes in other projects, also showed few effects. It appears that any intervention effects were diluted within the overall NSSE item and benchmark scores.

Inclusion of the deep learning scales was “opportunistic”: little analysis has been undertaken on these scales within Ontario universities. As the result of another project, the scales have recently been computed for a number of NSSE response files in Ontario and an exploration of scale variability and explanatory factors will be undertaken in the future.

4.10 University of Windsor (Intrusive Advising Program in First-Year School of Business)

4.10.1 Intervention Description and Background

Results from previous administrations of the Canadian University Survey Consortium (CUSC) survey indicated that only 50 per cent of new students at Windsor received academic advising prior to registering in their courses; that number dropped to 25 per cent during second year and was accompanied by a reduced level of satisfaction with advising services. Prior NSSE results showed that a lower-than-desired number of students discussed career plans with faculty members. In response, and consistent with Windsor's strategic plan, the University established the Advising Centre to develop and offer a wider range of advising services and to involve faculty more directly in student advising. In cooperation with the Advising Centre, Windsor's School of Business (whose CUSC and NSSE results mirrored those of the University overall) developed an "intrusive" advising plan for implementation in the Fall 2008 semester. The program involved the assignment of each new student to a faculty advisor (who did not instruct the student during the Fall term) and two senior student mentors (who had completed a training session). Students were informed of the program and were invited to participate both prior to course registration and again at the beginning of the Fall term. Weekly one-on-one meetings with student mentors focused on issues encountered during the early weeks of term (including academic problems and needs) and social activities and academic supports available to new students. Meetings with faculty advisors during the first week of classes and monthly thereafter involved discussions of academic issues and the development on an individualized student academic and advising plan. The overall objective of the project was to increase student engagement, social and academic integration and motivation.

4.10.2 Context Provided by NSSE Administrations

The University of Windsor administered NSSE in 2005, 2006 and 2008. First-year benchmark scores have generally increased over the three-year period, and particularly from the 2006 to 2008 administrations. The individual NSSE items selected by Windsor for analysis also show generally positive increases but with much greater variation, with some item means increasing by as much as nine per cent between 2006 and 2008 (Figure 17 and Table 45). Certain items are clearly central to the advising program – discussion of grades and career plans with faculty member, quality of relationships with faculty, quality of academic advising – but as with the other projects, interpretation of any results must be based on the preponderance of evidence rather than the significance of a few items.

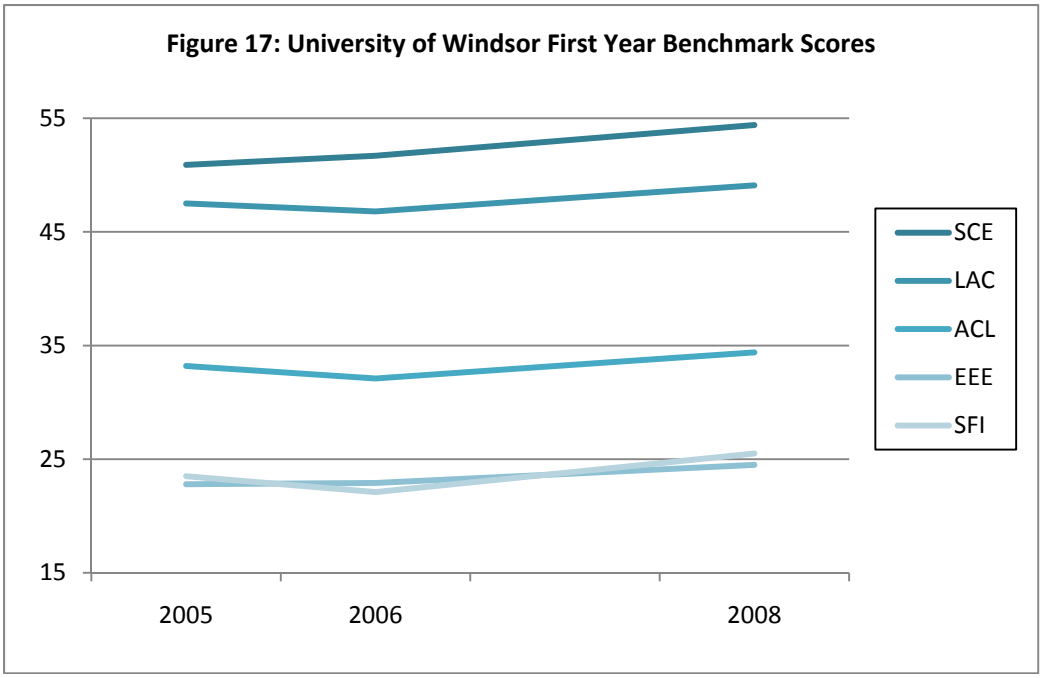


Table 45 : University of Windsor First Year Selected NSSE Item (Dependent Variable) Means			
NSSE Item	**2006**	**2008**	**% Change**
Came to class without completing readings or assignments	2.11	2.17	2.84%
Discussed ideas with others outside of class	2.28	2.41	5.70%
Frequency of tutoring or teaching other students	1.66	1.79	7.83%
Used electronic medium to discuss or complete assignments	2.54	2.71	6.69%
Used email to communicate with instructor	2.77	2.85	2.89%
Discussed grades with faculty member	2.14	2.19	2.34%
Talked about career plans with faculty member or advisor	1.60	1.74	8.75%
Discussed ideas from readings with faculty member outside class	1.68	1.77	5.36%
Received prompt feedback from instructor on academic performance	2.10	2.26	7.62%
Quality of relationships with faculty members	4.57	4.77	4.38%
Institution provides support for academic success	2.71	2.77	2.21%
Quality of academic advising	2.66	2.66	0.00%
Evaluation of entire educational experience at institution	2.82	2.83	0.35%
Would attend same institution if starting over again	2.94	2.93	-0.34%

It has been argued above that (specifically) course-based intervention results may be diluted through the use of NSSE because such interventions constitute only a small

portion of the experience students report on the survey. Windsor's project provides a somewhat different perspective. Student advising at Windsor occurs at the Faculty-wide level and the focus of the intervention is also Faculty-wide. The overall student advising experience at Windsor (with or without the advising intervention) is in fact what NSSE attempts to measure in the "advising" item. In the Windsor project, intervention intensity will likely be the primary factor in NSSE's ability to detect engagement changes, rather than the scope of the intervention itself.

4.10.3 Assessment Design

Analysis of Windsor's advising program involves both cross-sectional post-measure and successive cohort post-measure designs. NSSE was administered in 2008 with a 100 per cent sample of students in the Business School to provide a prior cohort control population. A targeted administration of NSSE in 2009 involved only Business students. Those students who self-selected for the project in 2008/09 constitute the experimental group; those who did not participate represent a concurrent control population. Nineteen of the advising project participants (about one-third of the total) completed NSSE. Experimental group records were propensity matched with both the current year and prior year control group records using basis of admission, age, domestic/international status, gender, full-/part-time status and entering average. Significant and near-significant pre-match differences (basis of admission and age in both designs, entering average in the successive cohort design) were eliminated post-match. The size of the control populations allowed both designs to use a 2:1 match ratio (Table 46).

A subjective intensity of involvement score was assigned to each of the participating students based on the number of meetings with faculty members and student mentors. Insufficient variation in the scores and the small number of records at each level prevented its use in this analysis.

Table 46: University of Windsor Propensity Matching Results									
		Pre-Match				Post-Match			
Design	Variable	Estimate	SE	Wald Chi-Square	p-value	Estimate	SE	Wald Chi-Square	p-value
Cross-Sectional	Basis of admission	2.02	0.77	6.84	0.009	-0.26	0.96	0.07	0.788
	Age	0.57	0.31	3.31	0.069	-0.02	0.43	0.00	0.956
	Domestic/international	1.17	0.78	2.24	0.134	-0.69	1.04	0.44	0.506
	Gender	0.14	0.50	0.08	0.773	-0.01	0.57	0.00	0.952
	PT/FT attendance	-12.37	357.36	0.00	0.972	all records in same group			
	Entering average	-0.05	0.04	1.78	0.182	-0.03	0.05	0.50	0.479
		(n=19 experimental, n=102 control)				(n=19 experimental, n=38 control)			
Successive Cohort	Basis of admission	1.97	0.77	6.61	0.010	0.03	1.26	0.00	0.981
	Age	0.65	0.29	5.06	0.025	0.49	0.58	0.72	0.395
	Domestic/international	1.17	0.77	2.30	0.130	-0.69	1.45	0.23	0.631
	Gender	0.53	0.49	1.19	0.276	-0.06	0.58	0.01	0.922
	PT/FT attendance	-12.31	314.83	0.00	0.969	all records in same group			
	Entering average	-0.10	0.04	5.22	0.022	-0.06	0.05	1.12	0.290
		(n=19 experimental, n=142 control)				(n=19 experimental, n=38 control)			

4.10.4 Assessment Results

Results for the two designs are provided in Table 47. Neither of the designs shows a significant experimental effect. This may be the result of small sample size, intervention intensity and/or the absence of an intensity of involvement measure.

Table 47: University of Windsor Regression Results for Engagement Effects										
	Cross-Sectional Design					Successive Cohort Design				
Dependent Variable	R^2	B-Est	SE	t-score	p-value	R^2	B-Est	SE	t-score	p-value
Came to class without completing readings or assignments	0.029	0.269	0.214	1.26	0.215	0.080	0.414	0.196	2.12	**0.039**
Discussed ideas with others outside of class	0.020	0.202	0.196	1.03	0.309	0.026	0.244	0.207	1.18	0.244
Frequency of tutoring or teaching other students	0.009	-0.173	0.251	-0.69	0.494	0.000	0.016	0.243	0.07	0.947
Used electronic medium to discuss or complete assignments	0.011	0.177	0.239	0.74	0.460	0.034	0.337	0.252	1.33	0.188
Used email to communicate with instructor	0.019	0.251	0.251	1.00	0.322	0.019	0.235	0.241	0.97	0.334
Discussed grades with faculty member	0.029	0.296	0.237	1.25	0.217	0.000	-0.042	0.296	-0.14	0.887
Talked about career plans with faculty member or advisor	0.006	0.117	0.213	0.55	0.585	0.004	0.108	0.235	0.46	0.648
Discussed ideas from readings with faculty member outside class	0.001	0.045	0.191	0.24	0.814	0.025	-0.265	0.235	-1.12	0.266
Received prompt feedback from instructor on academic performance	0.000	0.028	0.211	0.13	0.896	0.003	-0.094	0.222	-0.42	0.674
Quality of relationships with faculty members	0.001	-0.084	0.401	-0.21	0.834	0.007	-0.222	0.378	-0.59	0.559
Institution provides support for academic success	0.013	-0.197	0.242	-0.82	0.419	0.000	0.020	0.232	0.08	0.933
Quality of academic advising	0.000	0.008	0.247	0.03	0.976	0.003	0.085	0.222	0.38	0.703
Evaluation of entire educational experience at institution	0.008	-0.135	0.208	-0.65	0.519	0.018	-0.209	0.220	-0.95	0.347
Would attend same institution if starting over again	0.000	0.023	0.219	0.10	0.919	0.020	-0.216	0.216	-1.00	0.324
SCE benchmark	0.005	-2.527	5.120	-0.49	0.624	0.017	5.680	6.160	0.92	0.361
SFI benchmark	0.023	4.866	4.391	1.11	0.273	0.006	2.900	5.190	0.56	0.579

4.10.5 Summary

The University of Windsor project was subject to a relatively high degree of self-selection bias and involved a relatively small number of participants who were also NSSE respondents. While group differences were eliminated post-match, the small sample and the lack of a robust involvement score limited the power of the analysis. Although the NSSE question on advising corresponds to the level at which Windsor offers advising service, it is possible that a more detailed set of NSSE questions on advising, or a custom survey, might result in more promising analysis.

4.11 Queen's University (Enhanced On-Line Tutorial Support Across Fourth Year Electrical Engineering)

4.11.1 Intervention Description and Background

The fourth-year Electrical Engineering curriculum at Queen's University includes several courses in signal processing, communications and computer networks that have a common mathematics foundation. Three such courses are offered in the Fall term and three in the Winter term; the majority of students in the program register in at least several (n=60) if not all (n=40) of these courses. Tutorial support has traditionally been offered on a course-by-course basis with no integration among courses. In order to capitalize on the common course foundation and to improve access to tutorial support, an on-line tutorial support pilot program was introduced in 2008/09. The program consisted of instant messaging (IM) access to teaching assistants over extended hours (appropriate to student class schedules) and web enhancements to post FAQ's and other course content, supplemented with one-on-one tutorial support for issues that could not be resolved electronically. Conventional tutorial services were maintained throughout the pilot. Availability and details of the enhanced service were communicated to all eligible students via email and during classes in the first week of the Fall term. Teaching assistants were provided orientation to and training for the tutorial delivery model.

4.11.2 Context Provided by NSSE Administrations

Queen's University administered NSSE in 2004, 2006 and 2008, with a 100 per cent sample in the latter two years. Benchmark scores for the University overall are shown in Figure 18, and show overall stability with the exception of a decline in the SCE and SFI benchmarks in the 2006 administration. NSSE drilldown results are also available for the Electrical Engineering program and for the Faculty of Applied Science overall; selected results relevant to the advising intervention are shown in Table 48 and generally indicate engagement levels in Electrical Engineering slightly lower than those for the Faculty overall. As the result of a data sharing arrangement within the G13 (an association of research-intensive universities in Canada), program-level results are also available for the 13 member institutions overall; these results indicate that the Queen's program operates at close to the G13 average on the selected items.

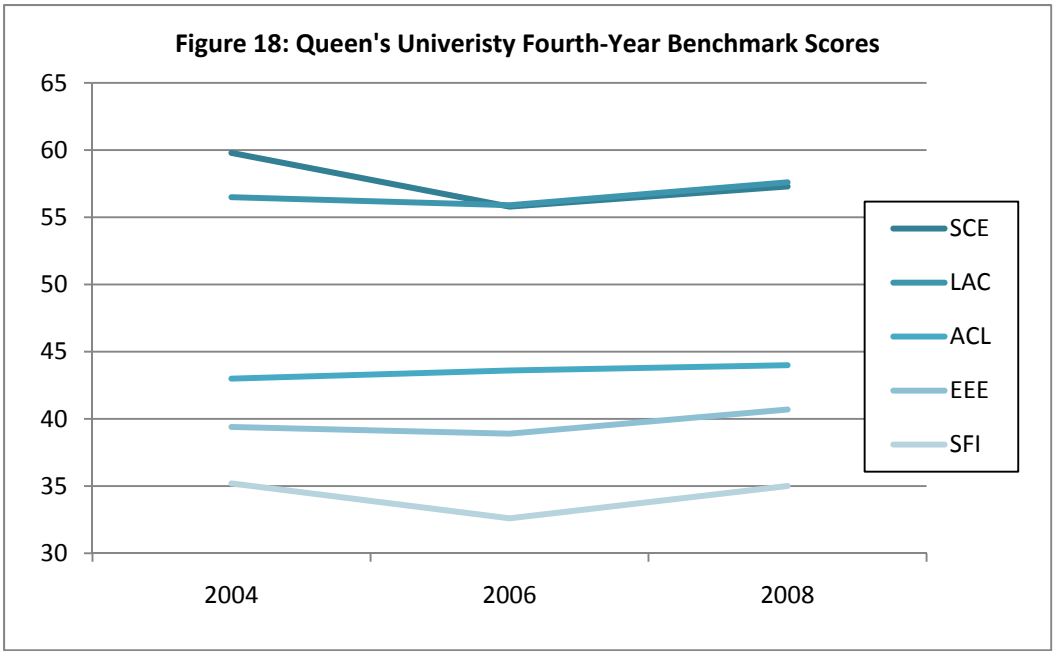


Table 48: Queen's University Selected Fourth-Year 2008 NSSE Item Means Comparisons for Engineering

NSSE Item	Queen's Electrical Engineering	Queen's Faculty of Applied Science	G13 Electrical Engineering
Institutional environment provides support needed for academic success	2.50	2.80	2.51
Used electronic medium to complete or discuss an assignment	3.13	3.15	3.04
Put together ideas from different courses when completing assignments	2.58	2.98	2.74
Worked with classmates outside of class to prepare assignments	3.04	3.46	3.17
Used email to communicate with an instructor	2.88	3.16	2.98
Received prompt feedback from faculty on academic performance	2.00	2.30	2.08
Problem sets taking more than one hour to complete	3.54	3.28	3.45
Problem sets taking less than one hour to complete	2.21	1.90	2.07
Institutional contribution to helping you to learn on your own	3.05	3.22	3.15
Deep learning composite scale	58.33	65.93	62.19

4.11.3 Assessment Design

The Electrical Engineering project was established to use NSSE in a successive cohort design and a slightly modified form of CLASSE in a cross-sectional design. NSSE was administered to 100 per cent of the students in the fourth-year of the program in both 2008 (control) and 2009 (experimental) and achieved a 45 per cent response rate in both years. CLASSE was administered and achieved a 35 per cent response rate in two of the three Winter term classes in which the vast majority of eligible students were registered.

4.11.4 Assessment Results

Initial take-up to the tutorial service was relatively low. A further round of invitations was distributed to students, and it was expected that utilization would increase during the Fall term examination period; however, usage remained low with only eight of the more than 60 eligible students accessing the service by the end of the Fall term. The major cost of the program was in TA stipends that had been fixed early in the year, so termination would not have resulted in cost savings. The service continued to operate throughout the Winter term; the same students continued to participate but no new students participated during the final four months. By itself, the program's low take-up rate prevented the kind of analysis undertaken for the other interventions. The problem was further aggravated by the lower-than-expected CLASSE response rate and the failure of most on-line tutoring service users to respond to NSSE, or to respond to (or self-identify on) CLASSE.

Further analysis within the Department may provide an explanation for the low take-up rate that appears to contradict the results of the various surveys that indicated a student desire for, and the academic value of the service. For the purposes of this report, fourth-year students did respond to the NSSE and/or CLASSE surveys, and a few limited forms of analysis are still possible:

- An exploration of the stability of NSSE scores for small unmatched samples over the two administrations;
- A "mapping" of content-matched questions on NSSE and CLASSE to determine the consistency of response;
- A more detailed comparison of NSSE responses in fourth-year Electrical Engineering at Queen's and across the G13 institutions to determine whether items of interest differ and to encourage conversation about discipline-specific strategies.

NSSE Score Stability over Successive Administrations:

Table 49 and Figure 19 present a comparison of unmatched group NSSE scores for the 2008 and 2009 administrations. The total number of completed responses was 24 and 29 in the two years respectively. The majority of the scores are highly consistent year over year (R^2 = .96) and the regression line aligns very closely with the 45-degree line with which it should coincide, despite differences in some item scores of 10 per cent or more.

These results are more consistent than those observed in the successive cohort characteristics of several other projects.

NSSE-CLASSE Content-Matched Question Consistency:

The results of some other projects (e.g., Western's, Queen's Psychology Department) indicated that course-specific experimental effects did not carry over to overall NSSE results. The Queen's Electrical Engineering project administered CLASSE in order to collect information about a cluster of common courses, however, which represented between about 20 per cent and 50 per cent of a student's full course load (rather than the five per cent - 10 per cent a single course would typically constitute). For the 15 content-matched questions on the NSSE and CLASSE instruments, the results indicate considerable consistency between the results of the two instruments (R^2 = .83) and slightly lower mean scores for CLASSE than for NSSE on each matched question (Table 50 and Figure 20). This consistency provides strong initial evidence for the validity of CLASSE in measuring the same engagement behaviours as NSSE and for the lessening of the apparent dilution effect of NSSE when applied to two or more courses.

Table 49: Queen's University NSSE Item Stability for Successive Unmatched Small Samples

NSSE Item	2008 Mean	2009 Mean
Asked questions/contributed to class discussions	2.36	2.54
Made class presentation	1.98	2.14
Prepared 2+ drafts of paper before handing it in	2.09	1.89
Worked on paper requiring integrating ideas from various sources	3.02	2.93
Came to class without completing readings or assignments	2.09	2.36
Worked with other students on projects during class	1.98	1.61
Worked with classmates outside class to prepare assignments	3.16	3.29
Put together ideas from different courses when completing assignments	2.96	2.71
Tutored or taught other students	2.30	2.04
Used an electronic medium to discuss/complete assignment	3.25	2.89
Used email to communicate with instructor	3.21	3.25
Discussed grades or assignments with instructor	2.14	2.39
Talked about career plans with faculty member or advisor	1.98	1.71
Discussed ideas with faculty members outside of class	1.96	1.71
Received prompt feedback on academic performance	2.21	2.21
Worked harder to meet instructor expectations	2.52	2.21
Worked with faculty members on activities other than coursework	1.66	1.57
Discussed ideas with others outside of class	2.77	2.50
Coursework emphasis on memorization	2.45	2.64
Coursework emphasis on analysis	3.36	3.18
Coursework emphasis on synthesis	3.16	2.86
Coursework emphasis on making judgments	2.71	2.61
Coursework emphasis on applying theories to practical problems	3.27	3.18
Number of assigned texts or packs of course readings	3.07	2.82
Number of written papers of 20+ pages	1.89	1.75
Number of written papers of 5 - 19 pages	2.61	2.68
Number of written papers of less than 5 pages	2.73	2.89
Number of problem sets taking 1+ hours to complete	2.43	3.21
Number of problem sets taking less than 1 hour to complete	2.11	2.00
Examinations challenged you to do your best work	5.11	5.50
Practicum, internship or field experience	0.42	0.54
Participation in learning community	0.33	0.25
Worked on research project with faculty member outside of program	0.21	0.14
Independent study or self-designed major	0.26	0.11
Culminating senior experience	0.54	0.43
Quality of relationships with other students	5.93	5.71
Quality of relationships with faculty members	4.81	4.89
Quality of relationships with admin personnel	5.23	5.36
Hours per week spent preparing for class	5.31	4.71
Institutional emphasis on providing support for academic success	2.71	2.85
Institutional contribution to helping you write clearly and effectively	2.77	2.75
Institutional contribution to helping you think critically and analytically	3.46	3.18
Institutional contribution to helping you learn effectively on your own	3.40	2.93
Institutional contribution to helping you solve complex problems	3.05	2.68

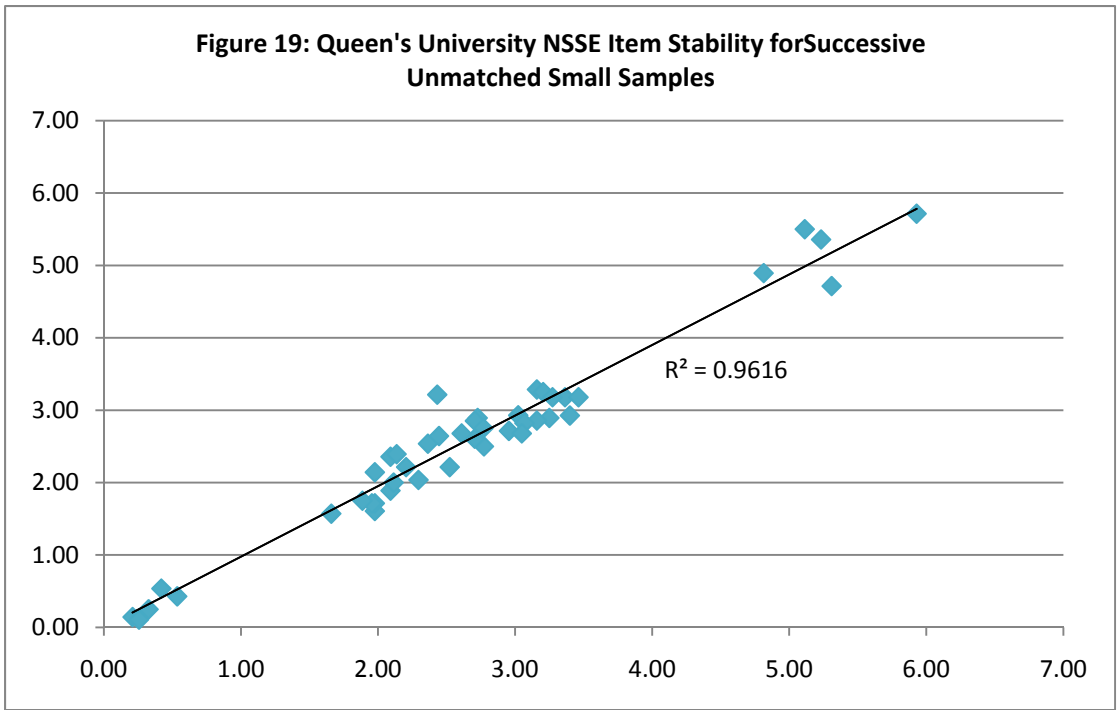
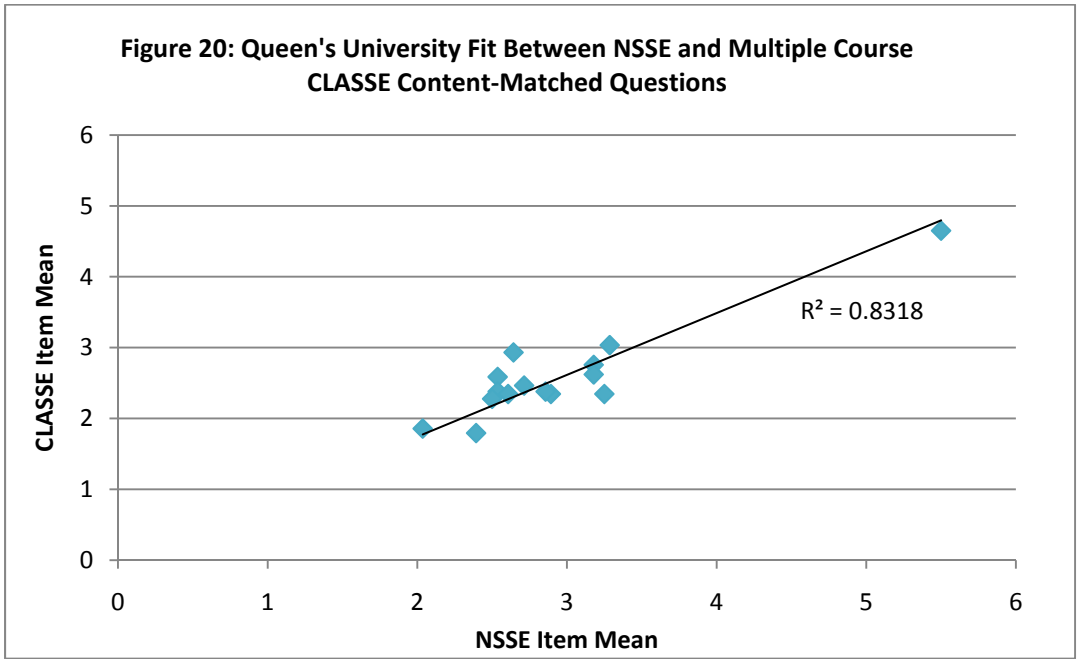


Table 50: Queen's University Comparison of NSSE and CLASSE Content-Matched Questions

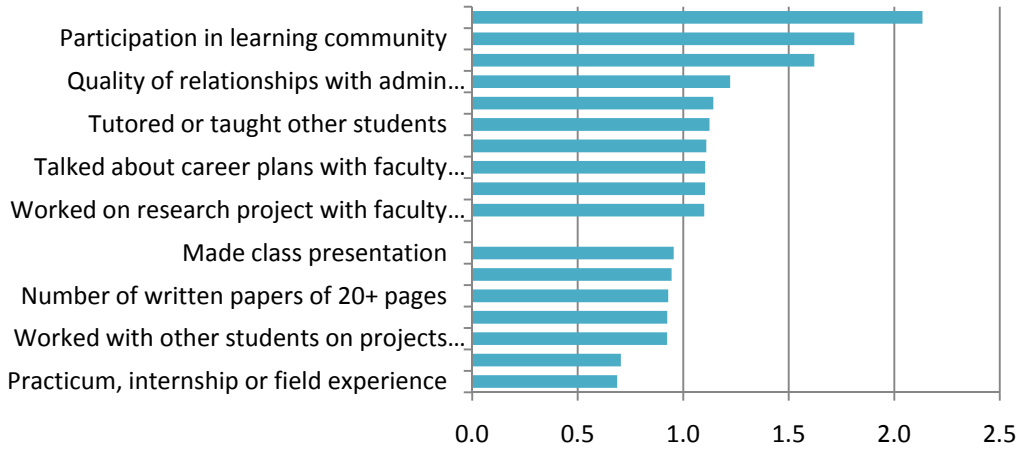
NSSE Item	NSSE Mean	CLASSE Mean
Asked questions/contributed to discussions (1 question	2.54	2.59
on NSSE; separate questions on CLASSE)	2.54	2.38
Worked with classmates outside class to prepare assignments	3.29	3.03
Put together ideas from different courses when completing assignments	2.71	2.46
Tutored or taught other students	2.04	1.86
Used an electronic medium to discuss/complete assignment	2.89	2.35
Used email to communicate with instructor	3.25	2.35
Discussed grades or assignments with instructor	2.39	1.79
Discussed ideas with others outside of class	2.50	2.28
Coursework emphasis on memorization	2.64	2.93
Coursework emphasis on analysis	3.18	2.76
Coursework emphasis on synthesis	2.86	2.38
Coursework emphasis on making judgments	2.61	2.35
Coursework emphasis on applying theories to practical problems	3.18	2.62
Examinations challenged you to do your best work *	5.50	4.65
(CLASSE item mean converted from 4-point to 7-point scale)		



Queen's and G13 NSSE Comparisons for Electrical Engineering:

Figure 21 provides a comparison of NSSE item scores for Queen's and (in the aggregate) the research-intensive universities in the G13, for selected items showing the greatest differences. NSSE does not contain a question bearing directly on the Queen's project (i.e., ease of access to out-of-class instructional assistance), but it does indicate that the Department's scores differ from the G13 average in a number of key areas, indicating possible areas of strength and weakness. Further analysis within the Department might determine whether these strengths and weaknesses might also form the basis for a future engagement intervention.

Figure 21: Queen's University Selected NSSE Item Scores Relative To G13 (Electrical Engineering)



4.11.5 Summary

Despite the low take-up rate that resulted in few intervention participants and an inability to undertake formal assessment, the Queen's project nonetheless provided an alternate perspective on the stability of successive cohorts that may warrant further analysis concerning the effect of year of study and program concentration on cohort differences; and it indicated that when CLASSE is administered over a cluster of courses rather than a single course, it begins to map much more closely to NSSE results.

4.12 Qualitative and Process Aspects of Intervention Assessment and the Project Overall

As noted earlier, most of the interventions involved qualitative and non-NSSE based assessment and analysis not reported on here, including additional surveys, student focus groups, and interviews with faculty and service providers. These activities generated significant amounts of data and information that were not directly linked to record-level NSSE responses and that were therefore outside the bounds of this report. Readers are encouraged to make contact with the individual sites to gain a better understanding of the contribution of these activities to overall intervention assessment.

The purpose of this section is to examine selected aspects of the intervention implementation and analysis process from both the site and central perspectives. This will complete the discussion presented in Section 2 above that documented activity up to the mid-point of intervention planning, and will address one of the four primary project

objectives – the sharing of experiences to support more effective implementation practice over the longer term. The primary source of information for this section is a survey administered to site participants in April 2009, at (for most sites) the completion of the implementation phase.

4.12.1 Intervention Cost

The ten sites received intervention project funding totaling approximately $145,000. Funding was limited to direct project expenses, including the hiring of temporary staff (e.g., TA's), equipment, office expenses and the additional cost associated with data collection and submission. This amount was about 70 per cent of total budget requests submitted; each institution accepted a reduction in funding (and a corresponding increase in in-kind contributions) in order to remain within the total funding available. Each site provided estimates of the total cost of their participation, both including and excluding the time commitment of pre-existing staff and faculty. Expenditures over and above project funding across all sites amounted to over $300,000 (including time commitments from existing staff) or $130,000 excluding these time commitments. Even a conservative estimate of total project costs, then, amounts to $275,000. This was, for all institutions involved, the first time a formal statistical assessment of the engagement impacts of interventions had been attempted. There is every reason to expect that subsequent interventions can be implemented at lower cost, as institutions climb the learning curve, data preparation and analysis become standardized and efficiencies in planning and delivery are found.

4.12.2 Organizational Learning

Institutions reported on the extent to which awareness of NSSE and engagement generally had increased in different areas of the university in the past two years, and whether factors other than the intervention had had an effect on awareness. Although a few institutions reported substantial awareness increases (particularly within senior academic administration and service provider groups), most indicated that awareness had increased only slightly in the two groups above and in the institutional research office, the teaching and learning centre, and among individual faculty members. Mechanisms for information dissemination and sources of heightened awareness during the preceding two years included:

- formal and informal discussions among department faculty, at retreats and in various committees;
- publication by *Maclean's Magazine* of benchmark scores for the majority of Canadian universities;
- an increased visibility of NSSE on institutional websites (including the incorporation of selected NSSE results in the Common University Data – Ontario (CUDO) presentation format;
- direct or indirect incorporation of engagement issues in strategic and academic plans and performance indicator documents;
- several larger-scale symposia and conferences on teaching and learning attended by faculty and staff;

- internal NSSE presentations to academic departments and Senate.

While this report will hopefully assist in generating increased awareness, it appears that engagement has not yet become a “top of mind” issue within the participating universities.

Another measure of the extent to which the nine participating universities have integrated an engagement focus into their operations is the kinds of activities that were performed prior to two years ago, those that have emerged during the past two years, and those that are still not a part of institutional operations. Table 48 presents a summary of the status of various activities. Engagement considerations remain largely separate from budget and internal academic assessment, and further experimentation/assessment has been undertaken at only a few universities. Most universities performed data drilldowns, made internal presentations and included NSSE in accountability reporting prior to 2007. A few institutions report that since 2007 they have begun making (or increased the number of) internal presentations, struck committees, incorporated engagement into budget decisions, undertaken other experiments or assessments, or commenced faculty and staff training activity. (In 2009 44 Canadian universities embarked on a NSSE data-sharing initiative involving program- and student subgroup-level drilldowns.) While there is no prescribed best approach to improved engagement practice, it does not appear that a common implementation or dissemination style has emerged within the participating universities in recent years.

Project participants at each of the universities also provided an informal assessment of the benefits of the intervention (regardless of whether the statistical analysis ultimately indicated an impact) in terms of increasing awareness among various groups, improving the student experience, identifying further opportunities for experimentation or research, enhancing external accountability or facilitating organizational or administrative change. Most expressed the belief that the intervention had in fact improved the student experience, with slightly fewer indicating they had identified opportunities for further implementation practice and research or organizational reform to advance implementation.

Table Fifty-One: Progress on Implementation Activity at Participating Institutions			
Activity	**Did not do two years ago and do not do now**	**Did not do two years ago but do now**	**Did do two years ago and still do now**
Drilldowns of NSSE responses by Faculty or program	0	0	**9**
NSSE presentations to individual faculty or groups of faculty	1	0	**8**
NSSE Presentations to service providers	1	1	**7**
Committees to study, promote or implement NSSE	4	2	3
Designation of senior academic to "champion" NSSE	3	2	4
Explicit budget allocations to NSSE-related activities	6	1	2
Incorporation of NSSE into academic program reviews	**7**	0	2
Incorporation of NSSE into instructor/course evaluations	**9**	0	0
Inclusion of NSSE results in accountability reporting (other than CUDO)	1	1	**7**
Formal exchange of NSSE results with other institutions	5	0	4
Other engagement-driven classroom experiments	6	1	2
Other engagement-driven non-classroom experiments	4	3	2
Other assessments of engagement impacts	6	1	2
Cognitive or focus group analysis on NSSE results	**7**	0	2
Engagement-based training or support to faculty	3	1	5
Engagement-based training or support to service providers	5	1	3

4.12.3 Intervention Implementation Issues

Nine of the original 13 approved projects were completed and assessed according to plan. However, site participants encountered a variety of problems ranging in severity, a discussion of which may assist in future planning and implementation:

- Unanticipated changes in instructor assignment and course scheduling affecting the delivery of a course-based project and the availability of a control group;
- Loss of design/implementation team members for various reasons, and a decline in the continuing commitment or availability of project staff;
- The projects as designed required significant effort not just in intervention and assessment planning and design, but also in terms of a targeted administration of NSSE, the design and administration of additional surveys, the merging of NSSE

response data with information from the student records system, and periodic progress reporting, putting occasional pressure on institutional research and project staff;
- One of the three terminated projects, and several of those that proceeded to completion, experienced lower-than-expected take-up rates;
- Despite relatively consistent NSSE response rates in prior administrations, a few of the projects came close to achieving an insufficient number of responses in the targeted 2009 administration;
- More broadly-based (e.g., Faculty-wide) interventions required faculty/staff participation and support well beyond direct project leaders that was sometimes difficult to generate and maintain;
- Inconsistent requirements by research ethics boards across the participating universities that limited disclosure and reporting of certain types of data and resulted in differing levels of flexibility in the administration of CLASSE.

The model employed for these projects involved submission of all data to, and performance of all analysis and reporting by, a "central project office" at Queen's University. The author's views on the experiences of the central office are:

- A higher level of detail is warranted in project proposals and in the proposal assessment process particularly with respect to the size and take-up rate of the target population(s) and a requirement for additional measurement tools;
- Data submission requirements should have been more standardized;
- The time required to assemble and organize the data, perform propensity matching and conduct the analysis was greater than that originally envisaged and ultimately necessitated an extension to the project completion date;
- One of the objectives of this project is the documentation and subsequent sharing of implementation experiences. This suggests a report with considerable detail on such issues as propensity matching, regression model results and plausible (in addition to probable) conclusions. While this report may contain more detail than some readers would want, it will hopefully convey the necessary information to those in need of a higher level of detail.

5. Conclusions

5.1 Overview

This concluding section attempts to pull together the key findings of this project, most of which have been introduced or alluded to earlier. The first eight topics deal with effective practice in intervention design, implementation and assessment and are offered as suggestions to those involved in engagement improvement efforts. They constitute relatively clear conclusions from this project, but will certainly be refined by others as engagement implementation practice continues to evolve. The final three topics are specifically related to the accountability environment in Ontario and present recommendations concerning the cost and funding of engagement implementation, an approach to engagement-based accountability, and the role of NSSE within the multi-year accountability agreements.

5.2 Intervention Scale and Intensity

One of the four objectives of this project was to assess whether NSSE and other engagement and academic outcome measurement tools were able to detect the existence of engagement-based interventions. The scale of interventions (i.e., the number of students affected) and their intensity (the degree to which behaviours are "pushed") are critical factors in this determination. The interventions in this project were modest in scale, affecting one course in a student's overall program of study, or one bundle of services in the overall academic and service experiences of students. They were also relatively low intensity: they involved adaptation of or enhancements to courses and services rather than their fundamental redesign. The response to small scale was the record-level targeting of intervention participants to prevent one type of measurement dilution; the response to low intensity was – in some cases – the administration of CLASSE to prevent the other form of measurement dilution.

To reiterate a point made earlier, NSSE has demonstrated value as a tool for identifying major engagement themes (benchmarks) and more specific engagement issues (items) through university-wide and program-level analysis of institutional differences; and it has triggered a truly significant level of discussion and exploration of the factors underlying such differences. However, NSSE is clearly not sufficiently sensitive to detect the scale and intensity of the interventions analyzed in this project. If there is an intensity threshold above which NSSE might reasonably be expected to detect effects, the interventions did not achieve it. It could be argued that the robustness and stability of NSSE are prerequisites for its primary purposes and that expectations of sensitivity even within targeted groups are inconsistent with this. The response to such an argument is that it was necessary to assess NSSE's sensitivity in order to assign it and other measurement tools their appropriate role in assessment and subsequent accountability activity.

CLASSE – which is really just another version of NSSE appropriate for course-specific analysis – was extremely useful in assessment of the course-based interventions. CLASSE did not detect all expected or desired engagement and academic outcome measures, and in some cases generated results that were puzzling enough to warrant

further analysis. However, the preponderance of evidence in projects utilizing CLASSE is that it is sufficiently sensitive to detect many key experimental effects. Student engagement results from a complex mix of course content and delivery, curriculum structure, service and "atmosphere", personal relationships and the integration of academic and social experiences. CLASSE clearly fits within the course-level dimension of the mix and was an effective measurement tool for those interventions able to use it.

5.3 Sample Sizes and Survey Response Rates

In several of the projects, original sample size objectives were not achieved because of lower-than-expected participation/take-up rates and survey response rates, and the occurrence of missing data on items necessary for propensity matching or regression covariates. In some cases, experimental group size exceeded candidate control group records resulting in matched groups of smaller than ideal size. In one intervention, extremely low survey response at the beginning of the administration period put the project at risk of termination. In another, take-up rates were low enough to warrant project termination. In retrospect, a more stringent net sample size criterion would have been appropriate in some cases.

There is no single "best" sample size criterion, and no firm rule for reconciling increased sample size and more powerful statistical analysis with higher cost. To some extent, sample size requirements are driven by intervention intensity and effectiveness, since detection of very large experimental differences can survive substantial loss in sample size. But sample size is also a function of intervention design and survey field practices.

One approach to intervention design is to work backwards from final sample size requirements to initial experimental and control group definitions (substituting a locally appropriate set of values for those used here as examples):

- Assume a minimum net sample requirement of 100 for each of the experimental and control groups;
- At 33 per cent survey response rate, a sample frame of 300/300 is necessary;
- At 20 per cent participation attrition or miss rate, revise sample frame upwards to 360/360;
- At 25 per cent record loss due to missing data required for propensity matching and for individual survey item non-response, revise estimate to 450/450;
- In order to perform propensity matching without loss of experimental "n", increase the size of candidate control group by 40 per cent, yielding sample size requirements of 450/630;
- In the special case of dual-survey administration where responses are to be linked (e.g., the Guelph project), increase sample sizes by 50 per cent to about 675/945;
- As appropriate, make further adjustments for non self-identified survey response.

An intervention target group requirement in excess of 400 students, and identification of an even larger control group, effectively limits the selection of course- and service-based interventions. Over time, as more and more interventions are undertaken and assessed and critical success factors are identified, sample size constraints and assessments of the type performed here could possibly be relaxed.

The University of Western Ontario project demonstrated the impact of non-response bias on survey-based assessment. Although this bias can be accommodated when interpreting intervention impacts, it is nonetheless appropriate to take steps to maximize survey response. Response rates appear to be part and parcel of institutional culture (with NSSE administrations typically generating 20 per cent - 70 per cent rates) and local conditions and regulations (with CLASSE administrations resulting in response rates of between 10 per cent and 60 per cent in the intervention projects).

5.4 Propensity Matching

Propensity matching proved to be an essential component of the assessments. Pre-match analysis of successive cohorts demonstrated that even large close-in-time cohorts often can differ significantly with respect to one or two variables, and show insignificant but by no means irrelevant differences on others. Pre-match analysis of cross-sectional control and experimental groups showed similar differences. In both situations, propensity matching eliminated all significant item differences, and removed concern over near-significance on other match variables.

In cross-sectional designs where the possibility of self-selection or predisposition bias exists, propensity matched groups appear unavoidable if experimental effects are to be isolated. As the Guelph example demonstrates, moving from unmatched samples to matched samples, and then to the inclusion of additional controls for any remaining predisposition effects provided greater confidence that the effects measured were in fact the results of the intervention.

The variables utilized for matching in the interventions were those that were available: demographic and academic items from the student records system. It is generally accepted that academic performance, gender, prior academic history and full-/part-time status affect both engagement and predisposition for self-selection, but they remain surrogates for this predisposition and are therefore not perfect controls. In some cases, the rate of missing data on these items prevented them from being employed in matching. Ideally, more variables would have been available and utilized in matching; however, student records systems may not be viable sources for them. As such, additional questions on customizable survey instruments seem a better option for identifying and assessing additional potential matching variables.

This project could have foregone propensity matching entirely in favour of multivariate regression analysis in which matching variables are included as covariates. (Though not reported in this analysis, multivariate models were run on several unmatched and matched group files in order to ensure that the matching had worked as intended. In many cases, the demographic and academic performance variables were significant pre-match; in all cases they were insignificant post-match.) The advantage of propensity matching followed by (generally) simple regression analysis over multivariate analysis is simplicity in model specification and interpretation without any loss of statistical power.

Matching in this project was conducted on control and experimental groups in the aggregate. It seems appropriate that intervention assessment also incorporate a second

level of matching within the experimental group to provide greater clarity with respect to intensity of involvement effects after controlling for possible intensity of involvement bias (which is, in effect, another dimension of self-selection bias).

Matching was generally achieved without significant loss of experimental records: utilization rates as high as 100 per cent and occasionally as low as 55 per cent were achieved (in the latter case because of limited control group size rather than fundamental differences between control and experimental groups). Slight loss of sample size due to matching appears to have been a reasonable tradeoff against possibly inconclusive analysis based on unmatched samples or more complex analysis utilizing a half-dozen or so covariates.

5.5 Intervention and Assessment Design

As noted earlier, the purpose of this project is not to assess interventions per se, but to explore the sensitivity of various measurement tools to intervention effects. Thus, while there are no "bad" interventions, there are interventions whose effects are more difficult to identify. The analysis has provided insight into five different dimensions of intervention and assessment design. The first (as noted above) is that larger sample sizes and greater intervention intensity make any effects more likely to be detected, and that the strength of NSSE is clearly not in identifying such small scale and low intensity intervention impacts.

The second is that successive cohort and cross-sectional designs both worked well, and that neither appears to have an intrinsic advantage. Successive cohort designs increase sample size potential by utilizing each cohort separately instead of splitting a single cohort. Survey activity and time commitment are effectively doubled; longer timeframes increase the potential for disruptive events to occur; and trends in the engagement background must be incorporated at least subjectively into the analysis. Cross-sectional designs may be more subject to self-selection bias (depending on participation rate) and because of cohort splitting may generate smaller experimental groups; however, survey field times are greatly reduced. Local circumstances appear to be the only key criterion in selection of one design over another.

Third, pre-post measurement can be applied in both successive cohort designs (as a more formal control for possible changes in the engagement background), and in cross-sectional designs (to accommodate predisposition effects). Only the Guelph project utilized a pre-post design. It provided significant insight into the impact and value of propensity matching, and greater confidence in the results because of assurances that predisposition bias had been accounted for.

Fourth, those projects relying solely on NSSE as a measurement tool showed no experimental effects, and absent any other quantitative measurement tools (e.g., other surveys, academic outcomes), little more about them can be explored or discussed. Until consensus emerges on the measurement tools most likely to demonstrate experimental effects in various circumstances, intervention design should as a matter of course incorporate additional measurement tools.

Finally, it is essential that interventions be designed to satisfy rigorous statistical criteria, and that those criteria be maintained throughout the implementation process. An intervention design involving compulsory participation (and hence, elimination of self-selection bias and at least some intensity of involvement measurement problems), for example, is unlikely to show experimental effects if a decision is made mid-point to move to a voluntary participation model with no tracking of participants or participation intensity.

5.6 Intensity of Participation Measures

The intensity of involvement analyses undertaken for several projects (e.g., Carleton, Guelph and Western Ontario) were suggestive but not entirely conclusive. Some intensity measures were highly subjective, while others were at best partial measures (e.g., frequency of attendance at, but not level of participation in, intervention activity). None were undertaken using within-group propensity matching. While the results suggest in general terms that intensity of involvement matters and that it can be detected with the right tool, this is clearly an area in which further work is needed: both in the construction of an appropriate intensity measure for a given intervention, development of an instrument containing the appropriate questions, the collection of participation data (through, for example, attendance/activity tracking or additional post-experiment surveying), and an analysis of factors associated with varying levels of involvement. Without distinguishing between the initial target group and the actual (participating) experimental group, and without differentiating intervention involvement levels, measurement dilution will occur that could mean the difference between conclusive and inconclusive analysis.

There is another obvious advantage to well-developed intensity measures. In the Ottawa intervention for example, FSS+ students could take advantage of up to a dozen specific services. Detailed participation data would permit not only statistical assessment of FSS+ based on varying levels and configurations of service usage, it would inform program revision based on these service usage patterns.

5.7 Analytical Methods and Findings

The decision to utilize propensity matched groups in (primarily) bivariate regressions was made early on in the project. As noted above, it simplifies model specification and interpretation relative to the use of multivariate analysis without propensity matching. Similar experimental effects would have been obtained with tests on the differences between means and proportions, but regression analysis provides a key piece of additional information – R^2 – that helps place the assessment in context.

R^2, the coefficient of determination, is a measure of the proportion of total variation in the dependent variable that is explained by the regression model. A model predicting the value of an engagement variable (e.g., frequency of involvement in class discussions) as a function of experimental participation where the R^2 value is .06, indicates that participation has explained six per cent of total engagement variation, leaving 94 per cent unexplained. Virtually all the regression models summarized in this report generated relatively low R^2 values. Propensity matching provides assurance that variation in engagement is not a function of (for example) age, gender or academic performance, which if included in the

models would have increased total R^2 but not "meaningful" R^2. Low R^2 reflects what is often referred to as a low "signal-to-noise" ratio. Non-modeled factors affecting academic engagement – noise – including student life circumstances, personality type, location (residence vs. commuter), and a host of others clearly generate significant engagement variation relative to the intervention itself – the signal. The approach used in this project was to judge intervention effects on the basis of a "preponderance of evidence", which requires a sufficiently large number of significant experimental effects in the appropriate direction. The NSSE analyses, which generally identified just one or a few significant variables, were by this approach deemed inconclusive. Analyses of other projects indicated significance not just for a number of individual variables, but also (given certain variables were multicollinear) for several distinct thematic clusters of variables. The preponderance of evidence approach suggests these results are fairly conclusive.

Most of the projects could have been subject to numerous additional forms of statistical analysis. The pooled course data in the Guelph analysis was performed using mixed model regression, an area in which the author is by no means an expert. A wider range of predictor variables could have been modeled for each intervention, up to and including virtually all engagement and experience items and benchmarks in the NSSE response set. And some of the analyses undertaken, including the analysis of non-response bias in the Western Ontario project, were not even envisaged at the outset of the project. Readers are invited to suggest improvements to the methods employed here, and to explore and document other methods in their own activities.

5.8 Planning and Design Uncertainties

Universities do not always provide a stable or predictable environment for engagement experimentation. An important response to uncertainty, contributed by one of the site participants in a progress report, is clear executive-level support to ensure ongoing commitment to project implementation and to assist participants in avoiding the kinds of events that would disrupt the project. The degree of "executive-level support" to the intervention projects was in some cases fairly substantial but in others virtually non-existent. It seems reasonable to suggest that future projects should be supported by the university's senior academic administration in order to assist in dissemination and follow-up, reinforce strategic relevance and incent others to adopt new practices. Uncertainties will likely continue to affect project completion, including those that affected the projects documented here: a labour dispute, low take-up rates, declining participation over time (attrition), unexpectedly low survey response, reassignment of instructors to different courses or courses to different semesters, and difficulties encountered in the classroom administration of CLASSE.

5.9 Organizational Learning

One identified dimension of project success (and a criterion in proposal approval) was the extent to which the project contributed to engagement-related organizational learning and commitment. While most proposals anticipated that such learning would occur, unfortunately most projects appear not (or not yet) to have achieved significant success in

this regard. Anecdotal accounts suggest that one project became known locally as "Fred's project", suggesting limited buy-in by others and limited positioning of the project within the university's broader quality agenda. Perhaps this response is typical of first-time pilot efforts, and it may change once this report is widely available and/or after more projects are implemented. It may also be a function of project scale, since larger or more intensive projects would almost by definition require department-wide (rather than individual or small group) involvement.

5.11 Intervention Cost and Funding Support

The Ontario post-secondary environment, as with several other jurisdictions, is experiencing serious financial constraint driven by a combination of grant funding levels, significant enrolment growth at both the graduate and undergraduate levels and the 2008 decline in financial markets. While commitment to quality remains an overriding issue in Ontario universities, engagement activity must nonetheless compete with unit budgeting and other pressing concerns. In this context, the extraordinary time commitment by all those involved at the participating universities over the past two and a half years is remarkable. It has also become clear that this project could not have been undertaken and would certainly not have been completed without the substantial financial support provided by the Higher Education Quality Council of Ontario. Now that this first round of activities is complete, attention must turn to maintaining and growing the universities' commitment to improved engagement practice.

The author's view is that ongoing experimentation and organization learning need to be encouraged and enabled through Ministry grant funding either through additional base funding or in the short term, through a special funding envelope allocated to approved projects following a peer review process coordinated by an informally constituted or formal (Council of Ontario Universities) university-led group.

5.12 Process vs. Outcome Accountability

Prior to this project, there was no firm evidence that the engagement or academic outcome impacts of modest interventions could be detected through the use of existing measurement tools. This project has shown that certain measurement tools, and it is hoped others that will be developed, are capable of detecting at least some of these impacts at least some of the time and as a result, that engagement interventions can become an explicit and demonstrable quality improvement strategy over time. The focus of accountability in Ontario in recent years has been on documenting quality efforts rather than outcomes. While the results of some of the assessments are certainly encouraging, it is the author's opinion that it would be premature to move immediately toward an outcomes-based accountability regime. Some interventions showed the desired results, but more are needed to identify general patterns in the types, scales and impacts of interventions – that is to identify best practice in design, implementation and measurement for a wider variety of interventions. And until and unless intervention scale is substantially increased, the dilution effect of NSSE-based measurement will remain a concern for university-wide performance measurement. A simple scenario demonstrates the difficulty

of moving to conventional (university-wide) NSSE-based outcomes accountability. If we assume

- 500 of 5,000 first year students are subject to a relatively focused intervention (a much larger number than most of the interventions studied here);
- Three of six items (measured on a 5-point Likert scale) within a benchmark show an increase in the experimental group averaging 0.2 (toward the high end of the range of coefficients in the regression models);
- A pre-experimental benchmark value of 50.0;

then the university-wide benchmark score would increase from 50.0 to 50.6, a difference far smaller than that typically observed in benchmarks over successive administrations. If limited to the experimental group alone, the benchmark would increase to 56.0 – a substantial increase but one limited to a small proportion of the total student population. Therefore, project-specific outcomes measurement (for specific courses or student subgroups) should remain the focus, along with an acceptance that meaningful aggregate benchmark movement will only occur as a result of repeated successful improvements.

5.13 NSSE and Multi-Year Agreements in Ontario

The structure of Ontario's multi-year accountability agreements is currently (as of February 2010) under review. NSSE played a limited role in the first round of MYA's. All Ontario universities are committed to periodically administering the survey; a limited number identified the maintenance or improvement of NSSE benchmark performance in the "strategies/targets/outcomes" portion of the agreement; and a few identified specific NSSE items as the intended focus of improvement efforts. As the next round of MYA's is being developed, several key issues should be kept in mind.

NSSE is an important tool in engagement and quality analysis, but should not be the only basis for engagement assessment or reporting. As several interventions indicated, other survey tools, objective skills development and knowledge acquisition "tests", and the CLASSE instrument itself provide useful and more "localized" measures of both quality processes and quality outcomes that can be both intuitively and empirically linked to the engagement agenda. The currently limited number of quality assessment tools can and should be increased over time through further research and experimentation.

The service dimension of engagement – which was the focus of the Ottawa and Windsor projects for example – did not have access to an engagement assessment tool as the course-based interventions did (though the availability of CLASSE came late enough that it could not be implemented in all those interventions that might have benefitted from it). While numerous questions in the NSSE instrument deal with the service aspects of engagement, they are relatively general (particularly those comprising the SCE benchmark) and cannot be linked with specific service initiatives to support service assessment. The MYA's should accommodate the limited number of standardized measurement tools.

As noted, NSSE has high value in broad benchmarking and comparative analysis and as a tool for generating discussion within the university community. However, its apparently limited value as an assessment tool for small targeted interventions suggests the need for a "second tier" of engagement surveys. As CLASSE does for single or multiple course engagement measurement, other more specific and perhaps customizable survey templates could support deeper analysis of various other engagement issues. NSSE would remain the "entry point" for engagement analysis and the tool of choice for periodic monitoring. However, there may be an opportunity for staff at NSSE, in concert with users, to construct a comprehensive second tier of engagement surveys under the overall NSSE umbrella, addressing, for example:

- Supportive services (counseling, program advising, institutional climate, deeper dimensions of staff and faculty interactions);
- Out-of-class peer interactions (additional aspects of study groups and peer support networks);
- Enriching experiences (differentiation of the scale of the experiences, and the intensity and frequency of student involvement);
- Intervention-focused engagement surveys (tailored to a limited number of intervention types).

While institutions could certainly develop such tools themselves, the obvious value of expert design and the research and practice value of (reasonably) standardized assessment measures would contribute to improved practice.

Many Ontario institutions appear to have moved beyond a pure process-based engagement accountability approach, but for the reasons noted above, cannot yet adopt a pure outcomes-based one either. The analysis and findings of this project suggest a middle path that is consistent with the proposed strategic orientation of future MYA's and with diverse institution-specific measures. It seems reasonable for Ontario institutions to

- Provide clear evidence that engagement issues (strengths and weaknesses) have been identified and communicated and that specific documented institutional efforts are in place (i.e., process-based accountability);
- Conduct periodic assessments of selected institutional efforts using measurement tools as available and relying on best practice in design and assessment (i.e., outcomes-based accountability, but specific to the outcomes that can be measured with the available tools);
- Administer NSSE at least every 3 – 5 years to monitor and benchmark overall institutional and system performance;
- To incorporate the three items above into MYA's within the context of each institution's distinct mission and strategy and the resources available for such activities.

6. Selected Web References

University Comparisons and Rankings

Maclean's Magazine	oncampus.macleans.ca/education/rankings/
US News and World Report	www.usnews.com/sections/rankings/
Canadian University Report	www.globecampus.ca/navigator
Times Higher Education Supplement	www.topuniversities.com/worlduniversityrankings
Shanghai Jiao Tong University	www.arwu.org/rank2008/EN2008.htm
Research Infosource	www.researchinfosource.com/top50.shtml
BusinessWeek Magazine	www.businessweek.com/bschools/rankings

National Survey of Student Engagement

Home Page	nsse.iub.edu/
Origins and Background	nsse.iub.edu/html/origins.cfm
Concepts and Psychometrics	nsse.iub.edu/pdf/conceptual_framework_2003.pdf
Annual Reports	nsse.iub.edu/html/annual_reports.cfm
Applications	nsse.iub.edu/pdf/Using_NSSE_Data.pdf
Annual Administration Info	nsse.iub.edu/nsse_2009/index.cfm (any year 1999 - 2009)
Survey Instruments	nsse.iub.edu/html/survey_instruments_2008.cfm
Stephen Porter paper At ASHE 2009	srporter.public.iastate.edu/surveys/porter_ashe_2009.pdf

Ontario-Specific Documents

Report of the Rae Review	www.edu.gov.on.ca/eng/document/reports/postsec.pdf
Reaching Higher Plan	www.premier.gov.on.ca/news/Product.asp?ProductID=114
HEQCO – Home Page	www.heqco.ca
HEQCO – NSSE Workshop	www.heqco.ca/en-CA/Events/Past per cent20Events/Pages/PastEvents_April27_2007.aspx
HEQCO – Research Reviews	www.heqco.ca/en-CA/Research/Review per cent20and per cent20Research per cent20Plan/Pages/default.aspx
Multi-Year Agreements (Samples)	www.queensu.ca/irp/accountability/reg_compliance.htm

Participating University Accountability and/or Institutional Research Web Sites

Carleton University	oirp.carleton.ca www2.carleton.ca/about/administrative/public-accountability.php www2.carleton.ca/surveys/html/surveys.htm
University of Guelph	www.uoguelph.ca/analysis_planning/
University of Ottawa	www.uottawa.ca/services/irp/eng/

www.uottawa.ca/services/irp/eng/research/accountabilityeng.html

Queen's University	www.queensu.ca/irp
Ryerson University	www.ryerson.ca/upo www.ryerson.ca/about/accountability
University of Waterloo	www.analysis.uwaterloo.ca www.uwaterloo.ca/accountability
University of Western Ontario	www.ipb.uwo.ca www.uwo.ca/pvp/accountability.html
Wilfrid Laurier University	www.wlu.ca/homepage.php?grp_id=2297&pv=1 www.wlu.ca/homepage.php?grp_id=169
University of Windsor	www.uwindsor.ca/info www.uwindsor.ca/president

Key Items from Previous Project Reports

Proposal Submission Form
Proposal Evaluation Form
US Implementation Practice Inventory
www.queensu.ca/irp/NSSEinterv/
Phase One Evaluation Survey
Data Submission Requirements
Post-Intervention Site Participant Questionnaire

University Contact Emails

Carleton University	Bruce Winer bruce_winer@carleton.ca Natasha MacDonald Natasha_macdonald@carleton.ca
University of Guelph	Brian Pettigrew brian.pettigrew@exec.uoguelph.ca Kelly Parkinson kparkins@registrar.uoguelph.ca
University of Ottawa	Pierre Mercier nitch@uottawa.ca
University of Windsor	Clayton Smith csmith@uwindsor.ca Rosemary Zanutto rosez@uwindsor.ca
Queen's University (Psychology)	Jill Atkinson jill.atkinson@queensu.ca
University of Waterloo	Mary Jane Jennings mjjennin@uwaterloo.ca Jennifer Kieffer j2kieffe@uwaterloo.ca
Ryerson University	Usha George ugeorge@ryerson.ca

	Janice Waddell jwaddell@ryerson.ca
Wilfrid Laurier University	Wally Pirker wpirker@wlu.ca
	Gail Forsyth gforsyth@wlu.ca
University of Ottawa	Pierre Mercier pierre.mercier@uottawa.ca
	Marie-Eve Meunier marie-eve.meunier@uottawa.ca
University of Western Ontario	Martin England england@uwo.ca
	Tom Haffie thaffie@uwo.ca
Queen's University (Engineering)	Naraig Manjikian naraig.manjikian@queensu.ca

Other

Propensity Matching Macro http://mayoresearch.mayo.edu/biostat/upload/gmatch.sas