University of Alberta

Confirmatory Factor Analysis of the Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q): Examining the Internal Structure within a Canadian Undergraduate Population

by

Ulemu Luhanga

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Education
in
Measurement, Evaluation and Cognition

Department of Educational Psychology

Fall 2012
Edmonton, Alberta

ABSTRACT

To ensure academic quality within the Canadian undergraduate education context, there is a need for student evaluation questionnaires that help to enhance students' teaching-learning environments. The Experiences of Teaching and Learning Questionnaire (ETL-Q) and the Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q) were designed to meet a similar need in undergraduate courses in UK. Though studies have examined the utility of these questionnaires across different countries, the utility of these tools has yet to be ascertained in Canadian undergraduate populations. Hence, the goal of this study was to examine the structure of SR-ETL-Q responses within a Canadian sample. This goal was achieved by comparing the fit of two competing models, a six-factor versus a five-factor model using confirmatory factor analyses. Results indicate that the theorized six-factor model provided a better fit for the Canadian sample. Implications of these results and recommendations for future research are also discussed.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my parents, Drs. Florence and Pearson Luhanga for being such amazing academics and role models! I hope to emulate you both in my own career. To the rest of my family, both immediate and extended, thank you for continuing to support my academic journey. A special thank you goes out to my brother, Penjani Luhanga, who played many roles (editor, mock supervisor, sounding board) as I wrote this thesis. To my “better” half and my rock, David On'gany, thank you for being my perpetual cheerleader as I continue to accumulate degrees ... I promise ... just one more ... then I am done!

To my mentors Dr. Jacqueline Leighton and Dr. Todd Rogers, thank you for being readily available to provide guidance and support during my time in the Centre for Research in Applied Measurement and Evaluation (CRAME) and for the continued support I get as I embark on the next chapter in my journey. To my co-supervisor, Dr. Cheryl Poth, thank you for bringing me on to the research project that lead to my thesis topic, and for your valuable comments as I worked through writing my thesis. To my committee members, Dr. Genevieve Gauthier and Dr. Paul Newton, I thank you.

Finally, to my fellow CRAMErs, past and present, thank you all for making my time in CRAME not only educational but also … lots of fun!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

In undergraduate education, academic quality is an important, complex and often times difficult-to-measure concept due to its multi-level, multi-dimensional, value-laden nature (Law, 2010; Vlăsceanu, Grunberg, & Parlea, 2007). It is multi-level in the sense that policies and practices surrounding academic quality impact all levels of an institution; from the lowest level (i.e. the course level) where the focus is generally on enhancement, to the highest level (i.e. the institution level) where the focus is generally on accountability (Law, 2010). Academic quality is also both multi-dimensional and value-laden due to various institutional stakeholders (e.g. students, faculty, administration, government) having different ideas about what constitutes as an indicator of quality (Law, 2010). Harvey and Green (1993) summarized the various conceptions of academic quality that a variety of academic stakeholders, including internal stakeholders (i.e. students, faculty and administration) and external stakeholders (i.e. government, funding agencies and accreditation agencies) had. The authors grouped stakeholders' conceptions of academic quality into five interrelated definitions: as exception, as perfection, as fitness for purpose, as value for money, and as transformative (Harvey & Green, 1993). These definitions were later merged, by Harvey and Knight (1996), into a single higher order conception: *quality as transformation*.

Quality as transformation refers to "a qualitative change; [wherein] education is about doing something to the student as opposed to something for the consumer [and] includes concepts of enhancing and empowering” (Kis, 2005, p. 4). Harvey and Knight (1996) argued that by adopting a unitary conception of academic quality as transformation, the other four definitions of quality would become “partial indicators of the transformation process at the heart of quality” (Harvey & Knight, 1996, p. 26). Under this unitary conception, Harvey and Knight (1996) went on to discover that both internal and external academic stakeholders agreed that academic quality should be related to the learning process; consequently these authors (and others) have called for the assessment of academic quality to pay particular attention to students’ experiences of learning and their learning environments (Harvey & Knight, 1996; Tam, 2001).

FEEDBACK FROM STUDENTS

In an attempt to assess academic quality through students’ perceptions of their experiences, *feedback from students* is often collected. Feedback from students refers to “the expressed opinions of students about the service they receive as students” (Harvey, 2011, p. 4). Traditionally feedback from students has been collected to provide: diagnostic feedback to faculty that will provide information for improvement of teaching; a measure of teaching effectiveness to be used in personnel and administrative decision making; information for students to use in selection of courses and teachers; and, as an outcome or process description for research on teaching (Kember, Leung, & Kwan, 2002; Law, 2010; Rowley, 1996, 2003). However, due to increasing recognition of the importance of feedback from students to assess academic quality, additional reasons for collecting this form of feedback have come to include: to provide auditable evidence that opportunities for student comment have been available for students to express their opinions on their courses, and that the data collected inform action and quality enhancement; to encourage students to reflect on their learning, and thereby to enhance their awareness of their own learning processes, and the factors that lead to positive or negative outcomes in such a way as to develop their learning competencies; to provide students, as customers, with an opportunity to express their level of satisfaction with a learning experience; and to benchmark institutions and generate other indicators of quality that may contribute to the marketplace reputation of the university (Rowley, 2003).

Though feedback from students, has the potential to provide a wealth of information for quality enhancement and improvement initiatives, there is evidence to suggest that the most commonly collected form of feedback from student, traditional student evaluation questionnaires, are failing to contribute to the improvement of teaching and learning (Kember et al., 2002; Rowley, 2003). Traditional student evaluation questionnaires have been criticized for reinforcing "traditional, lecturer-led modes of learning … rather than encouraging critical thought, and student-centred learning" (Rowley, 2003, p. 145). Furthermore, these questionnaires are often designed and used for staff appraisal processes. The emphasis on staff appraisal is often considered judgmental and summative and this may hinder the utility of the questionnaires as a means of improving teaching and learning due to faculty developing negative perceptions of these questionnaires (Kember & Leung, 2008; Kember et al., 2002; Richardson, 2005).

In an attempt to address one of these limitations, and in line with the conception of quality as transformative, Kember et al. (2002) and Kember and Leung (2008) suggest that there is a need for theory-based student evaluation questionnaires that are focused on measuring the quality of learning and teaching (i.e. academic quality). By designing theory-based questionnaires in this manner, the diagnostic power (i.e. the capability to distinguish strengths and weaknesses pertinent to quality concerns) of student evaluation questionnaires will be enhanced such that these instruments can then be used to identify aspects of students' teaching and learning environments that need attention and thus guide actions for quality improvement (Law, 2010). By adopting this type of theory-based questionnaire, higher education institutions (HEIs) can be provided with a proxy measure of academic quality.

IN THE CANADIAN HIGHER EDUCATION CONTEXT

In Canada, the method most commonly used by HEIs to collect evidence of academic quality is the minimum standards approach (Finnie & Usher, 2005). This approach "is the foundation of the periodic review process (often called "cyclical reviews" or "program reviews") that all Canadian universities use at a departmental level" (Finnie & Usher, 2005, p. 6). The process includes two stages: a self-audit and an external review (Baker & Miosi, 2010). The self-audit involves gathering information on the quality of a program from a number of different sources including students, alumni and faculty (Finnie & Usher, 2005). Within this self-audit stage, traditional student evaluation questionnaires are commonly used as a source of evidence and are considered a common tool for engaging students in the quality assessment process (Canadian Council on Learning, 2009).

Another means of collecting evidence of academic quality through feedback from students is through the learning impacts approach which can be used in conjunction with the minimum standards approach. The goal of the learning impacts approach is to judge academic quality by asking students about their overall experiences at HEIs (Finnie & Usher, 2005). Though the focus of the learning impacts approach is on students' experiences as a measure of academic quality, the most commonly used instrument, National Survey of Student Engagement (NSSE), is focused on students' overall experiences at the institution-level rather students' learning experiences at the programme or course level (Coates & Seifert, 2011).

To collect evidence of academic quality in Canadian HEIs, particularly at the course level, the most commonly used student evaluation questionnaires are student ratings of instruction (also commonly referred to as student evaluations of teaching – SETs) (Gravestock & Gregor-greenleaf, 2008). Though diligently collected, SETs like other traditional student evaluation questionnaires are designed primarily for staff appraisal processes. Consequently, SETs are not designed to identify aspects of students' teaching and learning environments that need attention, and therefore lack the diagnostic power needed to allow these instruments to be used as measures of the unitary conception of academic quality. To address this limitation within the Canadian undergraduate education context, there is a need to evaluate the usefulness of other student evaluation questionnaires that do possess the diagnostic power needed to enhance students' teaching and learning environments (Finnie & Usher, 2005; Gravestock & Gregor-greenleaf, 2008).

The *Experiences of Teaching and Learning Questionnaire* (ETLQ) (Entwistle, McCune, & Hounsell, 2002) and a shortened version of the ETLQ entitled the *Short Revised Experiences of Teaching and Learning Questionnaire* (SR-ETLQ) (Entwistle, 2009) are two questionnaires with sections that are purported to improve the quality of teaching and learning in undergraduate courses. Though several studies have provided various forms of validity evidence to support the use of ETLQ (and SR-ETLQ) responses for improving the quality of teaching and learning within undergraduate populations across different countries (e.g. UK, China, Finland), to date, the generalizability of this tool has yet to be ascertained in Canadian undergraduate populations.

Given the identified need within the Canadian undergraduate context for questionnaires that can provide diagnostic information that can be used to enhance students' teaching and learning environments, it is appears that the ETLQ or SR-ETLQ would be well suited to meet this need. However, before adopting either questionnaire to this Canadian context, cross-cultural adaptation and validation are required. Cross-cultural adaptation and validation is required when an instrument is to be used within a new culture, language and/or country to ascertain the equivalence of questionnaire interpretations and thus the utility of the questionnaire to the new setting (Beaton, Bombardier, Guillemin, & Ferraz, 2000). Thus, though the ETLQ and SR-ETLQ were developed in the UK, in the

English language, cultural differences between UK and Canada could affect the utility of this questionnaire due to item responses being interpreted differently.

To examine the equivalence of ETLQ response structures and interpretations, studies that have used the questionnaire in different countries (e.g. Parpala, Lindblom-ylänne, Komulainen, & Entwistle, in press; Xu, 2004), have conducted cross-cultural adaptation and validation that included the translation of the ETLQ to other languages followed by factor analyses to determine the equivalence of the questionnaire's underlying factor structure within these new settings. Consequently, in an attempt to begin to validate the use of ETLQ or SR-ETLQ responses within a Canadian undergraduate context, the goal of this study was to collect validity evidence that would examine the internal structure of responses within a Canadian undergraduate population.

CHAPTER TWO

LITERATURE REVIEW

INTRODUCTION

This chapter begins by introducing the theoretical and empirical background to the ETLQ (and SR-ETLQ). The chapter then focuses on assessing the utility of ETLQ (and SR-ETLQ) responses for improving the quality of teaching and learning within undergraduate populations. To build a case to support the implementation of either the ETLQ or the SR-ETLQ to a new context, the chapter discusses the validation process commonly used for student evaluation questionnaires, and then uses this validation process to organize the various sources of validity evidence currently available for ETLQ and SR-ETLQ responses. The chapter finally concludes with an examination of results from generalizability studies that analyzed the internal structure of ETLQ or SR-ETLQ responses in various undergraduate contexts.

STUDENT LEARNING RESEARCH

Within the umbrella term *student learning research* are four interrelated research areas: student approaches to learning (SAL), self-regulated learning (SRL), student engagement and, higher education teaching (Haggis, 2009; Trigwell, 2010). SAL research is focused on looking at the learning experience from the point of view of the students by using interviews and questionnaires; SRL research is focused on interventions that teach students to become effective self-regulating learners; student engagement research is focused on universities taking a holistic approach to student experiences (including their learning experiences); and finally, higher education teaching research is focused on the ways in which teachers act as medium to help students overcome barriers to learning (Trigwell, 2010). Using student learning research, coupled with academic quality research, a promising theory-based student evaluation questionnaire, the *Experiences of Teaching and Learning Questionnaire* (Entwistle et al., 2002), has been developed to measure the quality of teaching and learning (academic quality) in higher education environments.

EXPERIENCES OF TEACHING AND LEARNING QUESTIONNAIRE (ETLQ)

The ETLQ, was developed as part of a large scale UK project entitled *Enhancing Teaching and Learning in Undergraduate Courses (ETL)* project (Hounsell et al., 2005). The aim of the project was to "explore the idea of constructive alignment as a way of working with departmental colleagues to strengthen the teaching-learning environments (TLEs) experienced by undergraduate students, so as to enhance engagement, motivation, learning processes and outcomes, and levels of achievement" (Hounsell et al., 2005, p. 5). The ETLQ contains five sections that explore students' perceptions of the teaching-learning environment and their approaches to studying in a course using various Likert-type scales. The first section, entitled *Approaches to learning and studying* contains 18 items and uses a five-point Likert scale (1 = disagree to 5 = agree) to ask students to describe their approaches to studying within a course.

The second section, entitled *Experiences of teaching and learning* contains 40 items and uses a five-point Likert scale (1 = disagree to 5 = agree) to measure students' perceptions of the teaching and learning they experienced in the course. The third section, entitled ‘Demands made by the course unit’ contains 10 items that use a 5-point Likert scale (1=very difficult to 5 = very easy) and 1 open-ended item to ask about demands students felt that the course made. The fourth section, entitled ‘What you learned from this course unit’ contains 8 items that use a 5-point Likert scale (1=very little to 5=a lot) and 1 open-ended item to ask students what they felt they gained from the course. Finally the fifth section is a single 9-point scale item (1 = rather badly to 9 = very well) to ask students how they felt they had done overall in the course (Entwistle et al., 2002).

The 40-item section of the ETLQ, entitled *Experiences of teaching and learning* (henceforth referred to as the ETL-Q) is of relevance to the current study as it is this section of the ETLQ that has previously been used to monitor the quality of learning and teaching (i.e. academic quality) by focusing on students’ perceptions of their teaching-learning environments.

THE VALIDATION PROCESS

Validation studies provide an important means of collecting evidence to support the adoption of a questionnaire for particular intended uses (Kane, 2006). The validation process “involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations” (AERA, APA, & NCME, 1999, p. 9); it is therefore the responses (scores) from a questionnaire and their interpretation(s) that must be validated. Furthermore, as the responses are a function of items, persons responding to items and the context of measurement, it is the validated interpretation(s) that can provide meaning to an instrument within a given context, and moreover, evidence for the generalizability of these interpretation(s) over time and across groups or settings (Messick, 1989).

Kane (2006) conceptualizes the validation process as two interrelated arguments: an interpretive argument and a validity argument. The interpretive argument involves the development of inferences that lead from observed scores on an instrument to interpretations made based on those scores, while the validity argument tests the interpretive argument and its corresponding interpretation by integrating individual studies that focus on various sources of validity evidence such as statistical analyses or content analyses (Kane, 2006). Within this conceptualization of validation, two approaches are particularly relevant to student evaluation questionnaires: theory-based approach and qualitative approach.

Theory-based approach. In a theory-based approach, student evaluation questionnaire items are used as indicators of a questionnaire's underlying theoretical construct, thus item response interpretations estimate aspects of a questionnaire's underlying construct (Kane, 2006). The interpretive argument used for these indicators is made up of five major inferences: (1) scoring – from observed score to indicator score; (2) generalization – from indicator score to universe score; (3) extrapolation – from expected indicator score to target score for indicator; (4) theory-based interpretation – from target score for indicator to construct as defined by theory; and (5) implications – from construct to any implications suggested by construct label or description (Kane, 2006). The validity argument evaluates these inferences by collecting analytical and empirical evidence to support the appropriateness of the indicators and their interpretations (Kane, 2006). Analytical evidence (e.g. through a review of literature or content analysis) is collected during the development of the instrument whereas empirical evidence is collected by analysing responses (scores) using most commonly correlational analyses (e.g. confirmatory factor analysis or multi-trait multi-method matrix) and experimental studies (Kane, 2006).

Qualitative approach. In the case of a qualitative approach, student evaluation questionnaires and their items are thought to be qualitative assessments which focus on "the interpretation of performance in context … instead of [depending on] the observation-scoring-interpretation paradigm" (Kane, 2006, p. 47). As the focus of the qualitative approach is on making interpretations in context, this approach is particularly useful when one wants to take into account the interactions among various contextual factors that influence the performance being assessed. Kane (2006) provides an example of a three-part interpretive argument for classroom assessments. This three-part interpretive argument can be adapted to the case of student evaluation questionnaires by switching the focus from teachers evaluating students to students evaluating faculty as follows:

1. Development of initial views of faculty: students use their conceptual frameworks to develop a view of the faculty. They use these frameworks to make sense of their initial interactions with faculty and to develop their initial views of faculty. The conceptual frameworks provide a basis for

evaluating faculty performance and for anticipating various factors that might influence faculty learning and performance;

2. Refinement of the students' views: students modify their views of faculty, based on their ongoing interactions with the faculty; the students' evolving views generate expectations. These expectations provide working assumptions for subsequent instruction and assessment, and comparisons of the expectations to subsequent observations provide feedback on the accuracy of the students' views; and,
3. Extension of the students' evolving view to new contexts: students may also use their overall assessments of faculty to draw conclusions about expected performance in new contexts (e.g. in a new course).

The validity argument in the qualitative approach would therefore involve "an evaluation of the completeness and coherence of the interpretive argument and an evaluation of whether the interpretive argument provides a reasonable explication of the proposed interpretation" (Kane, 2006, p. 48). According to Kane (2006), within the qualitative approach, conceptual frameworks developed by individuals to assess performance can be thought of as informal theories. Consequently the conceptual frameworks students develop to evaluate faculty and respond to student evaluation questionnaires can be thought of as the theoretical constructs that item response interpretations estimate.

Although either the theory-based or qualitative approach can be used to validate theory-based student evaluation questionnaires, this study will focus on adopting only one approach, the theory-based validation approach, to ascertain the validity argument for the ETL-Q.

SOURCES OF VALIDITY EVIDENCE

Within a theory-based approach to validation, there are six sources of analytical and empirical evidence that can be collected to build a validity argument: Content (using judgemental or logical analyses); Response processes (using probing techniques such as protocol analysis); Internal structure of responses (using correlational analyses); Relationship with other variables/measures (using correlational analyses); Consequences of interpretations and uses (using qualitative analyses); and Generalizability (using variety of techniques including experimental manipulations) (Messick, 1989).

VALIDITY EVIDENCE FOR THE ETL-Q (AND SR-ETL-Q)

To establish the strength of the validity argument for ETL-Q responses and their interpretations, it was necessary to first examine evidence from previous analytical and empirical studies that have utilized the ETL-Q. This was done to obtain a comprehensive account of the degree to which existing empirical evidence and theory supported the intended use of this questionnaire as a measure of academic quality (i.e. quality of teaching and learning).

Evidence based on content. The ETL-Q was designed as a course-level student evaluation questionnaire that was to be used to monitor the quality of learning and teaching by focusing on students’ perceptions of their teaching-learning environments (Entwistle, 2009; Entwistle et al., 2002). Quality of learning and teaching was operationalized as those features of a teaching-learning environment "most likely to encourage engagement with subject matter, a deep approach, and high quality learning" (Entwistle et al., 2002, p. 10). A heuristic model of features of a teaching-learning environment that relate to the subject content and how it is taught (left-hand side), and features that not only influence student learning but are also themselves influenced by structures external to the course such as institutional/departmental missions, objectives and strategic plans (right-hand side) are displayed in Figure 2.1 below.

The left-hand side of the heuristic model (Figure 2.1) focuses on those features related to the subject content and how it is taught. In particular, this section of the model was designed to show a logical progression from the base of the model (course preparation) to the top (course delivery) (Entwistle, 2009). The beliefs an instructor has about teaching and the role of the teacher influences how they think about their subject matter and its underlying pedagogy. This in turn influences how an instructor interprets the formal requirements of a syllabus (i.e. target understanding), which ultimately leads them to decide on course content and how to go about teaching that content (Entwistle, 2009).

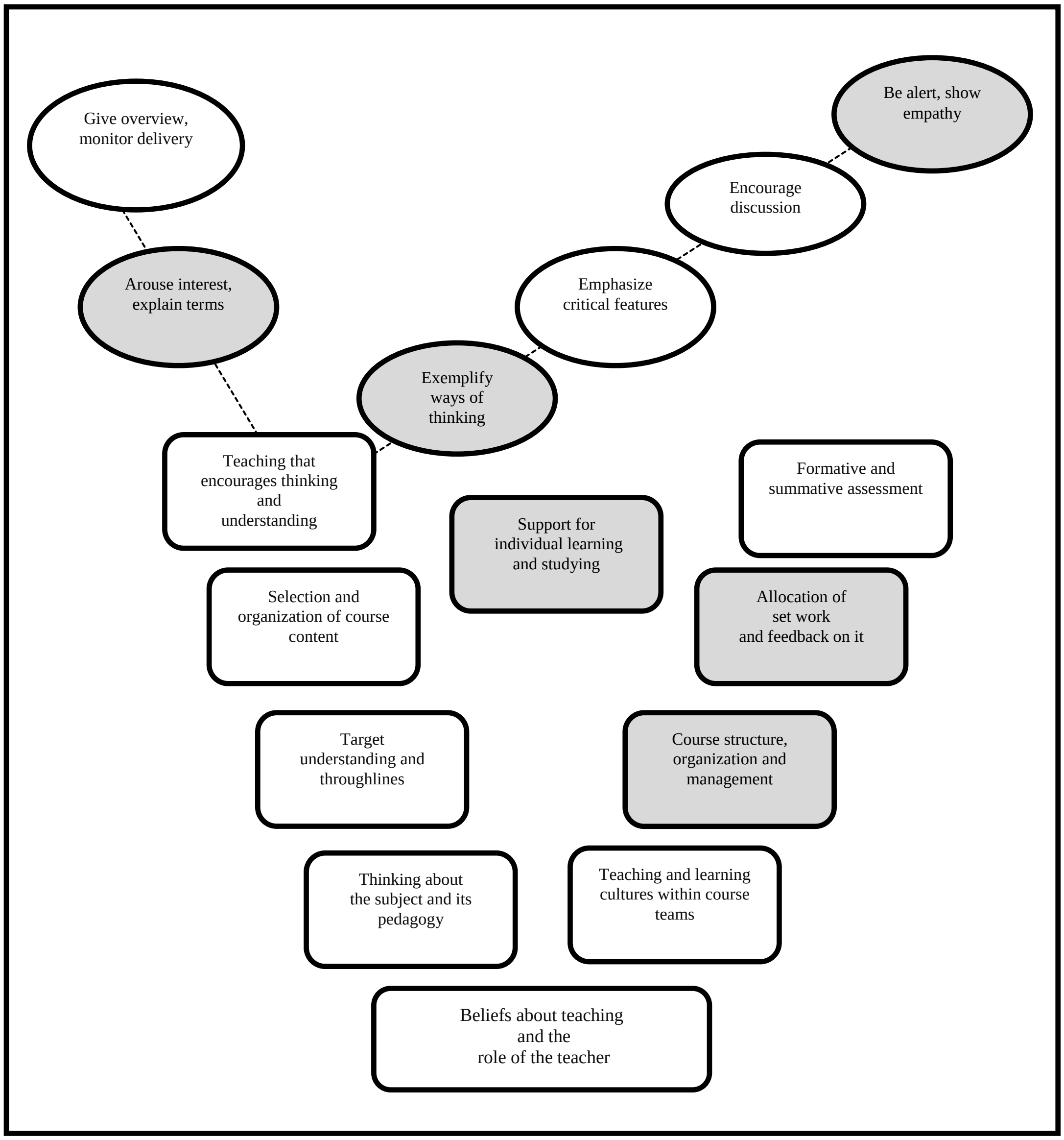

Figure 2.1. A heuristic model of features of teaching-learning environment that influence learning. Adapted from *Teaching for understanding at university: Deep approaches and distinctive ways of thinking* (p. 115), by N. Entwistle, 2009, New York, NY: Palgrave Macmillan. Copyright 2009 by Noel Entwistle.

With regard to teaching, the model expands on this concept by identifying critical aspects of teaching that encourage thinking (i.e. a deep approach) and understanding. These aspects (indicated by ovals in Figure 2.1) include: giving an overview and monitoring delivery; arousing interest, explaining terms and encouraging understanding; exemplifying ways of thinking and practising; emphasizing critical features and patterns of variation in content; encouraging individual reflection and group discussion; and, being alert while teaching and showing empathy with students (Entwistle, 2009).

On the right-hand side of the heuristic model (Figure 2.1) are those features of the teaching-learning environment that not only influence student learning but are also themselves influenced by structures external to the course such as institutional/departmental missions, objectives and strategic plans (Entwistle, 2009). In regard to course structure, organization and management, a course team (or individual instructor) must decide on the number and types of assignments and examinations, and their corresponding grading criteria. Furthermore, in making these decisions, an instructor or team must take into account perceived workloads and how to best ensure students receive timely feedback (Hounsell, Hounsell, Litjens, & McCune, 2005). Students expect assignments (formative assessment) and end of course examinations (summative assessments) to be transparent and fair, and will judge these facets by comparing their work with that of their fellow students, thus it is imperative that grading criteria be not only clear but also consistent (Entwistle, 2009). Finally, to provide students with support for individual learning and studying, instructors (or course teams) must encourage students to not only adopt skills such as self-regulated learning but to also seek support from their peers (Trigwell, 2010).

Although all these features work together to enhance student learning, the ETL-Q focuses on six features of the teaching-learning environment (highlighted in grey in Figure 2.1) that students are believed to perceive. Furthermore, to increase the diagnostic power of the ETL-Q, items representing these six features were designed to focus on particular teaching and learning activities believed to promote active, deep learning (Entwistle, 2003). By doing this, the aim was that the ETL-Q would be able to provide faculty with a means of identifying strengths and weaknesses in course preparation and delivery.

Evidence based on response processes. Within the literature on the ETL-Q, though the items were developed based on interviews with students (Entwistle, 2009; Hounsell et al., 2005), there are no studies to date that have researched the cognitive aspects of responding to the questionnaire items.

Evidence based on internal structure. Factor analyses were used to investigate the internal structure of responses for the original 40-item ETL-Q. Within the original context (British undergraduate programs), results of three studies provided evidence to suggest that two plausible solutions, a five-factor and a six-factor solution, could explain the internal structure of ETL-Q responses (Entwistle et al., 2002; McCune, 2003; Parpala, Lindblom-ylänne, Komulainen, & Entwistle, in press).

The initial exploratory factor analysis was conducted by Entwistle et al. (2002) using 472 British undergraduate students. Though the study was conducted with undergraduate students from Biology, Economics, Electronic Engineering, History, and Media programs in either their first or final years, the portion of students from each of these groups was not documented. In this analysis, exploratory factor analysis was conducted using maximum likelihood estimation followed by an oblique rotation to simple structure (Entwistle et al., 2002). This analysis resulted in five factors labelled: organisation and structure; encouraging learning; assessments and assignments; supportive climate; and, evoking interest. This five-factor solution explained 41% of the variance in responses (Entwistle et al., 2002).

The second exploratory factor analysis was conducted on a larger sample of 1828 British undergraduate students by McCune (2003). Once again, the distribution of students by program and year were not reported. Using maximum likelihood estimation with oblique rotation, once again a five-factor structure was found which explained 39% of the response variance. In this study, it was noted that the five-factor structure was slightly different as items related to supportive climate separated into two distinct factors related to student support and staff support respectively, while items related to encouraging learning and assessments and assignments came together to form one factor (McCune, 2003). The author chose to retain the original division between items related to encouraging learning and to assessments and assignments, but also introduced the split to items from the supportive climate factor thus resulting in six-subscales for which alpha values were as reported in Table 2.1 below:

Table 2.1

Alpha values for subscales of ETL-Q in McCune (2003) study

Factors (Entwistle et al., 2002)	Subscale (McCune, 2003)	Coeff. Alpha
Organisation and structure	Clear aims, organisation, alignment and integration	0.84
Encouraging learning	Teaching for understanding and choice	0.75
Assessments and assignments	Assessment for understanding, guidance and feedback	0.81
Supportive climate	Staff enthusiasm and support	0.77
Supportive climate	Support from other students	0.73
Evoking interest	Interest, enjoyment and relevance	0.81

The third study, conducted by Parpala et al. (in press), aimed to validate the ETL-Q in different countries. Within this study, the British sample used consisted of 2710 undergraduate students who came from eleven universities and were enrolled in Biology (30%), Economics (24%), Electronic Engineering (19%) and History (27%) programs (Parpala et al., in press). Exploratory factor analyses were conducted using principal axis factoring followed by an oblique, promax rotation. A six-factor solution was obtained in both samples, and this solution was similar to the six subscales created by McCune (2003). The percentage of response variance explained by this six-factor model was not reported, but authors did calculate factor reliabilities using the general reliabilities method (Tarkkonen & Vehkalahti, 2005). These reliabilities are displayed in Table 2.2 below.

Table 2.2

General reliabilities for six factor model of ETL-Q found in Parpala et al. (in press)

Factor	General reliability (British)
Teaching for understanding	0.76
Alignment	0.82
Staff enthusiasm and support	0.77
Interest and relevance	0.82
Constructive feedback	0.80
Support from other students	0.80

Finally, guided by previous analyses and using the results from the factor analysis of a sample of 1950 British undergraduate students, Entwistle (2009) shortened the ETLQ to create the Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETLQ). The sample of 1950 undergraduate students were enrolled in Biology (29%), Economics (23%), Electronic Engineering (19%), History (26%) and Media (3%) programs within the eleven universities, and consisted of first year students (81%) and final year students (19%) (Entwistle, 2005a; Hounsell, Entwistle, Anderson, et al., 2005). The SR-ETLQ was created by selecting those items that had the highest factor loadings on a single factor so as to maintain simple structure. The SR-ETLQ contained four sections: approaches to learning and studying; experience of teaching and learning (henceforth referred to as the SR-ETL-Q); demands made by the course; and, what was learned from the course (Entwistle, 2009). The resulting SR-ETL-Q contained 15 items that retained the six-factor structure of the original 40-item ETL-Q.

Evidence based on relationship with other variables. In conjunction with the ETLQ, a second questionnaire, the Learning and Studying Questionnaire (LSQ) was developed in the large scale UK project (Entwistle & McCune, 2004; Entwistle et al., 2002). A section of the LSQ was designed to "investigate students' broad aims and goals for higher education, using a set of items based on research into students' learning orientations" (McCune, 2003, p. 8). In order to focus some items on students' learning orientations within a particular course, shortened version of the LSQ was embedded into the ETLQ as the *Approaches to learning and studying* section (Entwistle et al., 2002). The approaches to learning and studying section consisted of four subscales: *deep approach, surface approach, monitoring studying*, and *organised studying* (McCune, 2003). By embedding this questionnaire into the ETLQ, the relationship between students' approaches to learning and their perceptions of teaching-learning environments, within a given course, could be investigated. The underlying assumption from theory was that the items (and corresponding subscales/factors) of the ETL-Q section would all be positively correlated with the deep approach, monitoring studying, and organised studying subscales of the LSQ, and negatively correlated with the surface approach subscale. If this were the case, this would suggest that the items of the ETL-Q did represent teaching-learning activities that encouraged a deep approach to learning. Correlational analyses conducted across a number of studies (e.g. Entwistle, 2005, 2008; Entwistle et al., 2002; McCune, 2003) provided evidence to support this assumption.

Two studies were also conducted to analyse the relationship between academic disciplines and students' perceptions of their teaching-learning environments within disciplines (Haarala-Muhonen, Ruohoniemi, Katajavuori, & Lindblom-Ylänne, 2011; Parpala, Lindblom-Ylänne, Komulainen, Litmanen, & Hirsto, 2010). These studies used two sections of the ETLQ (the LSQ and a modified 38-item version of the ETL-Q) to explore whether disciplinary variation influenced students' approaches to learning and their perceptions of their teaching-learning environments. Results of these studies provided support to suggest that unique aspects of a subject (discipline) and its corresponding pedagogy do influence both students' approaches to learning and their perceptions of their teaching-learning environments. As an example, it was found that at the University of Helsinki, students in Pharmacy and Veterinary Medicine perceived their environments more positively that those in Law (Haarala-Muhonen et al., 2011).

Evidence based on consequences. The ETL-Q was developed to monitor the quality of learning and teaching within a course. In particular, as the items are on a five-point Likert scale from disagree=1 to agree=5, the expectation is that if a course's teaching-learning environment is working effectively, all items (and therefore subscales) will be negatively distributed (i.e. most students' responses will cluster around agree). Items whose response distributions deviate from this expectation provide instructors or course teams with indicators of which aspects of the teaching-learning environment need to be reviewed (Entwistle, 2008; Entwistle et al., 2002).

Within the original UK-based ETL project, by design, data analysis results were used to facilitate discussions between researchers and course teams that led to collaborative initiatives that were implemented across subject areas, and subsequently, a second cycle of data-gathering was conducted to determine the impact of the initiatives (Hounsell et al., 2005). Consequences of using ETL-Q responses were reported for course units in Biology, Economics, Electronic Engineering and History programs and are described below.

Within Biology courses, initial results indicated that final year students tended to have higher mean scores across all subscales of the ETL-Q, suggesting that favourable perceptions were more likely to be found for final year students. Across the first year courses, where greater variability in ratings were noted, low scores (indicative of lower perceptions) were noted for items related to feedback and teaching for understanding/encouraging learning. Follow-up interviews conducted with students to analyse reasons for these low ratings lead to comments regarding the inadequacy of feedback and students dissatisfaction with delays in receiving feedback. Building on these findings, collaborative initiatives between researchers and course teams were implemented to attempt to resolve these concerns. Initiatives included providing greater clarity and guidance through the use of demonstrators or tutors. To determine the impact of initiatives, results from a second cycle of data-gathering were analysed. Although quantitative data collected during the second cycle using the ETL-Q did not provide statistically significant differences across all courses, results from two course units indicated improvement (higher scores) on items related to feedback. Follow-up interviews conducted in the second cycle provided further evidence, through students' positive comments, that collaborative initiatives implemented did indeed help improve students' learning (Hounsell et al., 2005; Hounsell, McCune, Litjens, & Hounsell, 2005).

Within Economics courses, initial results from the ETL-Q showed once again that final year students tended to perceive their overall teaching-learning environments more favourably than their first year counterparts, with the exception of peer support which first year students rated higher than final year students. It was noted however that in these courses, there were considerably lower rates of return for the ETL-Q thus results were taken with caution. Variability in ETL-Q scores was once again noted in first year courses with students providing low ratings for items related to encouraging learning and feedback and qualitative data from student interviews backed up the findings from the ETL-Q. In using both quantitative and qualitative data to come up with collaborative initiatives, it was noted that each Economics course had unique aspects, thus collaborative initiatives had to be context-specific rather than generalized. Collaborative initiatives implemented included reduction in pace of course delivery, focus on depth of learning and systematic revision of guidance and feedback (Land, Reimann, & Meyer, 2005; Reimann, 2004; Reimann, Land,

& Meyer, 2005). Consequences of adopting collaborative initiatives were not available for Economics courses due to departmental changes that arose during the period in which the ETL project was conducted (Reimann et al., 2005).

Within Electronic Engineering courses, initial results from the ETL-Q indicated that overall across both first year and final year courses, teaching-learning environments were perceived as effective. Thus the focus was on collaborative initiatives that could enhance those aspects of the teaching-learning environment that received slightly lower ratings: teaching for understanding/encouraging learning, clarity and coherence, feedback, and interest and enjoyment. Collaborative initiatives focused on analogue electronics units as these were found to be similar across universities. Qualitative interviews indicated that students concerns revolved around tutorial problems and the lack of work examples or feedback in relation to them; thus the main initiative introduced by course teams was tutorial workbooks that students used for problem solving. Within these workbooks students could comment on difficulties they had with a problem and tutors could then use this information to identify areas that needed further review (Entwistle, 2005b; Entwistle, Nisbet, & Bromage, 2005).

Finally, within History courses, initial results from the ETL-Q indicated that across both first year and final year courses students provided high ratings for all aspects of their teaching-learning environments. This trend was however not static, as scores in one year of collection were higher than those in a second year of collection, particularly in regard to feedback and staff support. Qualitative data from students supported these findings as students indicated concerns with the time taken for feedback to be delivered. To address these concerns, collaborative initiatives included changes to organization of course content, providing detailed guidance for assignments and increasing support and guidance from tutors. Results from post-collaboration data-gathering suggested that these initiatives were successful as slight increases were apparent in most ETL-Q subscales across all courses (Anderson & Day, 2005; Hounsell & Anderson, 2005).

Overall, across all subject areas the ETL-Q responses were successfully used to identify aspects of the various teaching-learning environments that required attention. Furthermore, results of using the ETL-Q, after various collaborative initiatives were implemented, also provided further evidence of the usefulness of this questionnaire across a range of undergraduate subject areas.

Evidence based on generalizability. Generalizability is concerned with the degree to which response interpretations are similar across different population groups (population generalizability), different situations or settings (ecological generalizability), different times (temporal generalizability) and different tasks (task generalizability) (Messick, 1989). Furthermore, when determining generalizability, Messick (1989) cautioned:

> It should be noted that generalizability of score meaning explicitly does *not* require that all the statistical relationships that a score displays with other variables in one group or context need be replicated in other groups or contexts. Indeed scores may interact with different variables in different contexts or with the same variables in different ways. (p. 56)

Guidelines for cross-cultural adaptation of questionnaires stress the importance of making a questionnaire culturally relevant to a new context by going through the processes of translation and cultural adaptation so that score interpretations can have meaning (Beaton et al., 2000; Berry, 1969; Johnson, 1998). Beaton et al. (2000) suggest that when a questionnaire is to be used in another country and another language, both changes in culture and language must occur, thus requiring the processes of translation and cultural adaptation to be implemented for the questionnaire to be useful. Furthermore, when using a questionnaire in a different country that uses the same language as the original context in which a questionnaire was developed, there is still a need to take into account cultural differences, thus cultural adaptation is still necessary (Beaton et al., 2000).

Following similar guidelines, evidence of population generalizability, with regard to the internal structure of ETL-Q responses, was collected through correlational analyses across a number of countries. The ETL-Q was translated into several languages and using factor analysis (both exploratory and confirmatory), evidence of the generalizability of the internal structure of ETL-Q responses was collected in Finland (Parpala et al., in press; Rytkönen, Parpala, Lindblom-Ylänne, Virtanen, & Postareff, 2011), China (Xu, 2004), and Belgium (Stes, De Maeyer, Gijbels, & Van Petegem, 2011).

In the Finnish and Chinese studies (Parpala et al., in press; Rytkönen et al., 2011; Xu, 2004), the internal structure of ETL-Q responses were found to be six-factor and five-factor structure respectively, similar to findings in the original

British context. The Finnish sample consisted of 2509 undergraduate students who came from 11 faculties at the University of Helskini; of those who reported their year of study, 54% were first-year students and 44% were third-year students while the Chinese sample consisted of 552 Economics undergraduate students from first (26%), second (19%), third (27%) and fourth year (28%) courses.

The greatest departure from this structure was evidenced in the study by Stes et al. (2011) where confirmatory factor analysis was conducted on ETL-Q data from 714 undergraduate students. Although they noted that multivariate normality tests for skewness and kurtosis clearly indicated skewed data, Stes et al. (2011) chose to use maximum likelihood estimation to test the theorized six-factor structure of the ETL-Q. In their study, they found that an eight-factor model provided the best fit for their data (Stes et al., 2011).

Overall, the various population generalizability studies, with the exception of Stes et al. (2011), provide support to suggest that even after adopting various procedures for cross-cultural adaptation, the internal structure of ETL-Q responses appears generalizable to a number of different undergraduate populations. Table 2.3, below, provides a summary of how findings related to the internal structure of ETL-Q (and SR-ETL-Q) responses in the aforementioned studies relate to the features of the teaching-learning environment (TLE) shown in Figure 1.

Table 2.3

Relationship between features of the TLE subscales of the ETL-Q from various population generalizability studies

Features of TLE (Figure 1)	**Entwistle, et al., 2002**	**McCune, 2003**	**Xu, 2004**	**Parpala, et al., in press**	**Entwistle, 2009**
Course structure, organization and management	Organisation and structure	Clear aims, organisation, alignment and integration	Clarity & Choice	Alignment	Congruence and coherence
Exemplify ways of thinking	Encouraging learning	Teaching for understanding and choice	Understanding, Challenge & Support	Teaching for understanding	Teaching for understanding
Arouse interest, explain terms	Evoking interest	Interest, enjoyment and relevance	Engagement	Interest and relevance	Interest and enjoyment
Be alert, show empathy	Supportive climate	Staff enthusiasm and support	Supportiveness	Staff enthusiasm and support	Staff enthusiasm and support
Support for individual learning and studying		Support from other students		Support from other students	Support from other students
Allocation of set work and feedback on it	Assessments and feedback	Assessment for understanding, guidance and feedback	Assessment focus	Constructive feedback	Constructive feedback

In summary, the ETL-Q and presumably the SR-ETL-Q are theory-based student evaluation questionnaires, based on student learning research, designed to measure the quality of teaching and learning (academic quality) in higher education. The empirical evidence collected provides support to suggest that the ETL-Q has accumulated validity evidence in favour of its use for improving the quality of teaching and learning within some undergraduate populations. Furthermore, previous population generalizability studies that have focused on the internal structure of ETL-Q responses, have been found the internal structure of responses to consistently result in a five- or six-factor structure across a range of undergraduate student populations in various countries. Overall, integrating these findings, from the available sources of validity evidence, strengthens the validity argument to support the interpretations of ETL-Q responses for their intended use: improving quality of teaching and learning in undergraduate education courses.

SIGNIFICANCE OF CURRENT STUDY

Messick (1989) stressed the importance of collecting generalizability evidence for an instrument across different groups, settings, times, and tasks. Generalizability studies are important in the process of building the validity argument for the ETL-Q as they require the collection of all other five sources of validity evidence within new contexts (group, setting, time or task) to which the questionnaire is to be applied. Though the ETL-Q has been shown to be a useful tool for improving the quality of teaching and learning across various undergraduate contexts, there remains a clear gap in the literature due to the lack of studies focusing on the collection of validity evidence for ETL-Q responses within North American undergraduate populations. This study is therefore opportune, as it begins to address this gap by collecting validity evidence based on the internal structure of ETL-Q responses within a new context: a Canadian undergraduate population.

CHAPTER THREE

METHODS

INTRODUCTION

This chapter begins by describing the main goal of the study. The goal is followed by a description of the participants and questionnaire that was selected for use in the study. The chapter concludes with a discussion of the data collection and analysis procedures that were used to achieve the goal of the study.

GOAL OF THE STUDY

The main goal of this study was to examine the internal structure of responses, from either the Experiences of Teaching and Learning Questionnaire (ETL-Q) or the Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q), within a Canadian undergraduate population. In particular, this study adopted a non-experimental quantitative research design to collect validity evidence toward establishing the population generalizability of responses within a new cultural context. Using confirmatory factor analyses, this study investigated whether the internal structure of responses from a Canadian sample would fit the theoretical six-factor model better than a plausible five-factor model alternative.

PARTICIPANTS

The dataset used in the current study was part of a larger project undertaken to inform ongoing course development in a large third-year multi-section course at a Western Canadian research-intensive university. Seven hundred Faculty of Education undergraduate students were enrolled in this multi-section course during the period in which the current study was conducted. The project was approved by the university's research ethics board and members of the research team complied with the university's standards for the protection of human research participants. Participation in the project was voluntary.

INSTRUMENT

The Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q) (Entwistle, 2009) was selected for use over the Experiences of Teaching and Learning Questionnaire (ETL-Q) due to the relatively small population accessible during the study period. The SR-ETL-Q contains 15 items and uses a 5-point Likert scale (1 = disagree to 5 = agree) to measure students' perceptions of various features of a course's teaching-learning environment. The 15 items are arranged into six sub-scales: Congruence and coherence (3 items), Teaching for understanding (3 items), Staff enthusiasm and support (3 items), Constructive feedback (2 items), Support from other students (2 items) and Interest and enjoyment (2 items).

These 15 SR-ETL-Q items were embedded in an online post-survey under the heading "What have been your overall experiences of the teaching and learning environment in this course?" Three minor modifications were made to these 15 items to ensure that the SR-ETL-Q was culturally relevant (Appendix A). The modifications were made due to the recognition that though the original SR-ETL-Q items used the term staff to refer to faculty, in the Canadian context this term could refer to both administration and faculty, thus instructor(s) was substituted for this term. Similarly, whereas in the original SR-ETL-Q the term course unit was used to refer to an individual course that made up a degree programme, using simply course was more suitable for the Canadian context. Finally the term set work was changed to assignments. The participants were asked to respond to the 15 items using the following key: (1) = disagree, (2) = disagree somewhat, (3) = neither disagree/agree, (4) = agree somewhat, and (5) = agree.

DATA COLLECTION

Data was collected during the fall 2011 semester. Students enrolled in the multi-section course were invited to complete an online survey, in which the SR-ETL-Q was embedded. The online survey was made available to participants on October 26, 2011, toward the end of their course. Participants were sent an email with a link to the survey and the link was also made available on their course's website. A research assistant also went in to classes during the final week to remind the students that the post-survey was available and to request students to participate in the study. The survey was then closed on November 6, 2011, two days after the course ended. The final dataset used in this study contained a subset of survey data from 202 undergraduate students who consented to have their answers used for research purposes.

DATA ANALYSIS

DATA SCREENING AND PREPARATION

The 15 items that made up the SR-ETL-Q were extracted and initial screening of data was conducted to determine the number of missing values by item and participant. The Missing Value Analysis (MVA) module of IBM SPSS Statistics Version 20 was used to conduct this analysis (IBM Corporation, 2011). Cases found to have over fifty percent of their data missing were removed from the analysis. Three cases were found to have over fifty percent of their data missing and were removed. This resulted in final set of one hundred and ninety nine cases. The remaining cases had between one to six missing values (i.e. between 6.7% and 40.0% of their data missing).

Several methods were available for imputing missing values in the remaining one hundred and ninety nine cases. Modern approaches which have come to be favoured include expectation-maximisation (EM) and multiple imputation (MI) (Schafer & Graham, 2002). The EM approach uses a two-step process that assumes a normal distribution for the partially missing data and estimates missing data based on the likelihood under that distribution (Little & Rubin, 2002). The MI approach is a flexible alternative to the EM approach, that uses Monte Carlo methods to handle a wider range of missing data types (e.g. ordinal data or continuous and ordinal data together) (Schafer & Graham, 2002). In the MI approach, missing values are replaced by a number of simulated versions though it is possible, too, to use the MI approach to run a single imputation by setting number of simulations to one (Barzi & Woodward, 2004;

Schafer, 1999). The EM and MI approaches were available within the MVA module of IBM SPSS Statistics Version 20.0. A study conducted to analyse the robustness of these two approaches to different sample sizes indicated that for small to moderate sample sizes, the MI approach worked better than EM (Barzi & Woodward, 2004). Thus given the sample size in this study, coupled with the ordinal nature of the data, the MI approach (using logistic regression and setting number of simulations to one) was used to impute missing values so as to have a complete dataset needed to perform confirmatory factor analysis.

DESCRIPTIVE STATISTICS

Item-level statistics. Given that the final dataset used in this study contained only 199 participants, it was necessary to examine mean, standard deviation, skewness and kurtosis for each of the 15 items. Levels of skewness and kurtosis were analysed so as to determine whether maximum likelihood (ML) estimation, which is commonly used in confirmatory factor analysis, could be used with these data. Skewness values were found to fall in the range -1.471 to 0.077, while kurtosis values ranged from -1.138 to 3.288. According to previous research, when univariate values of skewness are less than |2.0| and kurtosis less than |7.0|, as was the case within this dataset, ML estimation could be used (Curran, West, & Finch, 1996).

Inter-item correlations. Internal consistency refers to “the interrelatedness of a set of items” (Schmitt, 1996, p. 350). To analyse the internal consistency of SR-ETL-Q responses, and due to the ordinal nature of items, two inter-item correlation matrices, using Pearson and polychoric correlations, were generated with the SPSS R-Menu for Ordinal Factor Analysis (Basto & Pereira, 2012). Internal consistency results are central to correlational studies as they are the basis for methods such as factor analysis (Russell, 2002) and calculations of coefficient alpha (Cronbach, 1951). Traditionally correlational studies use Pearson correlations, which assume that data are continuously distributed, however to ensure accuracy of results when data are ordinal, polychoric correlations are preferred (Choi, Peters, & Mueller, 2010).

Following guidelines available for inter-item correlation matrices generated using Pearson correlations, the Pearson correlation matrix was analysed to identify correlations that were potentially too high (i.e. in excess of 0.70) as this could indicate item redundancy (John & Soto, 2007). Item redundancy is a concern as it leads to artificially high values for alpha (Streiner, 2003). Similarly, when conducting factor analyses, though all items in a multi-dimensional scale are expected to be correlated to some extent, it is necessary to compare correlations of items within a subscale (within-subscale inter-item correlations) versus correlations of items outside a subscale (between-subscale inter-item correlations) because between-subscale correlations that are larger than within-subscale correlations may be an indicator that an item does not clearly differentiate between two (or more) subscales (John & Soto, 2007).

Subscale statistics. As alpha is a function of internal consistency rather than a measure of internal consistency (though it is often incorrectly reported as one) (John & Soto, 2007; Sijtsma, 2009); when making comparisons between results of this study and the previous study by Entwistle (2009) that used SR-

ETL-Q, the mean inter-correlations for items within subscales was calculated and reported alongside alpha. Finally, using polychoric correlations, ordinal alpha, the ordinal equivalent of coefficient alpha, was also calculated (Zumbo, Gadermann, & Zeisser, 2007).

CONFIRMATORY FACTOR ANALYSIS

While a six-factor model was theorized from the literature for the SR-ETL-Q, a plausible alternative was the five-factor model formed by collapsing the two subscales related to support: instructor enthusiasm and support (3 items), and support from other students (2 items). Consequently, IBM SPSS Amos 20.0.0 was used to conduct confirmatory factor analyses with the aim of determining if either model provided a good fit to the given data (Arbuckle, 2011).

Though ordinal data is often treated as though it were continuous, so as to be able to use maximum likelihood estimation (ML), weighted least squares (WLS) estimation and unweighted least squares (ULS) are two alternative estimation methods that have been found to be more accurate when data are ordinal (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Olsson, Foss, Troye, & Howell, 2000).Within AMOS, all three estimation methods (WLS, ULS and ML) were available, though WLS estimation was labelled asymptotically distribution-free (ADF) (Arbuckle, 2011). WLS was not used for this study because it has been found to only perform adequately when large sample sizes (preferably over 1000) are available (Olsson et al., 2000). A Monte Carlo study that compared ULS and ML approaches found that when data were continuously distributed and weak factor loadings were present, though ULS generally outperformed ML, it was best to compute both ML and ULS solutions and compare them (Ximénez, 2006). Following this suggestion, both ML and ULS solutions were computed for each model to be tested.

Prior to running confirmatory factor analyses, the five-factor and six-factor models were hypothesized to fit the data and items were specified to load on specific latent factors. Furthermore, factors were also allowed to correlate with one another, but measurement errors were not allowed to correlate. Finally the error regression weights and variances of latent factors were set to equal 1.0. During confirmatory analysis, multiple fit indices, including Comparability Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), Non-Normed Fit Index (NNFI) (also known as the Tucker-Lewis Index – TLI) and Goodness of Fit Index (GFI) were examined to determine the best model.

When data is normally distributed, traditional cut-off values used to determine acceptable model fit based on ML method are as follows: 0.95 for CFI and TLI; 0.90 for GFI; 0.08 for SRMR; and, 0.05 for RMSEA (Marsh, Hau, & Wen, 2004). On the other hand, another study suggests that cut-off criteria for CFI, TLI and GFI should be set to 0.95 and RMSEA between 0.06 and 0.08 (Schreiber, Nora, Stage, Barlow, & King, 2006).

In the presence of increasing nonnormality and skewness, as is generally the case with ordinal data, all fit indices have been found to be negatively biased thus leading to indications of poorer model fit (Hutchinson & Olmos, 1998). For example, though often reported in confirmatory analysis results, GFI is negatively influenced by skewness with greater levels of skewness in the data leading to lower values of GFI. Furthermore, the fit indices least affected by nonnormality were found to be CFI, and NNFI (TLI), while the fit index least affected by sample size was found to be RMSEA (Hutchinson & Olmos, 1998). Finally, in regard to sample size considerations, Hu and Bentler (1999) suggest that for samples with fewer than 250 participants, a combination of CFI and SRMR should be reported with the following cut-off values, 0.95 for CFI and 0.09 for SRMR, used to determine model fit.

Integrating this knowledge, and taking into consideration that a single index cannot be used alone to compare models to ascertain which is model fits data better, all five fit indices will be compared heuristically (with traditional cut-off values only serving as guides) to determine which model best fits the data.

CHAPTER FOUR

RESULTS

INTRODUCTION

The main goal of this study was to examine the internal structure of the Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q) responses within a Canadian undergraduate population. This goal was achieved by comparing the fit of a theoretical six-factor model against a plausible five-factor model alternative for a Canadian sample. The results of data analyses are presented in this chapter and organised into two sections: descriptive statistics and testing the research hypothesis. Descriptive statistics include item-level statistics, inter-correlation matrices and subscale statistics. These statistics are followed by the results of confirmatory factor analyses of the two models using the two estimation procedures: maximum likelihood estimation and unweighted least squares estimation. The chapter ends with a comparison of model fit indices from the two models.

DESCRIPTIVE STATISTICS

Item-level statistics. Item-level univariate statistics (Table 4.1) were collected to determine whether data were approximately normally distributed. Item skewness values were found to range from -1.359 to 0.073 while kurtosis values ranged from -1.116 to 3.409. These findings fell within the guidelines set by Curran, West and Finch (1996) to indicate that maximum likelihood (ML) estimation could be used during confirmatory factor analyses.

Table 4.1

Univariate statistics for the 15 items from the SR-ETL-Q

Item	Mean	SD	Skewness	Kurtosis
Q01. It was clear to me what I was supposed to learn in this course.	4.362	0.791	-1.359	2.058
Q02. The topics seemed to follow each other in a way that made sense to me.	4.211	0.769	-1.323	3.409
Q03. What we were taught seemed to match what we were supposed to learn.	4.111	0.857	-0.943	0.969
Q04. This course has given me a sense of the thinking that goes on 'behind the scenes' in this subject area.	4.065	0.888	-0.959	1.034
Q05. The teaching in this course helped me to think about the evidence underpinning different views.	3.724	1.029	-0.575	-0.246
Q06. This course encouraged me to relate what I learned to issues in the wider world.	3.638	1.064	-0.532	-0.519
Q07. Instructor(s) tried to share their enthusiasm about the subject with us.	4.176	0.956	-1.166	0.958
Q08. Instructor(s) were patient in explaining things which seemed difficult to grasp.	3.658	1.195	-0.658	-0.513
Q09. Students' views were valued in this course.	3.794	1.152	-0.932	0.252
Q10. The feedback given on my assignments helped me to improve my ways of learning and studying.	2.608	1.262	0.073	-1.126
Q11. The feedback given on my assignments helped to clarify things I hadn't fully understood.	2.608	1.209	0.082	-0.978
Q12. Students supported each other and tried to give help when it was needed.	3.583	1.116	-0.552	-0.400
Q13. Talking with other students helped me to develop my understanding.	3.673	1.123	-0.581	-0.474
Q14. I found most of what I learned in this course really interesting.	3.472	1.193	-0.475	-0.643
Q15. I enjoyed being involved in this course.	3.427	1.241	-0.393	-0.830

Note. The items are labelled as Q01 to Q15 and are referred to by these labels throughout the study.

Inter-item correlations. Tables 4.2 and 4.3 present inter-item correlation matrices using Pearson correlations and polychoric correlations respectively. The matrices were generated to examine patterns of correlations between items within subscales (denoted by correlations within a border) versus items between subscales. These matrices were used to identify possible cases of item redundancy within scales and to determine whether any between-subscale correlations were larger than within-subscale correlations.

Table 4.2

Pearson inter-item correlation matrix

	Q01	Q02	Q03	Q04	Q05	Q06	Q07	Q08	Q09	Q10	Q11	Q12	Q13	Q14	Q15
Q01	1.000	.588	.566	.462	.353	.312	.470	.356	.443	.158	.123	.172	.122	.412	.459
Q02	.588	1.000	.577	.519	.406	.341	.396	.348	.426	.221	.220	.186	.057	.441	.524
Q03	.566	.577	1.000	.541	.487	.426	.488	.545	.545	.227	.232	.291	.174	.522	.634
Q04	.462	.519	.541	1.000	.644	.485	.534	.416	.458	.347	.259	.252	.118	.572	.557
Q05	.353	.406	.487	.644	1.000	.586	.460	.395	.386	.344	.270	.242	.249	.498	.532
Q06	.312	.341	.426	.485	.586	1.000	.356	.407	.347	.259	.266	.311	.260	.538	.535
Q07	.470	.396	.488	.534	.460	.356	1.000	.606	.593	.229	.204	.320	.223	.587	.550
Q08	.356	.348	.545	.416	.395	.407	.606	1.000	.635	.373	.333	.423	.289	.528	.658
Q09	.443	.426	.545	.458	.386	.347	.593	.635	1.000	.389	.279	.302	.198	.505	.592
Q10	.158	.221	.227	.347	.344	.259	.229	.373	.389	1.000	***.746***	.256	.187	.369	.333
Q11	.123	.220	.232	.259	.270	.266	.204	.333	.279	***.746***	1.000	.208	.270	.308	.304
Q12	.172	.186	.291	.252	.242	.311	.320	.423	.302	.256	.208	1.000	.512	.445	.381
Q13	.122	.057	.174	.118	.249	.260	.223	.289	.198	.187	.270	.512	1.000	.387	.344
Q14	.412	.441	.522	.572	.498	.538	.587	.528	.505	.369	.308	.445	.387	1.000	***.761***
Q15	.459	.524	**.634**	.557	.532	.535	.550	**.658**	.592	.333	.304	.381	.344	***.761***	1.000

Note. Correlations that are higher than expected are indicated in bold.

Table 4.3

Polychoric inter-item correlation matrix

	Q01	Q02	Q03	Q04	Q05	Q06	Q07	Q08	Q09	Q10	Q11	Q12	Q13	Q14	Q15
Q01	1.000	.736	.714	.537	.428	.426	.564	.436	.545	.196	.146	.232	.190	.493	.560
Q02	.736	1.000	.706	.646	.516	.433	.511	.419	.523	.266	.255	.227	.100	.522	.618
Q03	.714	.706	1.000	.619	.562	.531	.575	.624	.616	.269	.284	.366	.245	.597	.716
Q04	.537	.646	.619	1.000	.739	.577	.640	.482	.541	.410	.309	.276	.159	.652	.640
Q05	.428	.516	.562	.739	1.000	.658	.541	.447	.462	.387	.298	.302	.295	.552	.589
Q06	.426	.433	.531	.577	.658	1.000	.451	.463	.442	.297	.310	.370	.302	.596	.603
Q07	.564	.511	.575	.640	.541	.451	1.000	.699	.675	.279	.246	.373	.299	.685	.647
Q08	.436	.419	.624	.482	.447	.463	.699	1.000	.697	.428	.370	.475	.347	.589	.729
Q09	.545	.523	.616	.541	.462	.442	.675	.697	1.000	.456	.317	.371	.243	.576	.670
Q10	.196	.266	.269	.410	.387	.297	.279	.428	.456	1.000	.796	.292	.221	.433	.378
Q11	.146	.255	.284	.309	.298	.310	.246	.370	.317	.796	1.000	.238	.313	.357	.336
Q12	.232	.227	.366	.276	.302	.370	.373	.475	.371	.292	.238	1.000	.574	.491	.436
Q13	.190	.100	.245	.159	.295	.302	.299	.347	.243	.221	.313	.574	1.000	.437	.402
Q14	.493	.522	.597	.652	.552	.596	.685	.589	.576	.433	.357	.491	.437	1.000	.814
Q15	.560	.618	.716	.640	.589	.603	.647	.729	.670	.378	.336	.436	.402	.814	1.000

Subscale statistics. Table 4.4 reports the subscale statistics for the SR-ETL-Q responses within the current study. To allow a comparison between results from the current study and those from a previous study by Entwistle (2009), coefficient alpha was recorded. Additionally, because the SR-ETL-Q items are ordinal in nature, ordinal alpha (Zumbo et al., 2007) was also reported. Finally, to investigate how alpha values are affected by mean inter-item correlations, mean inter-item correlations were reported for both Pearson and polychoric correlations.

Table 4.4

SR-ETL-Q subscale statistics

Subscale (Based on Figure 1)	No. of items	$\bar{r}_{ij(Pear)}$	coeff. alpha	$\bar{r}_{ij(poly)}$	ordnl. alpha
(Congruence) Course structure, organization and management	3	0.577	0.802	0.719	0.885
(Teaching) Exemplify ways of thinking	3	0.572	0.797	0.658	0.852
(InstructorSupp) Be alert, show empathy	3	0.611	0.821	0.690	0.870
(Feedback) Allocation of set work and feedback on it	2	0.746	0.854	0.796	0.886
(StudentSupp) Support for individual learning and studying	2	0.512	0.677	0.574	0.729
(Interest) Arouse interest, explain terms	2	0.761	0.864	0.814	0.897

Note. $\bar{r}_{ij}$=mean inter-item correlation and under the heading subscale, terms in brackets represent the matching latent factor labels used in the theorized six-factor model.

Table 4.5 compares results from this study with those from the original study by Entwistle (2009). The Entwistle (2009) study only reported coefficient alpha values for the subscales; it was therefore necessary to compute mean inter-item correlations for comparison purposes. Because the Entwistle (2009) study used maximum likelihood estimation, comparisons were made using Pearson correlation-based statistics.

Table 4.5

Comparison of SR-ETL-Q subscale statistics for current study versus Entwistle (2009) study

		Current study (N = 199)		Entwistle, 2009 (N = 1950)	
Subscale (Based on Figure 1)	No. of items	$\bar{r}_{ij}$	Coeff. alpha	$\bar{r}_{ij}$	Coeff. alpha
(Congruence) Course structure, organization and management	3	.577	.802	.462	.724
(Teaching) Exemplify ways of thinking	3	.572	.797	.404	.668
(InstructorSupp) Be alert, show empathy	3	.611	.821	.449	.712
(Feedback) Allocation of set work and feedback on it	2	.746	.854	.626	.771
(StudentSupp) Support for individual learning and studying	2	.512	.677	.575	.734
(Interest) Arouse interest, explain terms	2	.761	.864	.709	.834

Note. $\bar{r}_{ij}$= Pearson mean inter-item correlations and under the heading subscale, terms in brackets represent the matching latent factor labels used in the theorized six-factor model.

TESTING THE RESEARCH HYPOTHESIS

FIVE-FACTOR MODEL

For the five-factor model, Table 4.5 provides a comparison of standardized and unstandardized parameter estimates and standard errors for maximum likelihood (ML) and unweighted least squares (ULS) solutions from confirmatory factor analyses (CFAs).

Table 4.6

Five-factor model standardized (β) and unstandardized (B) parameter estimates

		ML parameters			ULS parameters	
Item	Latent Factor	β	B	S.E.	β	B
Q01	Congruence	0.704	0.556	0.052	0.672	0.531
Q02	Congruence	0.724	0.555	0.05	0.711	0.545
Q03	Congruence	0.83	0.71	0.053	0.878	0.751
Q04	Teaching	0.801	0.709	0.056	0.801	0.709
Q05	Teaching	0.792	0.813	0.065	0.762	0.782
Q06	Teaching	0.682	0.724	0.071	0.715	0.758
Q07	Support	0.739	0.705	0.061	0.735	0.701
Q08	Support	0.813	0.969	0.073	0.794	0.946
Q09	Support	0.767	0.881	0.072	0.747	0.858
Q10	Feedback	0.954	1.201	0.095	0.925	1.165
Q11	Feedback	0.782	0.943	0.089	0.807	0.973
Q12	Support	0.494	0.55	0.079	0.521	0.58
Q13	Support	0.376	0.421	0.081	0.419	0.469
Q14	Interest	0.837	0.996	0.071	0.848	1.009
Q15	Interest	0.909	1.124	0.071	0.897	1.11

Note. S.E. = standard error

Maximum likelihood solution. Figure 4.1 provides the maximum likelihood (ML) solution for the five-factor model. Standardized parameter estimates, latent factor correlations and squared multiple correlation (SMC) values for observed variables are displayed in Figure 4.1. The ML-based fit indices for this model were found to be: Standardized Root Mean Square Residual (SRMR) = 0.057, Root Mean Square Error of Approximation (RMSEA) = 0.085, Comparability Fit Index (CFI) = 0.925, Tucker-Lewis Index (TLI) = 0.902 and Goodness of Fit Index (GFI) = 0.882.

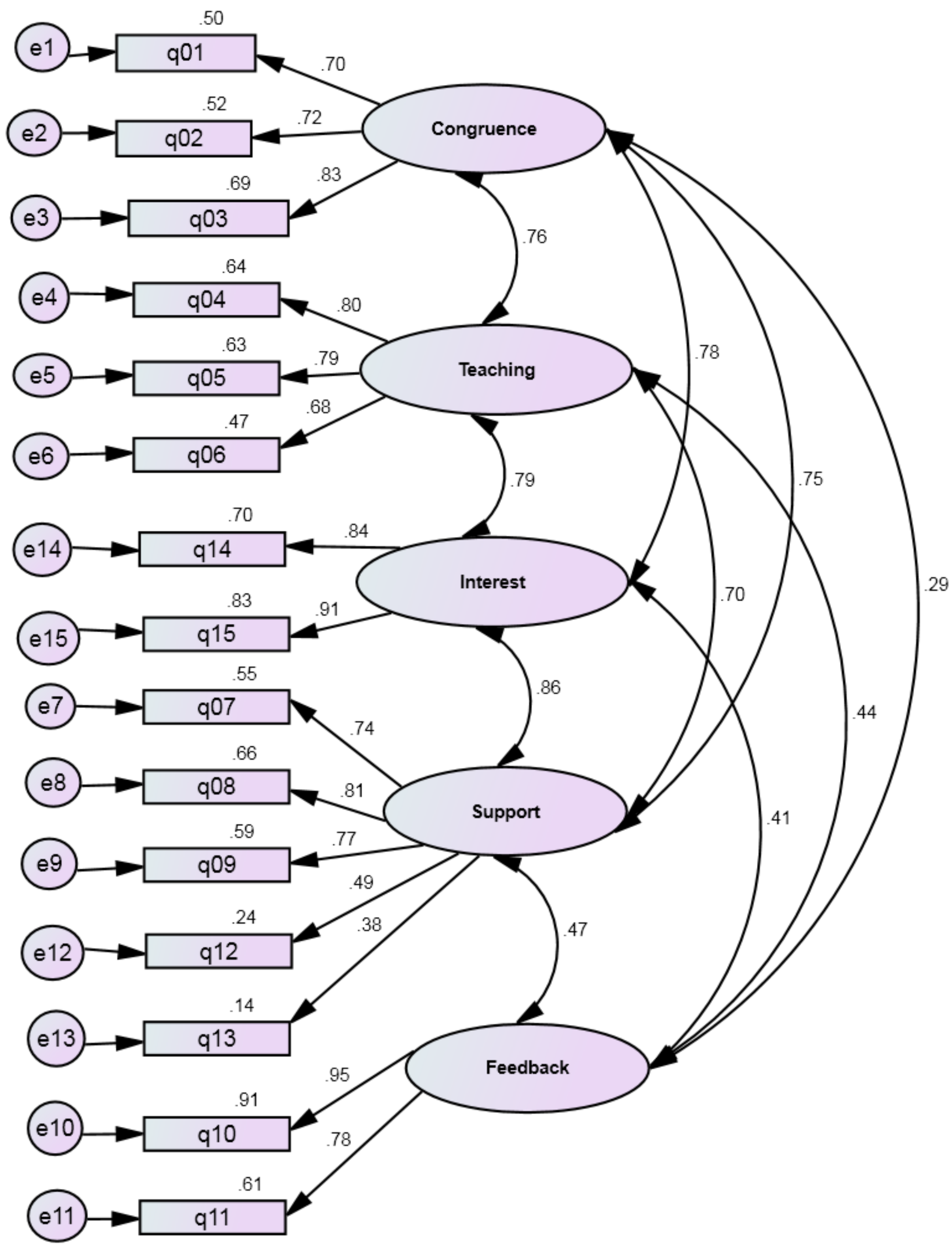

Figure 4.1. ML solution for five-factor CFA model

Unweighted least squares solution. Figure 4.2 provides the unweighted least squares (ULS) solution for the five-factor model. Standardized parameter estimates, latent factor correlations and squared multiple correlation (SMC) values for observed variables are displayed in Figure 4.2. Of the five fit indices of interest, only two of these indices were reported by AMOS for the ULS solution: Standardized Root Mean Square Residual (SRMR) = 0.056 and Goodness of Fit Index (GFI) = 0.987.

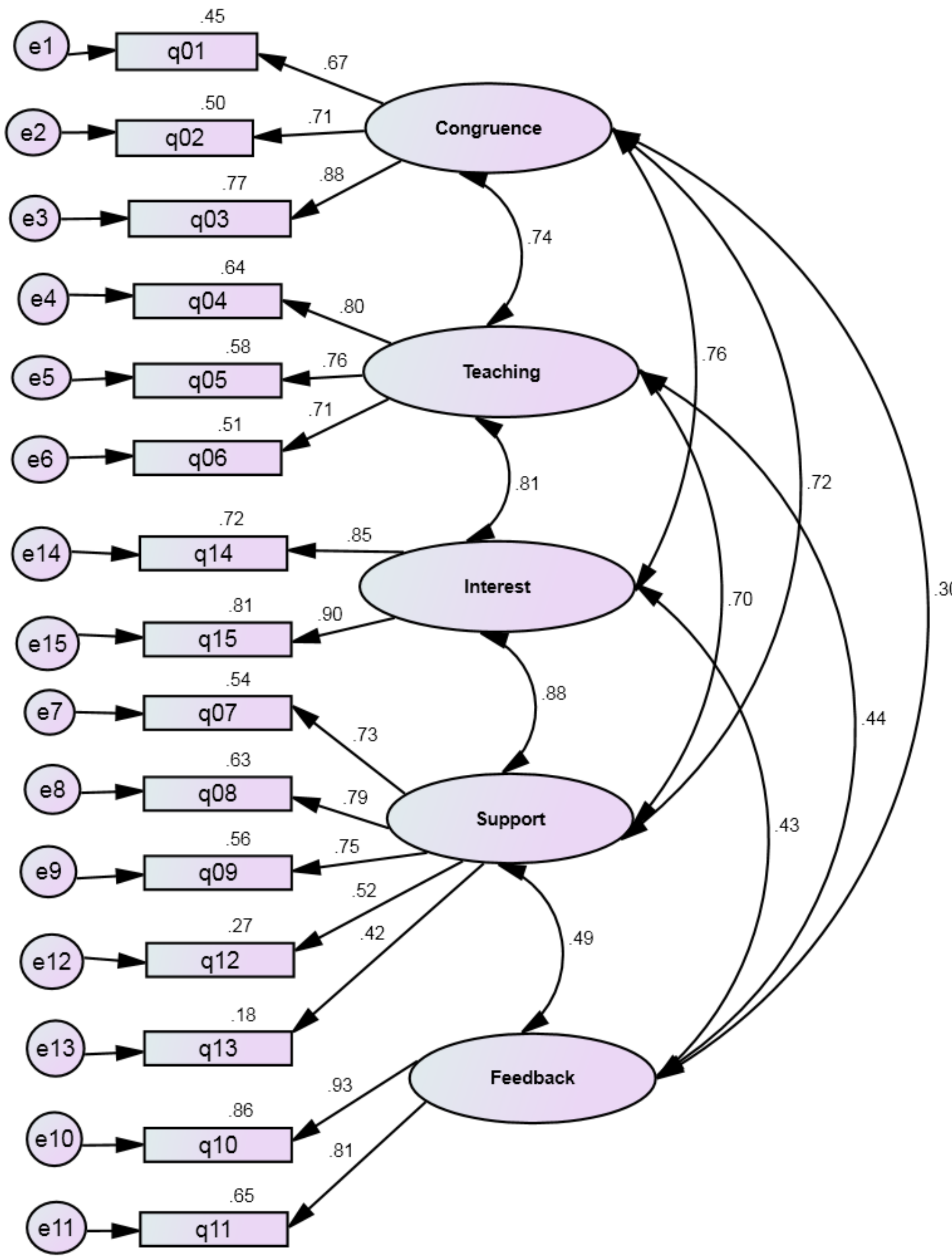

Figure 4.2. ULS solution for five-factor CFA model

SIX-FACTOR MODEL

For the six-factor model, Table 4.6 provides a comparison of standardized and unstandardized parameter estimates and standard errors ML and ULS solutions.

Table 4.7

Six-factor model standardized (β) and unstandardized (B) parameter estimates

		ML parameters			ULS parameters	
Item	Latent Factor	β	B	S.E.	β	B
Q01	Congruence	0.707	0.558	0.052	0.673	0.531
Q02	Congruence	0.728	0.558	0.05	0.712	0.546
Q03	Congruence	0.826	0.706	0.053	0.877	0.75
Q04	Teaching	0.802	0.711	0.056	0.801	0.71
Q05	Teaching	0.79	0.812	0.065	0.762	0.782
Q06	Teaching	0.681	0.723	0.071	0.714	0.758
Q07	InstructorSupp	0.752	0.717	0.061	0.757	0.722
Q08	InstructorSupp	0.81	0.965	0.074	0.817	0.974
Q09	InstructorSupp	0.784	0.901	0.072	0.773	0.887
Q10	Feedback	0.948	1.194	0.094	0.924	1.163
Q11	Feedback	0.787	0.949	0.089	0.808	0.974
Q12	StudentSupp	0.766	0.852	0.091	0.8	0.891
Q13	StudentSupp	0.668	0.748	0.089	0.639	0.716
Q14	Interest	0.846	1.006	0.07	0.849	1.01
Q15	Interest	0.899	1.113	0.071	0.896	1.109

Maximum likelihood solution. Figure 4.3 provides the ML solution for the six-factor model. Standardized parameter estimates, latent factor correlations and squared multiple correlation (SMC) values for observed variables are displayed in Figure 4.3. The ML-based fit indices for this model were found to be: Standardized Root Mean Square Residual (SRMR) = 0.043, Root Mean Square Error of Approximation (RMSEA) = 0.070, Comparability Fit Index (CFI) =

0.953, Tucker-Lewis Index (TLI) = 0.934 and Goodness of Fit Index (GFI) = 0.913.

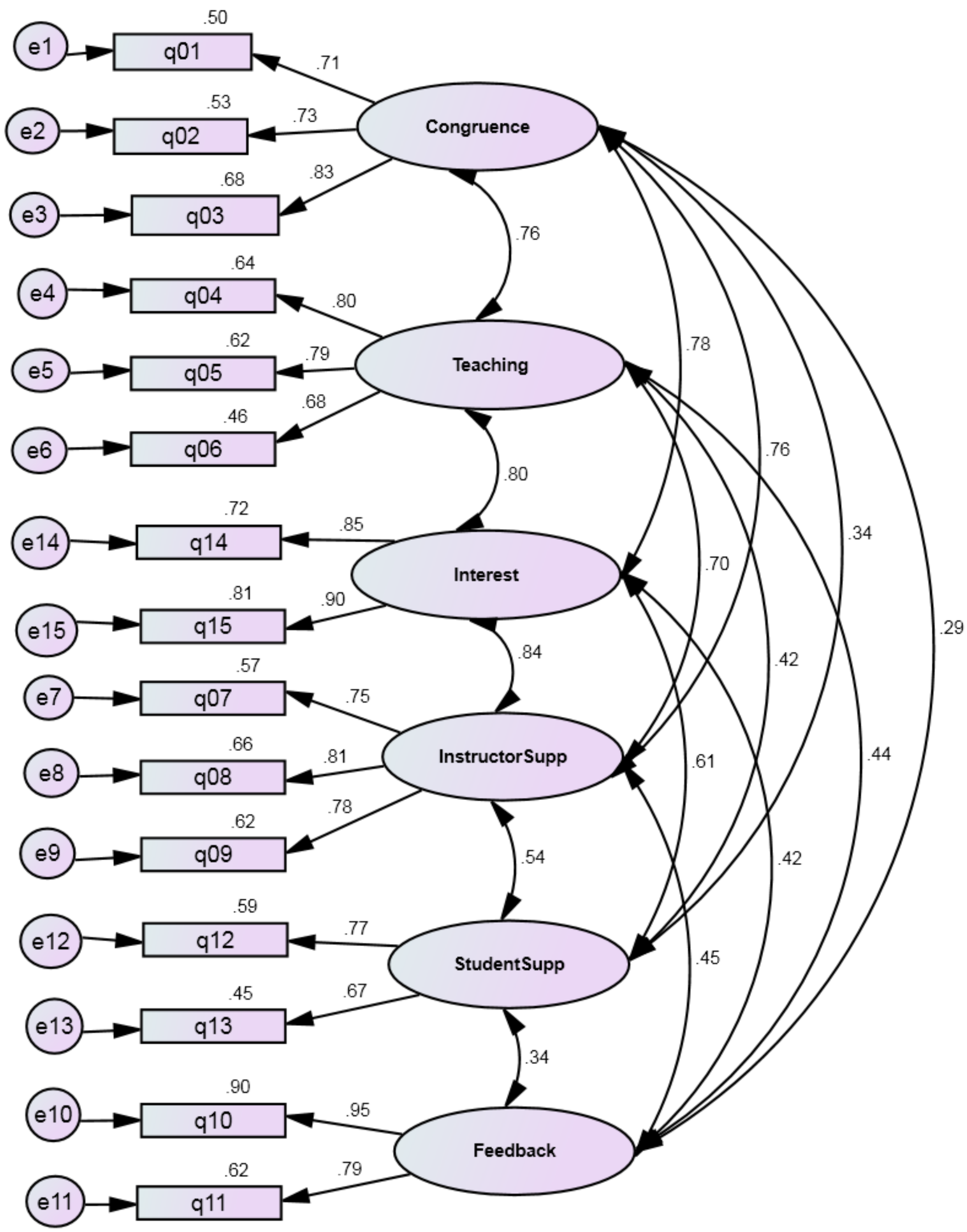

Figure 4.3. ML solution for six-factor CFA model

Unweighted least squares solution. Figure 4.4 provides the unweighted least squares (ULS) solution for the six-factor model. Standardized parameter estimates, latent factor correlations and squared multiple correlation (SMC) values for observed variables are displayed in Figure 4.4. Of the five fit indices of interest, only two of these indices were reported by AMOS for the ULS solution: Standardized Root Mean Square Residual (SRMR) = 0.041 and Goodness of Fit Index (GFI) = 0.994.

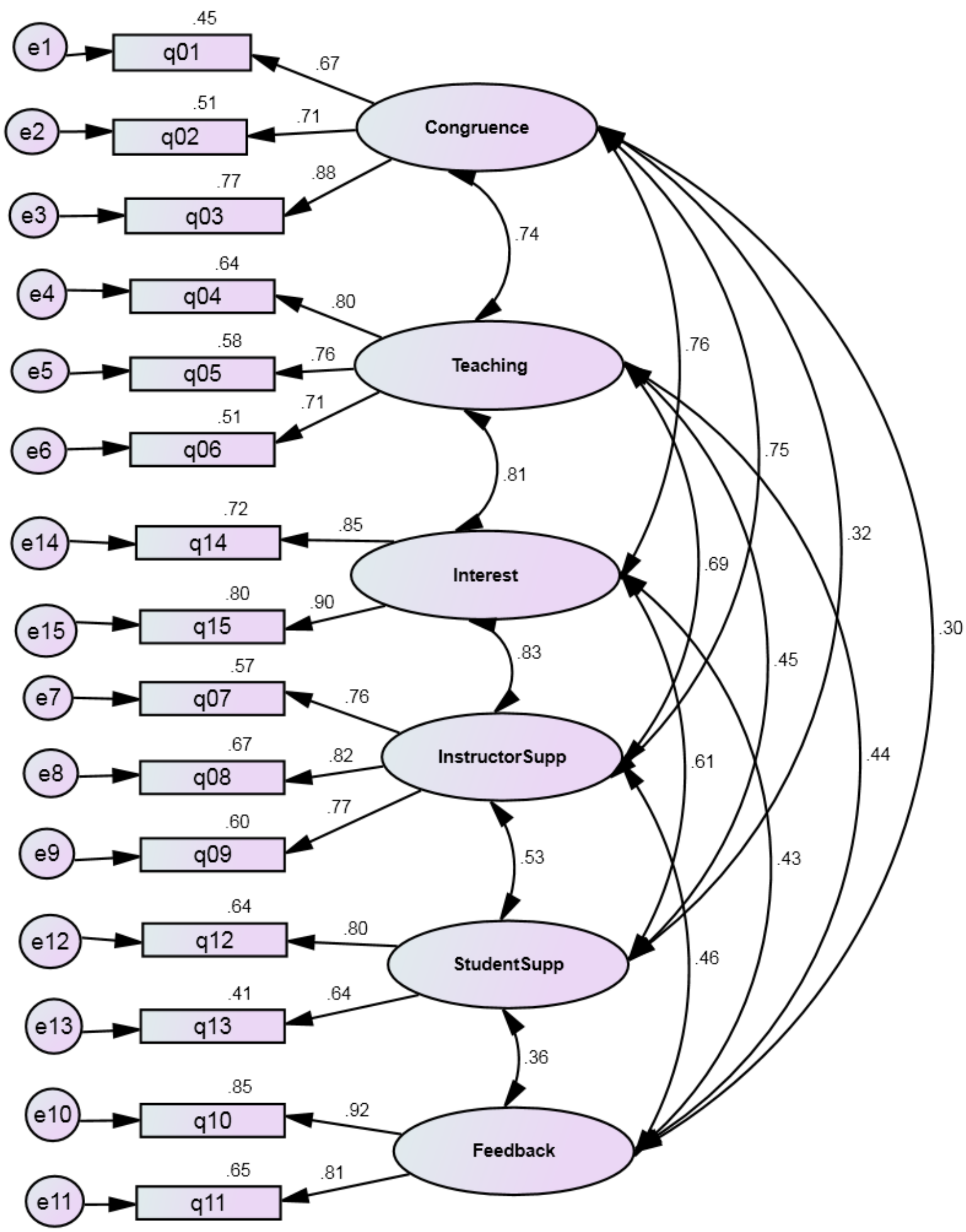

Figure 4.4. ULS solution for six-factor CFA model

COMPARING MODELS

To determine which model fit the Canadian data best, Tables 4.7 and 4.8 summarize goodness-of-fit statistics for the two competing models based on ML and ULS solutions respectively. Only two of the five selected goodness-of-fit indices were available when the ULS approach was selected in AMOS.

Table 4.8

Summary of ML-based goodness-of-fit indices for the two competing models

Model	SRMR	RMSEA	CFI	TLI	GFI
Five-factor	.057	.085	.925	.902	.882
Six-factor	**.043**	.070	**.953**	.934	**.913**

Note. Indices in bold indicate good model-to-data fit based on various ML-based criteria

Table 4.9

Summary of ULS-based goodness-of-fit indices for the two competing models

Model	SRMR	RMSEA	CFI	TLI	GFI
Five-factor	.057	-	-	-	.987
Six-factor	.041	-	-	-	.994

SUMMARY

Across both five-factor and six-factor solutions, as evidenced in Figures 4.1 to 4.4, latent factor correlations were found to all be positive (ranging between 0.29 and 0.88) providing support that these aspects of the teaching-learning environment, used to enhance quality of teaching and learning, were indeed interrelated.

High mean inter-item correlations were noted for items in the Feedback and Interest subscales. Comparing these results with those of the Entwistle (2009) study, in both studies (Table 4.5), high mean inter-item correlations (0.761 – current study and 0.709 – Entwistle study) were found for the Interest subscale suggesting that item redundancy could be a concern within this subscale. Furthermore, an item from Interest subscale, Q15, was also found to be highly correlated with items in other subscales (Table 4.2), providing further support to suggest that this item needed to be re-evaluated. Keeping in mind that correlation values are influenced by the population characteristics, the heterogeneous nature of the Entwistle (2009) study sample, which included students from five programs (Biology, Economics, Electronic Engineering, History and Media) versus the homogenous nature of the sample in the current study (all Education undergraduates) may provide some explanation as to why mean inter-item correlations for the current study were in general higher than those of the previous study.

Comparing the subscale statistics (Table 4.5) from the current study with those of the Entwistle (2009) study, the influence of internal consistency (measured by Pearson mean inter-item correlation) and number of items on the values of coefficient alpha can be clearly seen. For example, in the case of subscales with three items, the lowest value of alpha (0.404) corresponded to the lowest value of $\bar{r}_{ij}$ (0.668). This pattern was also present in the case of the remaining two-item subscales.

Finally, given that the sample size for this study was only one hundred and ninety nine ($N = 199$) and following the recommendations by Hu and Bentler (1999) for cases where $N < 250$, under ML-based goodness-of-fit indices (Table 4.7) only values of CFI and SRMR were compared because cut-off criteria based on RMSEA or TLI have been found to not be robust to the type of nonnormality present in ordinal data. Consequently, in the case of ordinal data and sample size less than two hundred and fifty, Hu and Bentler (1999) suggested that values if SRMR < 0.08 and CFI > 0.95 this would indicate good fit between a model and observed data. Guided by these recommendations, and comparing the five-factor and six-factor solutions across both ML and ULS solutions, the theorized six-factor model was found to provide a better fit to the Canadian sample than the alternative five-factor model.

CHAPTER FIVE

DISCUSSION

INTRODUCTION

Chapter five begins with a brief summary of the research goal and method of this study, followed by a discussion of the findings of the study in relation to the hypothesis that was to be tested. The next section discusses implications for research and practice, followed by limitations of the study. The chapter closes with recommendations for future research and conclusions drawn from results of this study in relation to previous studies.

SUMMARY OF THE STUDY

To enhance academic quality within the Canadian undergraduate education context, there is a need for student evaluation questionnaires that possess the diagnostic power needed to enhance students’ teaching and learning environments (Finnie & Usher, 2005; Gravestock & Gregor-greenleaf, 2008). The Short Revised Experiences of Teaching and Learning Questionnaire (SR-ETL-Q) (Entwistle, 2009) is a questionnaire that has accumulated validity evidence in favour of its use for improving the quality of teaching and learning within some undergraduate populations. Though population generalizability studies have been conducted to show that the SR-ETL-Q is useful for improving the quality of teaching and learning within undergraduate populations across different countries, to date, the generalizability of this tool had yet to be established in Canadian undergraduate populations. This study therefore adopted a non-experimental quantitative research design to collect validity evidence toward establishing the population generalizability of the internal structure of SR-ETL-Q responses to this new cultural context. Using confirmatory factor analyses, this study investigated whether the internal structure of SR-ETL-Q responses from a Canadian undergraduate sample would fit the theoretical six-factor model better than a plausible five-factor model alternative. Given that the theorized six-factor model found in the UK data also fit the Canadian sample data, this provided support for both the structural equivalence of SR-ETL-Q responses and for the equivalence of interpretations to be made using these responses within this new population.

IMPLICATIONS FOR RESEARCH

When building a validity argument for or against the use of an instrument, cross-cultural adaptation and validation are important as they provide evidence to support or refute the generalizability of an instrument for specific uses across various contexts. It is important to determine whether an instrument is generalizable as the generalizability of an instrument can affect the interpretations that can be made from its use within different contexts. If validity evidence does not support the generalizability of an instrument in different contexts, this may provide indication that the theory underlying an instrument cannot be supported elsewhere and thus interpretations made cannot be justified.

To date the generalizability of the ETL-Q and SR-ETL-Q has been analysed in European countries and China. The internal structure of ETL-Q responses was originally studied in UK (Entwistle et al., 2002; McCune, 2003).

Then, using cross-cultural adaptation and validation processes, researchers studied the generalizability of the ETL-Q within various undergraduate populations in China (Xu, 2004), Finland (Parpala et al., in press; Rytkönen, Parpala, Lindblom-Ylänne, Virtanen, & Postareff, 2011), and Belgium (Stes, De Maeyer, Gijbels, & Van Petegem, 2011). And finally, most recently, a study was conducted to ensure that the internal structure of SR-ETL-Q retained the structure of the ETL-Q (Entwistle, 2009).

The findings of the current study add to this growing body of literature surrounding the ETL-Q and SR-ETL-Q by providing evidence to support the generalizability of the SR-ETL-Q to a new context. While the validity evidence collected in the current study does support the generalizability of the SR-ETL-Q to a different context thus providing some indication that the theory underlying the instrument and the interpretations made using it may be justified within this Canadian population, it is important to remember that the population for this study came from a single course. There is therefore still a continued need to extend this body of work further so as to determine the generalizability of the SR-ETL-Q to the larger Canadian undergraduate context.

IMPLICATIONS FOR PRACTICE

To enhance academic quality in higher education institutions (HEIs), it is crucial that when designing quality teaching-learning environments, faculty remember that several interrelated factors influence student learning and if one element is in need of attention, this can interfere with students achieving high quality learning (Trigwell, 2010). To ensure the quality of teaching and learning in these environments, the 15 SR-ETL-Q items are designed to represent teaching-learning activities that encouraged high-quality learning (Entwistle, 2003). By designing the SR-ETL-Q in this manner, faculty are provided with a diagnostic tool that can give them access to information that may be used to inform their decisions surrounding teaching-learning environment design and implementation (Entwistle et al., 2002; Kember & Leung, 2008).

Within the Canadian undergraduate context, where course-level student evaluation questionnaires have been criticised for not providing the necessary diagnostic information that faculty need to enhance elements of students' teaching-learning environments (e.g. Gravestock & Gregor-greenleaf, 2008), the SR-ETL-Q may indeed be a promising tool that can be used to enhance students' teaching and learning environments. Due to the sample size and context (a single course) in which this study was conducted, there is an inability to generalize to the larger Canadian undergraduate population. However, within the given course under which this study was conducted, faculty may be encouraged to see that the finding from this current study does provide some support for the existing literature. If these faculty do choose to use the SR-ETL-Q, it could provide them with one way of getting students see the value of providing feedback if the results are used to for course enhancement. Furthermore, by providing students with evidence that feedback collected from them has been acted upon, this may not only encourage students to continue to provide feedback but can in turn enhance academic quality (Leckey & Neill, 2001; Rowley, 1996).

On a final note, giving feedback to students on their learning and obtaining feedback from students have traditionally been treated as separate tasks (Harvey, 2011), yet if faculty choose to utilise the SR-ETL-Q in their course, there is potential for symbiosis. By using the SR-ETL-Q to guide discussions between faculty and students about how to jointly improve students' teaching-learning experiences, students and faculty can begin to appreciate that both forms of feedback are focused on the same thing: feedback on learning (Harvey, 2011; Kember et al., 2002).

LIMITATIONS

Though the results provide indication that the SR-ETL-Q can be adapted to a Canadian sample, comparison between results of this study and previous studies were hampered by challenges surrounding data collection and data analysis procedures.

In the original study using the SR-ETL-Q, an exploratory factor analysis was conducted on a large sample (N = 1950) (Entwistle, 2009). A large sample was achieved by providing the questionnaire to students in paper-based format, during class time, toward the end of their courses. Moreover, students came from a variety of courses in different programs allowing for a large heterogeneous dataset in which to conduct the analysis. In this study, however, the SR-ETL-Q was embedded in an online survey that, though available to a population of seven hundred undergraduate students in a single multi-section course, was only completed by 28.4% (n = 199) of the students. Unfortunately, this response rate is consistent with findings in the literature (on paper-based versus online course evaluations) that indicate that in general response rates for online surveys are a great deal smaller than those possible with paper-based administrations (e.g. Dommeyer, Baum, Hanna, & Chapman, 2004; Lefever, Dal, & Matthíasdóttir, 2007; Nulty, 2008).

In regard to data analysis, though steps were taken to ensure that analyses were conducted using techniques for ordinal data (e.g. Basto & Pereira, 2012; Choi et al., 2010; Finch, 2010; Flora & Curran, 2004; Forero et al., 2009; Yang-Wallentin, Joreskog, & Luo, 2010; Zumbo et al., 2007), in order to be able to make comparisons with results from existing literature, it was necessary to use Pearson correlations and maximum likelihood estimation. Though Pearson correlations and maximum likelihood estimation are commonly used by researchers for factor analyses, particularly when large sample sizes are present, these techniques assume that data are continuous and normally distributed which is not usually the case with ordinal data.

RECOMMENDATIONS FOR FUTURE RESEARCH

To begin to address the limitations of this study, while also continuing to add to the two relevant bodies of literature, recommendations for future research include:

1. Conducting similar studies with larger sample sizes;
2. Where possible, an in-class paper-based version of the SR-ETL-Q should be used rather than making it available online so as to increase response rates. If an online version must be used, incentives, such as providing a

grade incentive (e.g. one-quarter of a percent) for completing a survey, may help to increase response rates (Dommeyer et al., 2004);

3. Conducting additional studies to collect other sources of validity evidence, particularly in regard to response processes. There is a body of research looking into cognitive aspects of response processes (e.g. Collins, 2003; Schwarz, 2007; Tourangeau & Rasinski, 1988), yet this line of research has yet to be used to analyse the SR-ETL-Q. Examining response processes through this line of research may help to determine whether Q15 indeed needs to be redesigned; and finally,
4. Conducting future analyses using procedures designed for ordinal data.

CONCLUSIONS

The goal of the study was to use confirmatory factor analysis to test whether the theorized six-factor model of SR-ETL-Q responses found in the Entwistle (2009) study was generalizable to a Canadian undergraduate population and this indeed was the finding. Although the results of this study provide support for this conclusion, this finding must be taken within the scope of this study and the aforementioned limitations. Furthermore, though this is but one small-scale validation study, the study does successfully begin to fill a gap in the literature by providing empirical evidence on the use of a theory-based student evaluation questionnaire within the Canadian higher education context. Finally, this study also adds to another body of literature by providing another source of validity evidence to strengthen the validity argument in support of the use of SR-ETL-Q responses for enhancing teaching-learning environments.

BIBLIOGRAPHY

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing* (2nd ed.). Washington, DC: AERA.

Anderson, C., & Day, K. (2005). *Subject Overiew Report: History*. *ETL Project*.

Arbuckle, J. L. (2011). IBM ® SPSS ® Amos ™ 20 User's Guide. IBM Corp.

Baker, D. N., & Miosi, T. (2010). The Quality Assurance of Degree Education in Canada. *Research in Comparative and International Education*, *5*(1), 32. doi:10.2304/rcie.2010.5.1.32

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. *American journal of epidemiology*, *160*(1), 34-45. doi:10.1093/aje/kwh175

Basto, M., & Pereira, J. M. (2012). An SPSS R-Menu for Ordinal Factor Analysis. *Journal Of Statistical Software*, *46*(4).

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186-3191.

Berry, J. (1969). On Cross-Cultural Comparability. *International Journal of Psychology*, *4*(2), 119-128. doi:10.1080/00207596908247261

Canadian Council on Learning. (2009). *“Up to Par: The Challenge of Demonstrating Quality in Canadian Post-Secondary Education,” Challenges in Canadian Post-secondary Education*. *Learning*. Ottawa, Ontario: Canadian Council on Learning.

Choi, J., Peters, M., & Mueller, R. O. (2010). Correlational analysis of ordinal data: from Pearson’s r to Bayesian polychoric correlation. *Asia Pacific Education Review*, *11*(4), 459-466. doi:10.1007/s12564-010-9096-y

Coates, H., & Seifert, T. (2011). Linking assessment for learning, improvement and accountability. *Quality in Higher Education*, *17*(2), 179-194. doi:10.1080/13538322.2011.554308

Collins, D. (2003). Pretesting survey instruments: an overview of cognitive methods. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, *12*(3), 229-38.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. Springer.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis. *Psychological Methods*, *1*(1), 16-29.

Dommeyer, C., Baum, P., Hanna, R., & Chapman, K. (2004). Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, *29*(5), 611-623. doi:10.1080/0260293041000168917l

Entwistle, N. (2003). *Concepts and conceptual frameworks underpinning the ETL project*. *Enhancing Teaching-Learning Environments in Undergraduate Courses Project*. Retrieved from http://www.tla.ed.ac.uk/etl/docs/ETLreport3.pdf

Entwistle, N. (2005a). Ways of thinking and ways of teaching across contrasting subject areas. *Curriculum Journal*, *16*(1), 67-82. New York, NY: Palgrave Macmillan. doi:10.1080/0958517042000336818

Entwistle, N. (2005b). *Enhancing Learning and Teaching in Electronic Engineering: A Digest of Research Findings and their Implications. ETL Project.* Retrieved from www.ed.ac.uk/etl/publications.html

Entwistle, N. (2008). Taking stock : teaching and learning research in higher education. *International symposium on "Teaching and Learning Research in Higher Education", Guelph, Ontario*.

Entwistle, N. (2009). *Teaching for Understanding at University: Deep approaches and distinctive ways of thinking*. New York, NY: Palgrave Macmillan.

Entwistle, N., & McCune, V. (2004). The Conceptual Bases of Study Strategy Inventories. *Educational Psychology Review*, *16*(4), 325-345. doi:10.1007/s10648-004-0003-0

Entwistle, N., McCune, V., & Hounsell, J. (2002). Approaches to studying and perceptions of university teaching-learning environments: Concepts, measures and preliminary findings. *Enhancing Teaching and Learning Environments in Undergraduate Courses Occasional Report*, *1*. Retrieved from http://www.etl.tla.ed.ac.uk/docs/etlreport1.pdf

Entwistle, N., Nisbet, J., & Bromage, A. (2005). *Subject Overview Report: Electronic Engineering. ETL Project.* Retrieved from www.ed.ac.uk/etl/publications.html

Finch, W. H. (2010). Imputation Methods for Missing Categorical Questionnaire Data: A Comparison of Approaches. *Journal of Data Science*, *8*, 361-378.

Finnie, R., & Usher, A. (2005). *Measuring the quality of post-secondary education: concepts, current practices and a strategic plan*. *Educational Policy*. Ottawa, Ontario: Canadian Policy Research Networks.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, *9*(4), 466-91. doi:10.1037/1082-989X.9.4.466

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625-641. doi:10.1080/10705510903203573

Gravestock, P., & Gregor-greenleaf, E. (2008). *Student course evaluations: Research, models and trends*. Toronto, Ontario: Higher Education Quality Council of Ontario.

Haarala-Muhonen, A., Ruohoniemi, M., Katajavuori, N., & Lindblom-Ylänne, S. (2011). Comparison of students' perceptions of their teaching–learning environments in three professional academic disciplines: A valuable tool for quality enhancement. *Learning Environments Research*, *14*(2), 155-169. doi:10.1007/s10984-011-9087-x

Haggis, T. (2009). Student Learning Research: A Broader View. In M. Tight, K. Ho Mok, J. Huisman, & C. Morphew (Eds.), *Routledge International Handbook of Higher Education* (Vol. 4, pp. 51-69). London, UK: Routledge.

Harvey, L. (2011). The nexus of feedback and improvement. In C. S. Nair & P. Mertova (Eds.), *Student Feedback: The cornerstone to an effective quality assurance system in higher education* (pp. 3-26). Oxford, UK: Chandos Publishing.

Harvey, L., & Green, D. (1993). Defining Quality. *Assessment & Evaluation in Higher Education*, *18*(1), 9-34. doi:10.1080/0260293930180102

Harvey, L., & Knight, P. T. (1996). *Transforming Higher Education.* Bristol, PA: Taylor & Francis. Retrieved from http://eric.ed.gov/ERICWebPortal/recordDetail?accno=ED418640

Hounsell, D., & Anderson, C. (2005). Ways of thinking and practising in biology and history: disciplinary aspects of teaching and learning environments. *Higher Education Colloquium, Centre for Teaching, Learning and Assessment, Teaching and Learning within the Disciplines, University of Edinburgh, June 10-11* (pp. 1-12).

Hounsell, D., Entwistle, N., Anderson, C., Bromage, A., Day, K., Hounsell, J., Land, R., Litjens, J., McCune, V., Meyer, E., Reimann, N., & Xu, R. (2005). *Enhancing teaching-learning environments in undergraduate courses: End-of-Award Report to ESRC on project L139251099. Summarizing statement. Retrieved November* (Vol. 11, pp. 1-20). Retrieved from http://www.tla.ed.ac.uk/etl/docs/etlfinalreport.pdf

Hounsell, D., Hounsell, J., Litjens, J., & McCune, V. (2005). Enhancing guidance and feedback to students: findings on the impact of evidence-informed initiatives. *European Association for Research on Learning and Instruction (EARLI) 11th Biennial Conference, Nicosia, Cyprus* (pp. 1-26). Retrieved from http://www.etl.tla.ed.ac.uk/docs/earliHHLM.pdf

Hounsell, D., McCune, V., Litjens, J., & Hounsell, J. (2005). *Subject Overview Report: Biosciences. ETL Project*. *of Edinburgh, Durham* (pp. 1-65). Edinburgh. Retrieved from http://www.ed.ac.uk/etl/docs/BiosciencesSR.pdf

Hutchinson, S. R., & Olmos, A. (1998). Behavior of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(4), 344-364. doi:10.1080/10705519809540111

IBM Corporation. (2011). *IBM SPSS Missing Values 20*. Armonk, NY: IBM Corp.

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Personality Psychology* (pp. 461-494). New York, NY: Guilford Press.

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten spezial*, *3*, 1–40. Retrieved from http://www.gesis.org/fileadmin/upload/forschung/publikationen/zeitschriften/zuma_nachrichten_spezial/znspezial3.pdf

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education/Praeger.

Kember, D., & Leung, D. Y. P. (2008). Establishing the validity and reliability of course evaluation questionnaires. *Assessment & Evaluation in Higher Education*, *33*(4), 341-353. doi:10.1080/02602930701563070

Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the Use of Student Feedback Questionnaires Improve the Overall Quality of Teaching? *Assessment & Evaluation in Higher Education*, *27*(5), 411-425. doi:10.1080/0260293022000009294

Kis, V. (2005). Quality assurance in tertiary education: Current practices in OECD countries and a literature review on potential effects. *Tertiary Review: A contribution to the OECD thematic review of tertiary education*. Retrieved from http://www.pisa.oecd.org/dataoecd/55/30/38006910.pdf

Land, R., Reimann, N., & Meyer, J. (2005). *Enhancing Learning and Teaching in Economics: A Digest of Research Findings and their Implications. ETL Project.* Retrieved from www.ed.ac.uk/etl/publications.html

Law, D. C. S. (2010). Quality assurance in post-secondary education: the student experience. *Quality Assurance in Education*, *18*(4), 250-270. doi:10.1108/09684881011016007

Leckey, J., & Neill, N. (2001). Quantifying Quality: The importance of student feedback. *Quality in Higher Education*, *7*(1), 19-32. doi:10.1080/13538320120045058

Lefever, S., Dal, M., & Matthíasdóttir, Á. (2007). Online data collection in academic research: advantages and limitations. *British Journal of Educational Technology*, *38*(4), 574-582. doi:10.1111/j.1467-8535.2006.00638.x

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Marsh, H. W., Hau, K.-tai, & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320-341. doi:10.1207/s15328007sem1103_2

McCune, V. (2003). Promoting high-quality learning: perspectives from the ETL project. *Norwegian Network in Higher Education 14th Conference* (pp. 1-29). Retrieved from http://en.scientificcommons.org/8677424

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education and Macmillan.

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: what can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301-314. doi:10.1080/02602930701293231

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The Performance of ML, GLS, and WLS Estimation in Structural Equation Modeling Under Conditions of Misspecification and Nonnormality. *Structural Equation Modeling*, *7*(4), 557-595.

Parpala, A., Lindblom-Ylänne, S., Komulainen, E., Litmanen, T., & Hirsto, L. (2010). Students’ approaches to learning and their experiences of the teaching-learning environment in different disciplines. *The British journal of educational psychology*, *80*(Pt 2), 269-82. doi:10.1348/000709909X476946

Parpala, A., Lindblom-ylänne, S., Komulainen, E., & Entwistle, N. (in press). Assessing students’ experiences of teaching-learning environments and approaches to learning: validation of a questionnaire used in different countries and varying contexts. *Learning Environments Research*.

Reimann, N. (2004). First Year Teaching-Learning Environments in Economics. *International Review of Economics Education*, *31*(1), 9-38.

Reimann, N., Land, R., & Meyer, J. (2005). *Subject Overview Report: Economics. ETL Project.* Retrieved from www.ed.ac.uk/etl/publications.html

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education*, *30*(4), 387-415. doi:10.1080/02602930500099193

Rowley, J. (1996). Measuring Quality in Higher Education. *Quality in Higher Education*, *2*(3), 237-255. doi:10.1080/1353832960020306

Rowley, J. (2003). Designing student feedback questionnaires. *Quality Assurance in Education*, *11*(3), 142-149. doi:10.1108/09684880310488454

Russell, D. W. (2002). In Search of Underlying Dimensions: The Use (and Abuse) of Factor Analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, *28*(12), 1629-1646. doi:10.1177/014616702237645

Rytkönen, H., Parpala, A., Lindblom-Ylänne, S., Virtanen, V., & Postareff, L. (2011). Factors affecting bioscience students' academic achievement. *Instructional Science*, *40*(2), 241-256. doi:10.1007/s11251-011-9176-3

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical methods in medical research*, *8*(1), 3-15. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10347857

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177. doi:10.1037//1082-989X.7.2.147

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, *99*(6), 323–338. Heldref Publications.

Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, *21*(2), 277-287. doi:10.1002/acp.1340

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, *74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Stes, A., De Maeyer, S., Gijbels, D., & Van Petegem, P. (2011). Instructional development for teachers in higher education: Effects on students' perceptions of the teaching-learning environment. *British Journal of Educational Psychology*, 1-22. doi:10.1111/j.2044-8279.2011.02032.x

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103. doi:10.1207/S15327752JPA8001_18

Tam, M. (2001). Measuring Quality and Performance in Higher Education. *Quality in Higher Education*, *7*(1), 47-54. doi:10.1080/13538320120045076

Tarkkonen, L., & Vehkalahti, K. (2005). Measurement errors in multivariate measurement scales. *Journal of Multivariate Analysis*, *96*(1), 172-189. doi:10.1016/j.jmva.2004.09.007

Tourangeau, R., & Rasinski, K. A. (1988). Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin*, *103*(3), 299-314.

Trigwell, K. (2010). Promoting effective student learning in higher education. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International encyclopedia of education: volume 4* (3rd ed., Vol. 4, pp. 461-466). Oxford, UK: Elsevier.

Vlăsceanu, L., Grunberg, L., & Parlea, D. (2007). *Quality assurance and accreditation: A glossary of basic terms and definitions*. Retrieved from http://www.aracis.ro/fileadmin/ARACIS/Publicatii_Aracis/Publicatii_ARACIS/Engleza/Glossary_07_05_2007.pdf

Ximénez, C. (2006). A Monte Carlo Study of Recovery of Weak Factor Loadings in Confirmatory Factor Analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(4), 587-614. doi:10.1207/s15328007sem1304_5

Xu, R. (2004). Chinese Mainland students' experiences of teaching and learning at a Chinese university: some emerging findings. *BERA 2004 Conference (UMIST, Manchester: BERA), 15th-18th* (pp. 1-19). Retrieved from http://www.etl.tla.ed.ac.uk/docs/XuBERA.pdf

Yang-Wallentin, F., Joreskog, K., & Luo, H. (2010). Confirmatory Factor Analysis of Ordinal Variables With Misspecified Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *17*(3), 392-423. doi:10.1080/10705511.2010.489003

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal Versions of Coefficients Alpha and Theta for Likert Rating Scales. *Journal of Modern Applied Statistical Methods*, *6*, 21-29.

APPENDIX

Modified SR-ETL-Q
Experiences of Teaching and Learning

This questionnaire has been designed to allow you to describe, in a systematic way, your reactions to the course you have been studying and how you have gone about learning it. There are a series of questions, some of which overlap so as to provide good overall coverage of different experiences. Most of the items are based on comments made by other students. Please respond truthfully, so that your answers describe your **actual** experiences in this **particular course**, working your way through the questionnaire **quickly**. It is important that you respond to **every** item. Please circle the appropriate number to indicate your response.

1 *= disagree* *2 = disagree somewhat* ***3*** *= unsure* *4 = agree somewhat* *5 = agree*

		disagree	*disagree somewhat*	*unsure*	*agree somewhat*	*agree*
1	It was clear to me what I was supposed to learn in this course.	**1**	**2**	**3**	**4**	**5**
2	The topics seemed to follow each other in a way that made sense to me.	**1**	**2**	**3**	**4**	**5**
3	What we were taught seemed to match what we were supposed to learn.	**1**	**2**	**3**	**4**	**5**
4	This course has given me a sense of the thinking that goes on 'behind the scenes' in this subject area.	**1**	**2**	**3**	**4**	**5**
5	The teaching in this course helped me to think about the evidence underpinning different views.	**1**	**2**	**3**	**4**	**5**
6	This course encouraged me to relate what I learned to issues in the wider world.	**1**	**2**	**3**	**4**	**5**
7	Instructor(s) tried to share their enthusiasm about the subject with us.	**1**	**2**	**3**	**4**	**5**
8	Instructor(s) were patient in explaining things which seemed difficult to grasp.	**1**	**2**	**3**	**4**	**5**
9	Students' views were valued in this course.	**1**	**2**	**3**	**4**	**5**
10	The feedback given on my assignments helped me to improve my ways of learning and studying.	**1**	**2**	**3**	**4**	**5**
11	The feedback given on my assignments helped to clarify things I hadn't fully understood.	**1**	**2**	**3**	**4**	**5**
12	Students supported each other and tried to give help when it was needed.	**1**	**2**	**3**	**4**	**5**
13	Talking with other students helped me to develop my understanding.	**1**	**2**	**3**	**4**	**5**
14	I found most of what I learned in this course really interesting.	**1**	**2**	**3**	**4**	**5**
15	I enjoyed being involved in this course.	**1**	**2**	**3**	**4**	**5**

Please check back to make sure that you have answered every question.

Thank you very much for spending time completing this questionnaire: it is much appreciated.

(Adapted from *Teaching for understanding at university: Deep approaches and distinctive ways of thinking* (p. 115), by N. Entwistle, 2009, New York, NY: Palgrave Macmillan.