PRELIMINARY PSYCHOMETRIC CHARACTERISTICS OF THE CRITICAL THINKING SELF-ASSESSMENT SCALE

A Thesis Submitted to the College of

Graduate Studies and Research

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy

in the College of Nursing

University of Saskatchewan

Saskatoon, CANADA.

By

Girija Gopinathan Nair

PERMISSION TO USE

College of Nursing
University of Saskatchewan
Saskatoon, Saskatchewan [S7N 5E5]
Canada

OR

Dean
College of Graduate Studies and Research
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan S7N 5A2
Canada

ABSTRACT

Critical thinking skills (CTS) have been conceptualized as having six core cognitive skills as purported by the American Philosophical Association (APA) through a Delphi survey. The APA report further provided directions for teaching - learning and evaluation of these cognitive skills. This scale construction study was based on the APA critical thinking skills construct definition. Using the APA evaluation directions as a guide, this researcher developed a self assessment scale for measuring the CTS of undergraduate nursing students with the intention of assisting students in developing and improving their thinking skills. The construction of the scale was based on Cronbach's Generalizability theory, and used Messick's (1990) unitary concept of construct validity framework for evaluating the psychometric properties of the scale.

The researcher developed 196 peer reviewed items for the Critical Thinking Self Assessment Scale (CTSAS) and the scale was subjected to experts' ratings to establish content relevance and representativeness of the items to the construct. Seventeen experts from different disciplines reviewed the items and rated the items as 3 or 4 if the items defined the construct. Mean, Median, range and Content Validity Index (*I-CVI*) and Aiken's Content Validity Coefficient (VI_k) were computed to retain, modify or reject the items. The decision criteria for retaining the items included a value of VI_k significant at $p < 0.05$, a value of $I\text{-}CVI \geq 0.75$, and a range value of < 2.75 for the '0 to 5' rating continuum. Statistical analysis of the item ratings resulted in reducing 196 items to 115. Following the rigorous content validation process, the 115 item CTSAS was tested through two developmental samples; one of 887 undergraduate nursing students from five Colleges of Nursing from Mahatma Gandhi University of Kerala State, India, and the second 144 undergraduate students from the College of Nursing, University of Saskatchewan, Canada. The questionnaire booklet also included an 18 item Need for Cognition

Scale (NCS-SF) developed by Cacioppo and Petty (1982) for testing convergent validity of CTSAS.

Exploratory data screening for the participants' responses resulted in the deletion of four items (both the samples showed similar results in these 4 items) and 19 cases from the Indian sample, which were either missing, skewed or outliers. The remaining 111 items were analyzed for internal consistency reliability with both Indian and Canadian samples and stability reliability with the retested Indian sample (251). Exploratory Factor Analysis (EFA) using Principal Component Analysis (PCA) with varimax and oblimin rotations was run for the six core scales separately, which were classified into 16 sub scales, with the Indian sample (868). This resulted in reducing 111 items to 90 items across 14 subscales. Two of the subscales failed to emerge in EFA. The item loadings to factors demonstrated homogeneity and loaded independently with large loading weights. The items loading were mostly consistent with the pre-designated scales.

The EFA retained 90 items were fixed in six path diagrams in the Analysis of Moment Structure (AMOS, added program in SPSS-PASW Statistics 18) graphics and Confirmatory Factor Analysis (CFA) was run with the 144 Canadian sample for each of the core scales to see the model fit. Three of the six core scales demonstrated acceptable goodness of fit indices and the remaining three reached almost reasonable to close fit. The Maximum Likelihood (ML) estimation-minimum discrepancy function-χ^2 values were significant for all six core scales. However, the three model fit scales had a ratio of χ^2 to degrees of freedom (CMIN / df) < 2 indicating good model fit. The Null hypothesis "not - close fit" ($H_0 = \varepsilon \geq 0.05$) was rejected in favour of the research hypothesis and it may be concluded that fit of the model in the population is close (i.e., $\varepsilon \leq 0.05$). The fit indices for the three core scales - Interpretation, Evaluation, and Inference, strongly supported the structural fidelity of the three core scales, and it is plausible to replicate similar findings in a comparable population. The results also supported the APA critical thinking construct definition for the three cognitive skills.

All the core scales revealed a reliability value ≥ 0.80 for the core scales. Some of the subscales achieved lower levels of correlation, but none were lower than 0.60. The total scale had very good internal consistency reliability; Cronbach α for the Indian sample was 0.961 and for the Canadian sample 0.975, and had high levels of communalities required for reducing the length of the scale. However, EFA and CFA gave strong results indicating further testing and analyzing the scale was necessary to refine the items. The convergent validity of the CTSAS tested with NCS-SF found significant correlations for five of the six core scales. The main limitation of the study was inadequate sample size for performing CFA. The socio-cultural influence on critical thinking was not tested. The study examined only some aspects of Messick's unitary concept of construct validity for establishing the psychometric of the CTSAS. However, the preliminary psychometrics results of the study were very appealing and the researcher is encouraged to further examine the usability of the scale and ensuring socio-cultural relevance of the CTSAS.

ACKNOWLEDGEMENTS

I am extremely grateful to, my supervisor, Dr. Lynnette Leeseberg Stamler for her sustained interest and enthusiasm in my project throughout the period of my study. Her supervision, guidance and support, both academically and personally, was invaluable in accomplishing this endeavor.

I extend my sincere gratitude to Dr. Donna Rennie, Chairperson, of my research committee for the support and guidance provided as and when required.

Dr. Laurie Hellsten has been an inspiring resource for the accomplishment of this project and I greatly appreciate the wisdom that was shared during our series of discussions and the timely guidance throughout the project even during her sabbatical.

I am indebted to Dr. Karen Semchuk, for the guidance and especially the support provided for facilitating my data collection in Canada.

I am grateful to Dr. Phil Woods and Dr. Sandra Bassandowski for their valuable discussions and continued support for my study especially during the planning stage. I thank Dr. Rose Roberts who was my co-supervisor for the beginning part of my project.

I humbly acknowledge the valuable contributions of all 17 content validation experts who willingly participated in the review and provided constructive suggestions for finalizing the items for the scale.

I place on record the discussions and assistance provided by Dr. Lucine M. Huckabey, Director and Professor, Department of Nursing, California State University, Long Beach, U S for generating the items for the scale and being a member of the panel for content validation.

I am indebted to Dr. Lisa Lix for the enriching discussions and guidance for planning and executing the analysis sections of the project.

I thank the SPSS Inc., Bangalore, India for assisting in fixing up the AMOS program without which the last part of the analysis would not have been possible.

I am deeply indebted to Mr. Pratap L. Murthy, Statistician & Data analyst, SPSS Inc., Bangalore, India for his invaluable assistance and discussions with the Confirmatory Factor Analysis in AMOS.

I am thankful to the Dean, Associate Deans, Faculty and administrative staff, College of Nursing, University of Saskatchewan for granting permission for data collection at three sites and facilitating my data collection.

I am thankful to the Directors and Principals of all five Colleges of Nursing, Kerala, India who consented to the data collection at their site.

I gratefully acknowledge my peer group in the instrument development course and my colleagues in the PhD program for their valuable comments and support.

I am highly obliged to all the student participants of U of S, Canada and the five Colleges of Nursing, Kerala, India for their frank responses and commitment to cooperate and complete the scale with care. Without their wholehearted participation this project would not have been a reality.

I am thankful to Mr. Richard Petit, Chief Executive Officer, Keewatin Yatthè Regional Health Authority for allowing my academic time and facilitating my research by granting leave and long vacation for data collection from work.

A special thanks to Sharon Kimbley, Director of Population Health, KYRHA, for facilitating my data collection by sharing my job.

I acknowledge the intellectual development that I have acquired from U of S through a well planned and executed doctoral study.

A great thank you to Tammy Morrison for the administrative assistance provided throughout my study. Thank you to Bruce Atamanenko, Nursing Technician for his valuable help with my presentations.

I am indebted to my friend, Patricia J. Zipchen who provided me with a comfortable stay in Saskatoon and her constant encouragement in completing this study which was invaluable.

I am grateful to Dr. Manuel James and family for facilitating my doctoral education in Canada. I am grateful to all my friends in India who have extended their help during data collection.

Finally, I thank my son, Gyby G. Nair, who has extended his wisdom and graphic skills in completing my project. I am also grateful to my family for encouraging me to fulfill my aspiration.

DEDICATION

I dedicate this dissertation to my loving husband Gopinathan Nair,

my son Gyby. G. Nair and daughter-in-law Arathi. G Nair

for their invaluable support and encouragement.

Further, I dedicate this project to all future undergraduate students

who would use this scale for enhancing their critical thinking.

TABLE OF CONTENTS

Note : Appendices **C, D, E & F** are the attachments for the Section 2, Manuscript 1

Appendices **E, F, & G** are the attachments for the Section 3, Manuscript 2

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

1. APA: American Philosophical Association
2. CTS: Critical Thinking Skills
3. CTSAS: Critical Thinking Self-Assessment Scale
4. CCTST: California Critical Thinking Skill Test
5. EDA: Exploratory Data Analysis
6. EFA: Exploratory Factor Analysis
7. CFA: Confirmatory Factor Analysis

DEVELOPING AND EVALUATING THE PRELIMINARY

PSYCHOMETRIC CHARACTERISTICS OF THE

CRITICAL THINKING SELF-ASSESSMENT SCALE FOR

UNDERGRADUATE NURSING STUDENTS

SECTION 1

INTRODUCTION

Developing and Evaluating the Preliminary Psychometric Characteristics of the Critical Thinking Self-Assessment Scale for Undergraduate Nursing Students

Sumner (1940) asserted that critical faculty or skill is a product of education and training and that Critical Thinking Skills (CTS) guarantee against delusion, deception, superstition, and misapprehension of our circumstances and ourselves (Paul, 2009). Critical thinking skills are an expected outcome of undergraduate nursing education programs in several countries because nurses need these skills to cope with the ever-changing complex health care system of 21st century. Programs in nursing education therefore, are required to include CTS as a learning outcome of the curriculum to achieve and maintain accreditation (Commission on Collegiate Nursing Education (CCNE), 2003; National League of Nursing Accrediting Commission (NLNAC), Inc., 2004). The NLNAC (2004 revised 2008), in its accreditation manual for post secondary and higher education in nursing, advocates that the nursing students be taught these cognitive skills in the program and provided adequate opportunities for their practice, and that nursing programs show evidence that their students have developed CTS.

Organization of the Dissertation

This dissertation encompasses a series of scientific steps undertaken to develop a Critical Thinking Self Assessment Scale (CTSAS) and evaluate its psychometric properties. This dissertation is organized and presented in a hybrid format. The two manuscripts and a section on analysis, results and interpretation, which were the intended outcomes of the study, and a section on summary, discussion, limitations, next steps, conclusions and implications were appropriately integrated into this dissertation. The first section of the dissertation includes the introduction, and the need and background of the study. Within this section of the dissertation a research plan outlines the developmental research, noting the purpose and objectives of the study, the methodology (design, setting, sample, and data collection instruments), the ethical considerations, and plan of analysis of data. The steps purported by DeVellis (2003) were followed in sequencing the sections of dissertation. In the second section, the first manuscript describes the process of identifying and deriving a conceptual framework for basing the CTSAS which was the first crucial step in the scale development. This included an extensive literature review to identify appropriate constructs for generating items for CTSAS and a proposed conceptual framework on which to base the scale.

In the section three, manuscript two is a methodological paper that reports the process and procedures employed in the content validation process for accumulating evidence for content validity of the CTSAS. Section four of the dissertation presents the analysis, results and interpretation of test scores obtained for the target population for establishing construct validity, which includes both internal (Exploratory Factor Analysis) and external validation (Confirmatory Factor Analysis), reliability, and convergent validity. The last section, section five, consists of a summary, discussion, limitations of the study, next steps for future research, implications of the findings and conclusions, followed by sections for references, and appendices. The data from these sections will be converted to manuscripts following the defense of the dissertation.

Need and Background of the Study

The Canadian Association of Schools of Nursing (CASN) and Saskatchewan Registered Nurses Association (SRNA) regulatory requirement states that critical inquiry and judgment are one of the educational outcomes in nurses for practice (CASN, 2005; SRNA, 2007). The Indian Nursing Council (INC) requires nursing programs to demonstrate evidence of accomplishment of CTS in graduates for accreditation of the program. The philosophy of the undergraduate program affirms that educational efforts be " directed to the development of critical thinking skills, competencies and standards required for practice of professional nursing and midwifery as envisaged in National Health Policy 2002" (INC, 2004, p.1). Several nursing educators have pointed out that nurses cannot realistically be expected to learn new and complex ways of thinking on their own without formal instruction and opportunities to practice these ways of thinking (Ibarreta & McLeod, 2004; Ironside, 2004; Simpson & Courtney, 2002).

Nursing education literature revealed a resurgence during the last two decades in developing learning settings that enhance nurses' critical thinking skills, as nursing education has shifted from a largely medically based model to a more holistic model (Carter & Rukholm, 2008; Colucciello, 1997; Daroszewski, Kinser, & Lloyd, 2004; Dickieson, Carter, & Wlash, 2008; Janicek, 2006; Johns, 1995; Kessler & Lund, 2004; Khosravani, Manoochehri, & Memarian, 2005; McGrath, 2003; McGrath, Sheskith, Lang & Estabrooks, 2003; Staib, 2003). Additional reasons why nurses must be competent critical thinkers include dramatic changes in health care related to "information technology, fiscal cutbacks, human resource limitations, and the acuity of many patient care situations" (Carter & Rukholm, 2008, p. 134). Nursing practice requires ongoing and interactive understanding of both the context of care and patients' experiences of wellness and illness (Ironside, 2003). Cody (2002) claimed that competent nursing practice requires much more than content knowledge. These factors have led to the development of learning opportunities that nurture students' growth as thinkers.

Developing CTS in students is an essential role of education. Teaching nurses and students to think critically requires mentors or teachers to be sound critical thinkers themselves. To be effective teachers of CTS, teachers need to equip themselves with the cognitive skills involved in CTS, and possess an understanding of the nature of learning opportunities that can foster CTS among students. Nursing educators often face challenges in deriving appropriate teaching and assessment strategies to show evidence of the achievement of CTS in students. They must be able to see the learners' thinking by observing the mental processes and the logic (thinking skills) that students use in deriving theory based clinical decisions (Brunt, 2005). Ensuring successful preparation of graduates who can think critically is a concern of assessment and evaluation strategies such as accreditation.

Nursing educators are now concerned about the strategies that target evidence of CTS as an outcome of education because such evidence is a requirement for accreditation of the programs and desired by employing agencies. Direct evidence for the development of CTS in students could be demonstrated using outcome assessments, validated instrumentation, course assessments based on CTS, and assessing clinical judgment (Facione & Facione, 1994). However, such tools and devices are sparingly available and those that are available are not easily accessible to students, teachers, and student researchers. For the last two decades, general educators, nurse educators, and critical thinking experts have recognized the need for developing CTS and assessing CTS in college graduates. However, research instruments that measure teaching methodologies that enhance CTS or progression of CTS are limited. Though most nursing faculties admit their responsibility to assess students' critical thinking skills, they face an ongoing challenge in measuring students' ability to think critically due to lack of appropriate instruments or strategies available to them (Morrison & Free, 2001). Thus, there is a dire need for scientifically constructing additional devices for measuring CTS.

Developing a valid and reliable instrument for measuring critical thinking is a challenge because the instrument needs to address the "subtleties and sophistication of critical thinking required for content rich and highly contextualized" nature of the discipline of practice (Facione & Facione, 1996, p. 42). There are a number of standardized tests available commercially to measure general CTS which are not economical or easily accessible to students to check their progress. Examples include the Watson Glaser Critical Thinking Appraisal (WGCTA), the California Critical Thinking Skills Test (CCTST), the Ennis-Weir Critical Thinking Essay Test (EWCTET), the Cornell Critical Thinking Test (CCTT) and the International Critical Thinking Essay Test. Researchers evaluating CTS of students using WGCTA have reported inconsistent results in assessing CTS (Adams, Whitlow, Stover & Johnson, 1996; Rubenfeld & Scheffer, 1999). More recently, Wagner and Harvey (2006) compared the WGCTA with the Wagner Assessment Test (WAT) using item response theory and found that standard errors of the WGCTA were 50% larger than the WAT. Their argument was that 80% of the WGCTA multiple choice items measuring the critical thinking skills had only two distracters allowing respondents greater chances for guessing resulting in almost similar performances of both strong and weak thinkers.

Several researchers reported that the CCTST is more appropriate than the WGCTA as the CCTST had established concurrent validity with SAT-Verbal, and had significant correlation between CCTST and College GPA, which accounted for 41% of variance in CCTST. A few of them claimed CCTST was better in discriminating cohorts in their acquisition of CTS and found, although not statistically significant, CCTST assisted in measuring CTS of graduates from entry to exit in a program (May, Edell, Butel, Doughty & Longford, 1999; McCarthey, Schuster, Zehr & McDougal, 1999; O’Sulliven, Belvins-Stephens, Smith & VaghanWrobel, 1997; Saucier, Stevens & Williums, 2000; Shin, Jung, Shin, & Kim, 2006; Spelic, Parsons, Hercinger, Andrews, Parks & Norris, 2001; Stone, Davidson, Evans & Hansen, 2001; Tiwari, Lai, So & Yuen, 2005).

However, Bondy, Koenigseder, Ishee, and Williams (2001) reported that CCTST did not have sufficient psychometric properties as the factor analysis did not conform to the subscale structure, and they questioned the stability and reliability of the test. Further, Leppa (1997) reported that the CCTST demonstrated poor stability-reliability and placed a greater psychological burden on students and, as a result, chose to discontinue use of the CCTST with nursing students. Stein, Hynes and Understein (2003) assessed CTS of senior college students and reported that the CCTST had a low reliability, low item-total correlations, Principal Component Analyses (PCA) did not support item classification, and there were some indications of cultural bias. Although measuring CTS has been the focus of nursing education, and research, as well as nursing licensing and accrediting agencies (Facione & Facione, 1996; Oreman, 1997; Twibell, Ryan, & Hermiz, 2005), little research has been conducted to evaluate critical thinking as demonstrated by a lack of valid and reliable critical thinking instruments for assessing CTS in nursing (Simpson & Courtney, 2002).

Developing instruments to measure critical thinking is difficult. One difficulty is that critical thinking is a latent trait not amenable to direct observation, but inferred from a number of behaviours demonstrated by students. Another difficulty is that test developers venturing into measuring CTS are often confronted with issues related to theoretical underpinning, design, and practical aspects of developing an assessment strategy that captures the full scope of CTS. Critical thinking experts claim that self-assessment is an essential feature in developing CTS as the students go through a program of study (Paul & Elder, 2006).

Self-assessment is a process of self-directed monitoring that is initiated and driven by the individual and is intended for ongoing improvement (Galbraith, Hawkins, & Holmboe, 2008). Currently, many physicians use self-assessment for life-long learning, and it is now mandatory for the continuing competence among physicians. Medical professionals now show a renewed interest in self-assessment for enhancing their competencies thus ensuring quality care to clients (Austin, Gregory, & Galli, 2008). Research evidence reveals that many professionals are not proficient in self-assessment as they rarely practice it. A general trend is to use multiple-choice tests for measuring knowledge, which is easier, and attributes such as CTS, communication, interpersonal, and professional behaviours are difficult to measure and often unattended (Austin, Gregory & Chiu, 2008; Galbraith et al., 2008).

Critical thinking skills are developed over time, hence, the progression from unreflective through challenging, beginning, practicing, and advanced thinker to a master critical thinker can be monitored along the program of study. The experts claimed that CTS is fundamental to, if not essential for, a rational and democratic society (Facione, 1990). Thus, constant monitoring and ensuring the progressive development of this trait is essential. Ensuring the development of CTS is significant for all occupations in a rapidly changing society in the 21st century. Thus, there is a need for an easy and self-monitoring instrument to help students understand the cognitive process and self- regulate the progress of CTS.

A review of the literature on the existing instruments for measuring CTS revealed no self-assessment scale for assessing the development and progress of CTS. The existing standardized instruments do not provide an understanding of the processes and functions of cognitive skills by which the student is able to self-assess and to improve upon these skills. Further, the existing commercially available tests are not easily accessible and economical for students and student researchers. For example, the graduate student cost of getting a paper and pencil CCTST test for dissertation work (discounted) amounts to US $ 9/- per student when requesting 300 tests, and above all, the publisher retains the control of testing and analysis. This makes student retesting through the years of nursing study very problematic and supports the need for a self-assessment tool.

The researcher, with her lengthy experience in nursing education as a teacher and an administrator, has often tried several teaching methodologies with diverse student populations to enhance the students' ability to think analytically with information and evidence. The researcher faced many challenges in working with the students in accomplishing this task. As a passionate teacher, it is the researcher's desire to help students instil in themselves a habit of thinking by understanding the ontology and epistemology of CTS so that they enjoy thinking as a valuable asset. Thus, the researcher was motivated to develop an easy, accessible tool that could be used by students and teachers for improving their thinking.

In the current study, it was assumed that this instrument would be an educating and self-motivating force as this can create an awareness in the students about the cognitive skills they need to master in their thinking and its relevance to their learning. This Critical Thinking Self-Assessment Scale (CTSAS), if developed with scientific soundness in demonstrating evidence of validity and reliability, could be used as a tool to encourage self-development, not only by diagnosing how the learners think, but also helping them to understand how to take charge of their thinking, and to improve upon it. Thus, the study aimed to develop and evaluate preliminary psychometric characteristics of the CTSAS in measuring critical thinking skills of undergraduate nursing students.

Research Plan

Purpose of the study: The overall purpose of the study was to develop a sufficient number of items for CTSAS for the critical thinking constructs and evaluate its preliminary psychometric characteristics based on Messick's (1990) unitary concepts of construct validity.

Objectives. The objectives of the study were: (i) To develop sufficient number of items for the CTSAS based on the constructs identified in the conceptual framework; (ii) to establish content validity evidence for the CTSAS; (iii) to establish evidence of construct validity (structural validity), both internal and external validity, of the CTSAS; (iv) to establish reliability, both internal consistency and stability reliability, of the CTSAS; and (v) to establish convergent validity of CTSAS.

Design. The researcher used a staged non-experimental design for the development, validation, and evaluation of the CTSAS based on Messick's (1990) unitary concept of construct validity framework for measuring CTS. A non-experimental multi-staged design was considered appropriate because "most methodologic studies are non-experimental, often focusing on instrument development and testing", in this case, the development and initial testing of the CTSAS (Polit & Beck, 2008, p. 329).

Setting. For the content validity section, an expert panel was chosen from North America and India. For the other psychometric testing, the setting included both the Canadian and Indian perspective. The College of Nursing, University of Saskatchewan (U of S) was the first site of data collection. This included all the three campuses (Saskatoon, Regina and Prince Albert) in Saskatchewan. In India, several Colleges of Nursing in the state of Kerala formed the second site for data collection. The duration of the undergraduate study is four years in both countries. The researcher chose the two sites because both countries participate in accreditation processes where one of the main objectives is the demonstration of CTS in graduates. These sites were chosen because the researcher had links to both areas and was sure of the presence of only two curricula.

Population and sample. To achieve the content validity objective, the researcher selected a panel of 20 experts from different fields of discipline. The central aim was to include experts with experience in teaching critical thinking skills, researching in these areas, and their willingness to be a panel member. The choice of a multi-disciplinary approach followed the guide of the American Philosophical Association (APA) Delphi survey methodology. The experts included were from the fields of education, medical education, nursing, science, psychology, philosophy, kinesiology and theology. The panel members were involved in teaching inter-professional problem based learning (IPPBL), and critical thinking courses, and were experts who had experience in developing instruments from North America and India.

Two samples were used for the psychometric testing of the CTSAS, one from India and one from Canada. In Canada, the data were collected from 144 year third and fourth year students of University of Saskatchewan from the three sites. In India, 887 students representing all four years of study from five Colleges of Nursing participated in the data collection. The year three and year four students (251) of three colleges took a repeat test after two weeks. Refer to Table 1 for details of the sample distribution across the two countries. All participants who volunteered to participate were included in the study.

Table 1

Distribution of Participants by Country, College and Site

Canada (n=144)			India (n=887)		
Site	Enrolled	No. and % Participation	College	Enrolled	No. and % Participation
Prince Albert	63	44 (12.7)	CN*** -Aswini	146	140 (14.7)
Regina	95	48 (13.6)	CN -Dharmagiri	198	154 (16.2)
Saskatoon	195	52 (14.7)	CN –Kangazha	197	193 (20.3)
			CN –KVM	214	212 (22.3)
			CN –Pushpagiri	195	188 (20.5)
Total	353	144** (41%)		950	887* (94%)

*Absenteeism (6%)

**declined (59%)

***CN: College of Nursing

Data collection tools and techniques. The literature review for identifying the construct and definition of CTS started with the concept analysis of CTS as part of the doctoral course work in Nursing Science Philosophy. The actual item generation occurred as part of a course on Advanced Test Theory and Instrument Construction, which was completed during the researcher's doctoral course work. During the course work, the researcher generated a pool of preliminary items for the core six skills and the sub-skills using the definitions and indicators from the APA Delphi report (for a detailed description of construct definition refer to Section Two (Manuscript One) titled "A Conceptual Framework for Developing A Critical Thinking Self-Assessment Scale for Undergraduate Nursing Students").

In addition to the APA report, the researcher used other sources as inspiration for developing the preliminary items. First, the researcher had access and permission to use a rating scale that was developed in California School of Nursing for evaluating the nursing program. This scale provided an illustration of a format and nature of the items that could measure the core skills, but it only had face validity. Second, Rubenfeld and Scheffer`s THINK model, its descriptions and situations in the book-*Critical thinking in nursing: An integrative approach* (1999) provided additional information for writing the items for the various categories.

Above all, the keynote address delivered by Paul (2007) at the 27th International Conference on Excellence in Critical Thinking provided a wealth of information to derive the items for each scale. The detailed description of CTS, its elements, standards, and stage theory was very useful in generating the items. Writing items, reviewing and critiquing existing scales as part of the assignments and in-class group, and individual activities during the course on instrument development enhanced the researcher’s skill in writing the items for the scale (DeVellis, 2003; Frey, Petersen, Edwards, Pedrotti, & Peyton, 2005; Horan, DiStefano, & Motl, 2003).

Data collection

The data collection period extended from July 2010 to September 2010 in India and from October 2010 to December 2010 in Canada. The schedule for data collection is presented in Appendix A.

Data analysis

The analysis for establishing content validity included assessing and analysing the expert judges’ ratings on the items. The methodology and process of content validation is described in detail in Section Three (Manuscript Two), titled Content Validity of the CTSAS.

The analysis for the other psychometric testing comprised a series of steps involved in statistical testing, which included data screening, reliability, and data reduction strategies.

These steps include the following

1. Step 1: Exploratory Data Analysis for screening and cleaning data.

2. Step 2: Reliability analysis
3. Step 3: Exploratory Factor Analysis
4. Step 4: Confirmatory Factor Analysis
5. Step 5: Correlation statistics for establishing convergent validity of the scale.

Section four of the report presents the detailed description of the plan for each step, the statistical analysis, results, and interpretation of the findings for establishing construct validity.

Ethics

The researcher ensured human subject protection by obtaining ethical approval by the University of Saskatchewan Behavioural Research Ethics Board for both sections of the study (work with the experts and testing with the students). Approval for data collection was obtained from College of Nursing, University of Saskatchewan, Canada. Based on that approval, a copy of which was provided to them, five colleges in the state of Kerala, India provided their approval for data collection. The participation of content validation experts was voluntary and the researcher assured confidentiality of the information shared. The content validation package included a consent form to indicate their willingness to participate as a panel member.

The consent form for the student participants included a statement of assurance that participation in the study was voluntary and their decision to participate or refuse to participate would have no effect on their course grades. The consent form also included assurance of confidentiality of information provided by stating that the results will not reveal the participant's identity and will be reported on to as aggregate group data. The approval for the use of incentives to encourage participation (e.g. a draw for a bookstore gift certificate) was sought through the ethics application. (Refer to Appendix B and Appendix B1 for Certificates of Approval).

Theoretical Perspective of Construction of the CTSAS

The development of the Critical Thinking Self-Assessment Scale (CTSAS) for measuring this latent trait, CTS, assumes the logical positivist view that "these phenomena exist and they can influence behaviour, but they are intangible, and it may be appropriate to infer their existence from their behavioural consequences" (DeVellis, 2003, p. 9). The development of the CTSAS was based on Cronbach's vision of classical theory (Generalizability Theory), theory regarding the adequacy with which one can generalize from a sample of observations to a universe of observations from which it was randomly sampled (Brennan, 2006). According to Haertel, the assumption of classical theory is "Observed score $X_{pf,}$ obtained when *p* is administered a form of *f* test, *X* is the sum of a true score component and error component, i.e. $X = tp + E_{pf}$" (Brennan, 2006, p. 68). Using this theory, the steps of validating the CTSAS included: (i) testing the scale on the target population of undergraduate students, (ii) establishing reliability coefficient by Cronbach's alpha, (iii) testing the validity of the scale using data reduction strategies both Exploratory Factor Analysis and Confirmatory Factor Analysis.

Evaluation of Psychometric properties of the CTSAS

Evaluation of the psychometric properties of the CTSAS was based on Messick's Unitary Concept of Validity. Messick defined validity as an "integrated evaluative judgment of the degree to which evidence and theory support the adequacy and appropriateness of interpretations and actions based on the test scores" (Messick, 1990, p. 1). According to Messick, "the unitary concept of validity integrates content, criteria and consequences into a construct framework for the empirical testing of rational hypotheses about score meaning and theoretically relevant relationships" (1995, p. 741). Messick further described the six aspects of construct validity. These are:

i) “the content aspect of construct validity includes evidence of content relevance and representativeness and it is attained through expert judgment; ii) the substantive aspect refers to the theory and process modeling in identifying the content domain and the correlation patterns among part scores and response consistencies; iii) the structural aspect appraises the fidelity of the scoring structure to the structure of the content domain; iv) the generalizability aspect examines the extent to which score properties and interpretations are generalizable to the population groups, setting and tasks (criterion validity); v) the external aspect include the convergent and discriminant validity evidence; vi) the consequential aspect appraises the value implications of score interpretation and test use” (Messick, 1995, p. 745).

The researcher intended to accumulate evidence for the construct validity framework, which included content validity, structural validity and reliability. The external aspect of construct validity was targeted only for convergent validity. However, divergent validity, criterion validity and consequential validity were not considered at this stage of the study. The first step of assessing content validity is to define the construct on which to base the CTSAS. This process was almost completed before the actual research. The first manuscript entitled “A Conceptual Framework for developing CTSAS for Undergraduate Nursing Students” presents in detail the first step of the scale development (*determine what it is you want to measure)* and includes the literature review, construct definition, and a proposed conceptual framework on which to base the CTSAS.

SECTION 2

MANUSCRIPT 1

A CONCEPTUAL FRAMEWORK FOR DEVELOPING A CRITICAL THINKING SELF-ASSESSMENT SCALE FOR UNDERGRADUATE NURSING STUDENTS

Abstract

In today's health care realm, nurses must be talented critical thinkers to cope with the challenges in the ever changing health care system, changing population trends, and extended role expectations. Thus, several countries now recognize Critical Thinking Skills (CTS) as an expected outcome of nursing education programs. Critical thinking has been defined in multiple ways by philosophers, critical thinking experts, and educators. Nursing experts conceptualize critical thinking as a process involving cognitive and affective domains of reasoning. There are a plethora of definitions available in the literature. Nurse educators are often challenged with teaching and measuring CTS because of its latent nature and lack of a uniform definition of the concept. The purpose of this review of critical thinking literature is to examine various definitions and to identify a set of constructs that define critical thinking to derive a conceptual framework on which to base a self-assessment scale for measuring CTS.

Key words: critical thinking skills, conceptual framework, construct, self-assessment scale.

This manuscript is prepared according to the author guidelines of the *Journal of Nursing Education.*

Critical thinking skills (CTS) are required for nurses to cope with the ever-changing complex health care system of the 21st century. The ability to think critically is an expected outcome of undergraduate nursing education programs in several countries throughout the world (Canadian Association of Schools of Nursing (CASN), 2005; Commission on Collegiate Nursing Education (CCNE), 2003; National League of Nursing Accrediting Commission (NLNAC), 2004; Indian Nursing Council (INC), 2004). Programs in nursing education, therefore, mandate teaching and assessing learning outcomes that evidence acquisition of CTS in students.

During the past two decades, the nursing literature revealed a resurgence of developing learning settings that enhance students' and nurses' critical thinking skills (Carter & Rukholm, 2008; Daroszewski, Kinser, & Lloyd, 2004; Dickerson, Carter, & Walsh, 2008; Janicek, 2006; Kessler & Lund, 2004; Khosravani, Manoochehri, & Memarian, 2005; McGrath, 2003; McGrath, Hesketh, Lang & Estabrooks, 2003; Romeo, 2010; Staib, 2003). Nursing educators affirm that nursing education mandates deliberate actions for formal instruction in CTS and opportunities to practice these ways of thinking for effective learning. Due to a shift in nursing education from a medical model to a more holistic model students are required to reflect and integrate an array of determinants that impact human health and caring (Ibarreta & McLeod, 2004; Ironside, 2004; Simpson & Courtney, 2002) and are encouraged to move towards life-long learning (Romeo, 2010). Furthermore, nurses must be critical thinkers in order to effectively cope with advancing technologies, fiscal cutbacks, human resource limitations, and the increased acuity seen in many patient care situations (Carter & Rukholm, 2008). In addition, competent clinical practice demands reflective thinking beyond content knowledge (Cody, 2002); and the need for continuous interactive understanding of both the context of care and patients' experiences of wellness and illness (Ironside, 2003).

Teaching nurses and students to think critically requires mentors or teachers to be sound critical thinkers themselves. A study of critical thinking focused on current knowledge and practices among teachers of 27 teacher education colleges in California revealed inconsistent results in that the teachers (89%) claimed that CTS were the primary objective of their instruction, yet only 19% gave clear definitions of CTS (Paul, Elder, & Bartell, 1997; Paul, 2007). Cosgrove (2011), in a recent study, concluded that tutors who participated in the study were less focused on fostering essential CTS and dispositions such as intellectual analysis, intellectual evaluation and intellectual traits of mind. Nursing faculty in schools and colleges continuously strive to teach and develop this skill in students. The teaching-learning process involving an abstract concept or skill that is open to many interpretations mandates that students understand the ontology and epistemology of the concept and skill for effective learning. Thus, those teaching CTS need to equip themselves with the cognitive skills involved in critical thinking, and possess an understanding of the nature of learning opportunities and instructional designs that can foster CTS among students.

Nursing educators often face challenges in deriving appropriate teaching and assessment strategies to show evidence of the achievement of CTS. They must be able to see the learners' thinking by observing the mental processes and the logic (thinking skills) that students use in deriving theory based clinical decisions (Brunt, 2005). Direct evidence of the effect of pedagogical strategies for the development of CTS in students could be demonstrated by using outcome assessments, validated instrumentation, course assessments based on CTS, and assessing clinical judgment (Facione & Facione, 1994). However, such instruments and devices are largely unavailable and those that are available are not easily accessible or economical for use by students, teachers, and student researchers. Further, there are some concerns reported about their reliability and validity. They are also not helpful in self-educating students to monitor, self-regulate, and improve their critical thinking. Research using relevant instruments to measure teaching methodologies that enhance CTS or assess CTS for self-improvement is very limited. This review of critical thinking focuses on examining the existing instruments used for measuring critical thinking and the challenges in defining the concept for developing a valid and reliable instrument. The author proposes the need for a self-assessment scale and reviews various definitions and constructs to derive a conceptual framework for such a self assessment scale.

Literature Review: Current Thinking

Test development. Lack of appropriate instruments to measure CTS continues to pose challenges for nurse educators (Morrison & Free, 2001). Developing instruments to measure critical thinking is a challenge because critical thinking is a latent trait not amenable to direct observation, but must be inferred from a number of behaviours demonstrated by students. Thus, a valid and reliable instrument measuring CTS needs to address the "subtleties and sophistication of critical thinking required for content rich and highly contextualized" nature of the discipline of practice (Facione & Facione, 1996, p. 42). That is, the instrument needs to include the cognitive abilities involved in reflective thinking and demarcate the unreflective thinking that leads to biased, distorted, and prejudiced thinking (Paul, 2007).

Existing instruments. There are a number of standardized tests available to measure general CTS. Most of these instruments were developed during the eighties and nineties. Examples include the Watson Glaser Critical Thinking Appraisal (WGCTA) (Watson & Glaser, 1980), the Ennis-Weir Critical Thinking Essay Test (EWCTET) (Ennis & Weir, 1985), the Cornell Critical Thinking Test (CCTT) (Ennis, Milllman & Tomko, 1985), the California Critical

Thinking Skills test (CCTST) (Facione, Facione, Blohm & Gittens, 2008), and the International Critical Thinking Essay Test (Paul & Elder, 2007).

Adams, Whitlow, Stover and Johnson (1996) conducted a comprehensive review of studies using the WGCTA, CCTST, EWCTTE, and CCTT with nursing students, and reported inconsistent results in assessing CTS. The authors concluded that although WGCTA is an extensively used tool in nursing, findings are inconsistent in areas such as testing the development of CTS from program entry to exit, and CTS and clinical judgment. The major criticism was that the instruments are not discipline specific, and difficult to tap the CTS applied within a nursing context (Adams et al., 1996; Rubenfeld & Scheffer, 1999). More recently, Wagner and Harvey (2006) compared the WGCTA with the Wagner Assessment Test (WAT) using item response theory and found that standard errors of the WGCTA were 50% larger than with the WAT. Their argument was that 80% of the WGCTA multiple choice items measuring the skills had only two distracters allowing respondents greater chances for guessing, resulting in both strong and weak thinkers performing similarly.

A variety of studies since 1996 have discussed the CCTST and their findings indicated that the CCTST is more appropriate than the WGCTA in measuring critical thinking skills as the CCTST had established concurrent validity with SAT-Verbal, and had significant correlation between the CCTST and College GPA, which accounted for 41% of variance in the CCTST. A few of them claimed CCTST was better in discriminating cohorts in their acquisition of CTS and found, although not statistically significant, that the CCTST was helpful in measuring the CTS progress among graduates from entry to exit in a program (May, Edell, Butel, Doughty & Longford, 1999; McCarthey, Schuster, Zehr & McDougal, 1999; Saucier, Stevens & Williums, 2000; Shin, Jung, Shin, & Kim, 2006; Spelic, Parsons, Hercinger, Andrews, Parks & Norris, 2001; Stone, Davidson, Evans & Hansen, 2001; Tiwari, Lai, So & Yuen, 2006).

However, Bondy, Koenigseder, Ishee and Williams (2001) reported that the CCTST does not have sufficient psychometric properties as the factor analysis did not conform to the subscale structure, and they questioned the stability and reliability of the scale. Stein, Hynes, and Understein (2003) assessed CTS of senior college students and reported that CCTST had a low reliability, low item-total correlations, Principal Component Analysis did not support item classification, and there were some indications of cultural bias. Further, Leppa (1997) reported that the CCTST demonstrated poor stability-reliability and placed a greater psychological burden on students and, as a result, chose to discontinue use of the scale with nursing students. It is interesting to note that the CCTST has a counterpart test called the California Critical Thinking Disposition Inventory (CCTDI). Both Bondy et al. (2001) and Leppa (1997) found that the CCTDI was more reliable in usage.

Romeo (2010) analyzed recent quantitative research findings relevant to measuring critical thinking abilities and skills in undergraduate nursing students and critical thinking's role as a predictor of National Council Licensure Examination-Registered Nurse (NCLEX-RN) performance. The findings included that the majority of the instruments available to measure critical thinking skills and abilities are not sufficiently specific for use with nursing students. The most frequently used tool was the CCTST, "which yielded mixed findings in the reviewed studies" (p. 4). Although some of the studies showed predictor relationships of CTS with NCLEX-RN, Romeo found that these findings were not generalizable due to either the lack of scientific soundness in the conduct of the study or an undersized sample size. Romeo recommended further research to develop scientifically valid and reliable measures or additional testing of the existing tools with rigorous research methods.

Morrison and Free (2001) designed guidelines for content based multiple-choice test items using multi-logic thinking that promotes and measures CTS. Others have used alternative forms of evaluation such as concept mapping (Billings & Kowalski, 2008; Ellermann, Yahiro & Wong, 2006; King & Shell, 2002; Wilgis & McConnell, 2008); a visual analogue scale (Foley, 2008); portfolio writing (Facione & Facione, 1996); reflective reasoning (Carter & Rukholm, 2008; Dickerson, et al., 2008; Kessler & Lund, 2004); and group dynamics (Khosravani, et.al., 2005) to both foster and assess CTS. Since 2000, several nurse researchers have directed their attention towards assessing CTS through reflective processes (qualitative approaches) with specific references to assessing CTS through assignments, essays, and clinical practicum using rubrics (Baker, 2001; Brunt, 2005; Carter & Rukholm, 2008; Dickerson, et al., 2008; Gray, 2003; King & Shell, 2002; Morrison & Free, 2001; Sorensen & Yankech, 2008; Twibell, Ryan, & Hermiz, 2005; Wilgis & McConnell, 2008).

Self-assessment. No self-assessment tools for measuring CTS could be located. Although measuring CTS has been the focus of nursing education and research as well as nursing licensing and accrediting agencies (Facione & Facione, 1996; Twibell, et al., 2005), there has been a dearth of conclusive research on the evaluation of critical thinking as demonstrated by a lack of acceptable critical thinking instruments specific to nursing for assessing CTS. A literature review on CTS by Simpson and Courtney (2002) recommended the need for multiple devices for measuring CTS. Critical thinking experts have claimed that self-assessment is an essential feature in developing CTS as the students go through the program of study (Facione, 1990; Paul, et al., 1997; Paul & Elder 2006).

Self-assessment is a process of self-directed monitoring that is initiated and driven by the individual and is intended for ongoing improvement (Galbraith, Hawkins, & Holmboe, 2008). Self-regulation is an outcome of meta-cognition, that is, constant monitoring of one's own thinking, being knowledgeable about one's own thought process, accuracy in describing one's own thinking, control of one's own actions, beliefs and intuitions and a continuous improvement in thinking (Flavel, 1979). Currently, many physicians use self-assessment for life-long learning, and such self-assessments are now mandatory for the continuing competence among physicians, nurses and other health science professionals for ensuring quality care to clients (Austin, Gregory, & Galli, 2008; Bassendowski and Petrucka, 2009). Research evidence reveals that many professionals lack proficiency in self-assessment (Austin et.al, 2008). While it is much easier to use knowledge based multiple-choice tests, attributes such as CTS, communication, interpersonal, and professional behaviours are difficult to measure and therefore often unattended (Austin, Gregory & Chiu, 2008; Galbraith et al., 2008).

The intent of developing a Critical Thinking Self-Assessment Scale (CTSAS) is to help students enhance their meta-cognitive capabilities of objectively monitoring their own thinking and improving their CTS. The CTSAS further aims to infuse self-responsibility and accountability in students for self-regulating and strengthening their CTS. The author believed that CTS are universal and once mastered can enable one to think through a variety of situations regardless of one's discipline with specific reference to meeting the professional as well as the personal challenges that an individual may face in life. A general assumption is that fostering the development of CTS requires a clear understanding of the ontology of CTS and the epistemology of acquisition of this skill. Human capacity for reasoning can be nurtured and developed through educational programs. We assume that CTSAS will assist students in the engagement of effortful thinking that is reflective, reasonable and directed on to what to believe or do.

Identifying Constructs for the Conceptual Framework

Literature searches were conducted using CINAHL, Pub med, MEDLINE, ERIC databases and Google search from 1990 to 2010 on CTS for the purpose of identifying and defining the constructs for developing a conceptual framework on which to base the CTSAS. The key words used were critical thinking skills-definitions, constructs of CTS, measurement of CTS, CTS instrument development, self-assessment and psychometrics. A large number of research and non-research articles were retrieved, and those which examined definitions, concepts, constructs and measurement of CTS were targeted for detailed review in addition to the books written by nurse authors on critical thinking (Alfaro-LeFevre, 2009; Facione & Facione, 2008; Rubenfeld & Scheffer, 1999) and seminal articles and books by critical thinking experts (Facione, 1990; Facione & Facione, 1994; 2007; 2008; Paul, 1995; Paul & Elder, 2002; 2006).

Definitions and Concepts of Critical Thinking Skills

It is evident from the review that there are several ways to define CTS. For the National Council for Excellence in Critical Thinking, Scriven and Paul defined critical thinking as “the intellectually disciplined process of actively and skilfully conceptualizing, applying, analyzing, evaluating or synthesizing information gathered from observation, experience, reflection, reasoning, or communication, as a guide to belief and action” (cited in Paul et al., 1997, p. 4). Paul and Elder (2002) have further indicated that critical thinking is self-directed, self-disciplined, self-monitored and is self-corrective thinking. It entails effective communication and problem solving to overcome our naïve egocentrism - a tendency to think ourselves at the center of the world, and sociocentrism - a tendency to think within the confines of our groups. In contrast, nurse authors and educators have conceptualized CTS as: 1) an essential tool for sound clinical judgment in nursing (Facione, Facione & Sasnchez, 1994); 2) being able to respond to a problem using the nursing process (Young, 1998); 3) involving integrative thinking (Tanner, 2007); and 4) exhibiting habits of mind such as confidence, creativity, inquisitiveness, and open mindedness (Scheffer & Rubenfield, 2000).

Transferability of critical thinking skills. Experts debate the transferability of CTS from one setting or situation to another. Some argue that critical thinking proficiency can be general and applicable across subject matter disciplines, in a variety of situations and context, and a wide range of human activities (Paul. et al,, 1997); and that CTS can be tested within a general context (Ennis, 1989; Facione & Facione, 1994, 2008; Paul, & Noshich, 1993). Others have asserted, and some have changed their views, that critical thinking mastery improves if developed, and assessed within the context of the discipline (McPeck, 1990; Paul, 2007). McPeck (1990) stated that thinkers evaluate situations, issues, and events in light of background knowledge, context, and reflective scepticism.

Universal vs. subject-specific critical thinking. Bandman and Bandman (1995) claimed universality of critical thinking and viewed CTS as both subject-specific and general. The National Academy of Education (NAE) advocated for the development of subject-specific higher order thinking tests (Morrison & Free, 2001). Some of the nurse authors and educators argue that CTS is contextual, for example, nurses are involved in complex situations that require in-depth considerations of the workplace expectations, and family and client expectations; to effectively use specific knowledge that is vital to keeping patients safe and helping them promote their wellness demands deep and integrative thinking skills (Alfaro-LeFevre, 1995).

In contrast, Rubenfeld and Scheffer (1999) claimed that "critical thinking is the same in every discipline" (p. 376); CTS is the same but the task or content for thinking can be discipline specific (Rubenfeld & Scheffer). While there are these controversies, the authors believe that CTS is universal, and critical thinking is not a method to be learned but a process involving cognitive and affective domains of reasoning. Critical thinking has two dimensions: a frame of mind or a quest for thinking (disposition), and a set of operational cognitive skills (Facione, 1990). It involves self-examination or a meta-cognitive process in identifying the flaws in thinking, and continuously effecting change in improving the quality of thinking. A critical thinker is confident to suspend a decision or a claim advanced by another if it is not supported by a valid evidential reasoning, and seeks further information. A critical thinker is able to change his or her own ideas and opinions when warranted. Critical thinking is the awakening of the mind to discriminate right from wrong, and the exhibiting of intellectual courage and integrity in advocating for the right (Paul & Elder, 2006).

Domains of critical thinking skills

Critical thinking includes a set of skills and abilities for generating and processing information and beliefs (Paul & Nosich, 1993). Facione and Facione (2008) described CTS as an interactive, reflective, reasoning process of making a judgment about what to believe and what to do. Paul and Elder (2006) described four key domains of CTS: elements of thought, abilities, affective domain, and intellectual standards. Paul and colleagues (1997) further noted that "these four domains are interrelated and interdependent, functioning as complex skills, practices, dispositions, attitudes and values" (p. 1).

Domain of elements of thought. The ontology of elements of thought implicit in all reasoning become the building blocks of thinking and they shape reasoning and give logic to reason. These elements include purpose, goal or end in view, question at issue, frame of reference or point of view, information, empirical dimensions (interpretation and inference), and conceptual dimension (theories, axioms, laws, principles, assumptions, implications and consequences) (Paul & Elder, 2006).

Domain of abilities. The epistemology of CTS entails the critical thinker's approach to specific issues, questions, or problems. These abilities include: recognition of the problems, finding workable means for meeting those problems, gathering and marshalling pertinent information, and recognizing unstated assumptions and values. Additional useful abilities involve comprehension and the use of language with clarity, accuracy, and discrimination to interpret data, appraise evidence, evaluate arguments, and examine the logical relationships of presuppositions in order to draw conclusions and evaluate generalizations (Facione, 2007; Paul & Elder, 2006).

Affective domain. The affective domain of CTS characterizes the attitudes, dispositions, passions, and traits or habits of mind. These traits are: thinking independently, exercising fair-mindedness, developing intellectual humility, courage, faith, integrity, and perseverance, exploring the feeling underlying the thoughts, developing confidence in reason, and intellectual curiosity (Paul & Elder, 2002, 2006).

Domain of intellectual standards. The domain of intellectual standards includes the values and intellectual standards that impose discipline and restraint on the thinking. These values and standards, included in the National Council Statement, are "clarity, accuracy, precision, logic, fairness, relevance, consistency, breadth, depth, and comprehensive in thinking" (Paul & Elder, 2006, p. 43).

Critical Thinking and Motivation

An individual's motivation can vary the use of critical thinking. When an individual's thinking is non-substantive, and grounded in selfish motives, ideas tend to be manipulated for the sake of self or the group. Motivation can perpetrate a fraud or deliberately confuse or confound, and frustrate a group or project. Critical thinking, when substantive, and based on fair-mindedness and intellectual integrity, is an intellectually higher order. A person with a probing inquisitiveness, a keenness of mind, a zealous dedication to reason, and eagerness for reliable information possesses a *critical spirit*, or a disposition to think critically (Facione, 2007). In contrast, a person without a critical spirit may have low reasoning abilities, be closed minded, inflexible, insensitive, and unable to understand what others think or jump to conclusions. Critical thinkers are seldom influenced by cultural beliefs, religious tenets, social mores, or political orientations. Rather, they deliberately commit themselves to reason things with evidence, and seek information with objectivity, integrity and fair-mindedness (Giancarlo & Facione, 2001). Thus the trait of a critical spirit is a prerequisite to utilizing the cognitive skills of critical thinking.

Effective practices in thinking involve learners engaging in cognitive activities, more precisely, an arousal of the need for cognition. That is, the individual's "tendency to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). Individuals who are high in need for cognition demonstrate a quest for information, and think about and reflect on information to make a sense of the world around them. In contrast, those who are in low need for cognition depend on others, intuition, and need social comparison to arrive at meaning. For the former, meta-cognition becomes an automatic process to advance through the stages of thinking (Evans, Kirby & Fabrigar, 2003). Bassendowski and Petrucka (2009) studied registered nurses' (RN) perceptions of a continuing competence program (CCP) and reported that RNs with high internal locus of control were able to influence their professional activities such as the CCPs.

Social Cognitive Learning Theory and CTS.

The social cognitive learning theory (SCLT) explains how people acquire and maintain certain behavioural patterns (Bandura, 2001). The SCLT assumes a multitude of capabilities in an individual, and claims that a person is neither driven by inner forces nor automatically shaped by external stimuli. Rather, human functioning is a result of triadic reciprocal interaction among behaviour, cognitive capabilities and personal factors, and environmental events. Ontological perspectives of multitude of capabilities are symbolizing capability, forethought capability, vicarious capability, self-reflective capability, and self-regulating capability (Bandura, 1986).

Symbolizing capability helps learners process, internalize, and transform experiences to form new information and is a medium for thinking and communicating our thoughts, beliefs, and ideas. *Forethought capability* enables the development of a critical spirit, a disposition to using critical thinking. The freedom to think, self-accomplishments, and satisfaction enhance confidence. Fear, egocentric, and socio-centric thinking and cultural sanctions may inhibit the development of the critical spirit. *Vicarious capability* helps in observational learning and learning by experience. In order to make the observation more critical and meaningful, the observation should precede the analysis and interpretation of the situation or event, and extrapolation of the findings to derive decisions (Bandura, 1986).

Self-reflective capabilities enable people to analyze their own thought processes by reflecting on their experiences. Through self-reflection, people can evaluate, and alter their own thinking, and they can monitor their ideas and the resulting behaviour or predict behaviour. Finally, *self-regulating capability* assumes that people do not behave just to suit the preferences of others. Internal standards motivate and regulate much of the behaviour. Self-evaluation and self-directness wield influence over the external environment through spontaneity and independence. For example, an individual spontaneously calls upon the cognitive skills without any external prompt when he or she possesses a disposition to think productively and critically (Bandura, 1986). Personality traits such as open-mindedness, cognitive complexity, need for cognition, tolerance of ambiguity, and reflectiveness characterize an effective critical thinker (Kurt, 1995). Thus, facilitative environmental conditions, effective cognitive guides, and incentives for their own efforts are essential to accomplish self-regulation (Bandura, 1986).

The ontological perspective of cognitive capabilities indicates the antecedents of the domains of critical thinking. The domains of linguistic and reasoning ability, elements of thoughts, traits of mind, and the intellectual standards, become the decisive outcomes of the reciprocal interaction of the triadic components referred to in social cognitive learning. These cognitive skills and dispositional attributes of CTS are core to nursing and they epitomize a search for best knowledge in a given context. Critical thinking skills open doors to new perspectives about the world, foster self-confidence, and encourage life-long learning.

An analysis of definitions of CT reveals that CT is not a method to be learned but a process involving cognitive and affective domains of reasoning. The differing definitions of critical thinking have posed constraints to the development and assessment of CTS due to the lack of a uniform and clear understanding of the concept. In 1987, The American Philosophical Association (APA) sponsored a project, and appointed Peter Facione, a philosopher and writer in the field of critical thinking, to head a systematic inquiry on the status of critical thinking and critical thinking assessment. During the qualitative Delphi research study (February 1988 – November 1989), 46 experts from the fields of education, philosophy, social sciences, and physical sciences rigorously engaged in six rounds of responses and deliberations to questions from the chair, and arguments, comments, agreements and disagreements to the statements synthesized on each round from those responses. The final result was a majority consensus (87-95%) on the six core cognitive skills of critical thinking - interpretation, analysis, evaluation, inference, explanation and self-regulation. The APA Delphi survey final outcome was a consensus definition on critical thinking skills with directions for curriculum development, instruction, and assessment (Facione, 1990).

APA Definition of Critical Thinking Skills

The APA consensus definition of critical thinking is seen as a "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as an explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which judgment is based" (Facione, 1990, p. 2). Critical thinking is an essential tool of inquiry. An ideal thinker is characterized not merely by her or his cognitive skills, but also by how she or he approaches life and living. The approaches to life and living that characterize critical thinking include:

> Inquisitiveness with regard to a wide range of issues, concern to become and remain well-informed, alertness to opportunities to use critical thinking, trust in the processes of reasoned inquiry, self-confidence in one's own abilities to reason, open-mindedness regarding divergent world views, flexibility in considering alternatives and opinions, understanding of the opinions of other people, fair-mindedness in appraising reasoning, honesty in facing one's own biases, prejudices, stereotypes, or egocentric tendencies, prudence in suspending, making or altering judgments, willingness to reconsider and revise views where honest reflection suggests that change is warranted (Facione, 2007,

p. 9).

According to APA, the six core cognitive skills of CTS are Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-regulation. The definitions of core skills are presented in Box1. <Insert Box1 here> (Appendix C). Each of these core skills is further classified with two to three sub-skills. The core skills and corresponding sub-skills are presented in Box 2 (Facione, 1990). <Insert Box 2 about here> (Appendix D).

Reasons for Selecting the APA Definition for the Scale

The APA document clearly explicates the critical thinking core constructs and the sub-skills of each constructs with criteria and indicators for teaching and assessment. Paul and Elder purported similar concepts in their writing and emphasized a variety of pedagogical strategies for developing CTS (Paul, 2009; Paul & Elder, 2002, 2006). The APA recommendations included assessment criteria and indicators from which items could be generated for the self-assessment scale. Thus, the theory and philosophical concepts of CTS have been thoroughly discussed and debated over the decades to reach valid constructs for measuring CTS.

Conceptual Framework

Psychometricians warrant that instrument developers report the information on the process of instrument development to maintain rigour and the scientific soundness of scale development. The authors have developed this conceptual framework as a first essential step in the instrument development process and to test the structural validity. The proposed conceptual framework adopts the APA consensus definition of critical thinking to define the construct.

Critical thinking is an essential tool of inquiry. Beyond the cognitive skills, the definition emphasises the intellectual standards implicit in using these skills, and the prerequisite need for cognition or the disposition to use the cognitive skills. The experts on critical thinking also examined the definition’s application in a variety of situations. These included the areas of life and learning, with reading and writing, or all programs rich with discipline-specific content (Facione, 1990). Nursing has a rich content base for practice and students need to think through the content and understand and reflect on its application in patient care situations.

The execution of core skills requires a blend of operations of the sub-skills for its useful application in situations demanding assessing and analyzing events and information to arrive at accurate inferences. Although the sub-skills are arbitrary differentiations, when in use, some skills may influence the development of others in different situations. For example, the individual demonstrates interdependency of the core and sub-skills in a unique manner when critical thinking is used in situations such as effective communication, problem solving or decision making to draw valid evidential reasoning for an inference.

The core cognitive skills are conceptualized as interrelated and interdependent. Attainment of the each of the core skill is in relation to the others, and depends on the development of the others. Although stages of CTS development appear to be hierarchical, the pattern of development of its elements (issues, premises, claims, points of view, and opinions), cognitive traits and intellectual standards are interdependent and interrelated. For example, the development of one’s ability to interpret a piece of information or data requires accuracy, precision, and the evidential support for a valid reasoning to draw inference.

The schematic presentation created by Facione of the conceptual framework is illustrated in Figure 2.1 with the core cognitive skills (Facione, 1990), and Figure 2. 2 shows the researcher created schematic of the core cognitive skills and their sub-skills. Each of the sub-skill definitions purported will be effectively used for generating the items for the self assessment scale. The interdependent relationship among the core and sub-skills is hypothesized to emerge into a structural pattern from the observed scores when administering the scale on a designated population (undergraduate nursing students) (Appendix E & F). (Insert Figure 2.1 and 2. 2 here)

The development of critical thinking skills is a gradual process. The need for cognition or a mental habit to engage in effortful thinking and enjoy thinking is a vital prerequisite for developing and advancing towards higher levels of critical thinking and learning (Cacioppo & Petty, 1982; Culhane, Morera, & Watson, 2006; Evans et al., 2003). Changing one's habit of thinking happens over the years. In addition to the definitions and concepts proposed by Paul and Elder (2002, 2006, 2007), and Paul and Nosich (1993), which supported the APA definition and framework, their stage theory (Paul & Elder, 2006) explicates the six stages of the developmental hierarchy of CTS from unreflective thinker through challenged, beginning, practicing, advanced and master thinker. The stage theory demonstrates a linear relationship progressing from stage one to six intertwining with the six core cognitive skills accomplishment at each stage. The core cognitive skills exhibit interrelationship and interdependency at each of the six stages which can be monitored across a program of study that targets the outcome of CTS.

Conclusion

Critical thinking is essential to all walks of life. Critical thinking is in every discipline and the discipline, context or situation forms the content for thinking. Experts agree that critical thinking is vital for survival in the complex system of functioning in every discipline and profession; in particular, health sciences professions, as this involve dealing with complex human issues and concerns. Understanding of the philosophical underpinnings of the concept of CTS is imperative for those who intend to develop, sustain and improve one's thinking skills.

Several studies have reported inconsistent results with the CCTST, which was developed based on the APA definition of CTS. Despite the inconsistencies and disagreements within the research literature, the APA definition of CTS was a product of rigorous scientific process. The present research, using a different path to develop a measure of CTS, aims to examine if the items generated based on APA definition hold valid in defining the CTS construct. While the CCTST based on the APA consensus definition yielded inconsistent results upon measurement in various situations, we believe the addition of Paul and Elder's (2006) stages of CTS and rigorous development of the CTSAS can result in an instrument that students and faculty can use.

We hold that students can be trained to develop CTS and that the development and progression of this skill can be measured. We believe this self-assessment scale will motivate and educate students to develop and improve this skill through understanding the ontology and epistemology as well as the reasons why this is valuable for their learning. The authors are hopeful that such an instrument will assist undergraduate nursing students to self-monitor their thinking with the aim of improvement. Through this and other elements of undergraduate nursing education, graduates with strong CTS will be able to utilize and appropriately apply the knowledge and information to their work in maximizing the health of global populations.

References

Adams, M. H., Whitlow, J. F., Stover, L. M., & Johnson, K. W. (1996). Critical thinking as an educational outcome: An evaluation of current tools of measurement. *Nurse Educator, 21*(3), 23-32.

Alfaro-LeFevre, R. (2009). *Critical thinking and clinical judgment* (4th ed.). St. Louise: Elsevier, Sounders.

Austin, Z., Gregory, P. A. M., & Chiu, S. (2008). Use of reflection-in-action and self- assessment to promote critical thinking among pharmacy students. *American Journal of Pharmaceutical Education, 72(3),* 1-8.

Austin, Z., Gregory, P. A. M., & Galli, M. (2008). "I just don't know what I'm supposed to know": Evaluating self-assessment skills of international pharmacy graduates in Canada. *Research in Social and Administrative Pharmacy, 4,* 115-124.

Baker, D. (2001). Nursing reasoning model. *Nurse Educator, 26(5),* 203-204.

Bandman, E. L., & Bandman, B. (1995). *Critical Thinking in Nursing*. Norwalk, CN: Appleton & Lange.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood: Prentice Hall, Inc.

Bandura, A. (2001). Social cognitive theory: An agentive perspective. *Annual Review of Psychology, 52,* 1-26.

Bassandowski, S. & Petrucka, P. (2009). Perceptions of select registered nurses of the continuing competence program of the Saskatchewan Registered Nurses` Association. *Journal of Continuing Education in Nursing, 40*(12), 553-559.

Billings, D. M., & Kowalski, K. (2008). Augment mapping. *The Journal of Continuing Education in Nursing, 39(6),* 246-247.

Bondy, K. N., Koenigseder, L. A., Ishee, J. H., & Williams, B. G. (2001). Psychometric properties of California Critical Thinking Tests. *Journal of Nursing measurement*, *9*(3), 309-328.

Brunt, B. A. (2005). Models, measurement, and strategies in developing critical thinking skills. *The Journal of Continuing Education in Nursing*, *36*(6), 255-262.

Cacioppo, J. T. & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*(1), 116-131.

Canadian Association of School of Nursing (2005). CASN Accreditation-recognition of excellence. Retrieved from www.casn.ca

Carter, L. M., & Rukholm, E. (2008). A study of critical thinking, teacher-student interaction, and discipline-specific writing in an online educational setting for registered nurses. *The Journal of Continuing Education in Nursing*, *39*(3), 133-138.

Cody, W., K. (2002). Critical thinking and nursing science: Judgment or Vision? *Nursing Science Quarterly*, *15*, 184-189.

Collucciello, M. L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students-a conceptual model for evaluation. *Journal of Professional Nursing*, *13* (4), 236-245.

Commission on Collegiate Nursing Education. (2003). *Standards for accreditation of baccalaureate and graduate nursing programs* Retrieved from www.aacn.nche.edu/Accreditation/NEW_STANDARDS.htm

Cosgrove, R (2011). Critical thinking in the Oxford tutorial: a call for an explicit and systematic approach. *Higher Education Research & Development*, *30*(3), 343-356.

Culhane, S. E., Morera, O. F., & Watson, P. J. (2006). The Assessment of factorial invariance in need for cognition using Hispanic and Anglo samples. *The Journal of Psychology, 140*(1), 53-67.

Daroszewski, E. B., Kinser, A. G., & Loyd, S. L. (2004) Online, directed journaling in community health advanced practice nursing clinical education. *Journal of Nursing Education*, *43*, 175-180.

Dickieson, P., Carter, L. M., & Walsh, M. (2008). Integrative thinking and learning in undergraduate nursing education: Three strategies. *International Journal of Nursing Education Scholarship. 5*(1), 1-13.

Ellermann, C. R., Yahiro, M. R. K., & Wong, S. (2006) Logic models used to enhance critical thinking. *Journal of Nursing Education, 45*(6), 220-227.

Ennis, R.H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18,* 4-10.

Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test.* Pacific grove, CA: Midwest publications.

Ennis, R. H., Milllman, J., & Tomko, T. N. (1985). *Cornell critical thinking test level X and level Z manual.* 3rd. ed. Pacific grove, CA: Midwest publications.

Evans, C. J., Kirby, J. R., & Fabrigar, L. R. (2003). Approaches to learning, need for cognition, and strategic flexibility among university students. *British Journal of Educational Psychology, 73,* 507-528.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction, Executive summary, The Delphi report. American Philosophical Association, California Academic Press. Retrieved from http://www.isigntassessment.com/article.html

Facione, P. A. (2007). Critical Thinking what it is and why it counts. Update 2007, Retrieved from http://www.insightassessment.com/articles.html

Facione P. A., & Facione, N. (1994, August). *Class Room Assessment of Higher Order Thinking.* Paper presented at the 14th Annual International Conference on Critical Thinking and Educational Reform, Sonoma, CA.

Facione, N. C., & Facione, P. A. (1996). Assessment design issues for evaluating critical thinking in nursing. *Holistic Nursing Practice, 10*(3), 41-53.

Facione, N. C., & Facione, P. A. (2008). *Critical thinking and clinical reasoning in the health sciences: An international multidisciplinary teaching anthology.* Millbrae, CA: The California Academic Press.

Facione P. A., Facione, N. C., & Sanchez C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of California Critical Thinking Disposition Inventory. *Journal of Nursing Education, 33*, 345-350.

Facione, N. C., Facione, P., Blohm, S. W., & Gittens, C. A. (2008). California critical thinking skills test manual. CA: Insight assessment, The California Academic Press.

Flavel, J. H. (1979). Meta-cognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, *34*, 906-911.

Foley, D. K. (2008). Development of a visual analogue scale to measure curriculum outcome. *Journal of Nursing Education, 47*(5), 209-213.

Galbraith, R. M., Hawkins, R. E., & Holmboe, E. S. (2008). Making self-assessment more effective. *Journal of Continuing Education in the Health Professions, 28(1),* 20-24.

Giancarlo, C. A., & Facione, P. A. (2001). A look across four years at the disposition toward critical thinking among under graduate students. *The Journal of General Education, 50* (1), 29-55.

Gray, M. T. (2003). Beyond content: Generating critical thinking in the classroom. *Nurse Educator, 28(3),* 136-140.

Ibarreta, G. I., & McLeod, L. (2004). Thinking aloud on paper: An experience in journal writing. *Journal of Nursing Education, 43*(3), 134-137.

Indian Nursing Council. (2004). Syllabus: Basic B.Sc. Nursing. New Delhi: INC.

Ironside, P. M. (2003). New pedagogies for teaching thinking: The Lived experiences of students and teachers enacting narrative pedagogy. *Journal of Nursing Education, 42*(11), 509-516.

Ironside, P. M. (2004). Covering content and teaching thinking: Deconstructing the additive curriculum. *Journal of Nursing Education, 43*(1), 5-12.

Janicek, M. (2006). How to read, understand, and write discussions section in medical articles. An exercise in critical thinking. *Medical Science Monitor, 12*(6), 28-36.

Johns, C. (1995). Framing learning through reflection with Carper's fundamental ways of knowing in nursing. *Advanced Nursing Practice, 22*(2), 226-234.

Kessler, P., & Lund, C. (2004). Reflective journaling: Developing online journal for distance education. *Nurse Educator, 29*(1), 20-24.

Khosravani, S. Manoochehri, H., & Memarian, R. (2005). Developing critical thinking skills in nursing students by group dynamics. *The Internet Journal of Advanced Nursing Practice, 7* (2), Retrieved from http://www.ispub.com/ostia/index.php?xmlFilePath=Journals/ijanp/vol7n2/skills/xml

King, M. & Shell, R. (2002). Critical thinking strategies: Teaching and evaluating critical thinking with concept maps. *Nurse Educator, 27(5)*, 214-216.

Kurt, T. (1995). *Critical thinking ability and disposition as factors of performance on a written critical thinking test.* Paper presented at the Annual meeting of the American Educational research Association. San Francisco, CA. Retrieved from http://www.eric.ed.gov/ERICWebPortal/Home.portal;jsessionid

Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with California Critical Thinking Tests. *Nurse Educator, 22* (5), 29-33.

May, B. A., Edell, V., Butell, S., Doughty, J., & Longford, C. (1999). Critical thinking and clinical competence: A study of their relationship in BSN seniors. *Journal of Nursing Education*, *38,* 100-110.

McCarthey, P., Schuster, P., Zehr, P., & McDougal, D. (1999). Evaluation of critical thinking in a baccalaureate nursing program. *Journal of Nursing Education*, *38* (3), 142-144.

McGrath, P. J. (2003). Issues and innovations in nursing education. *Journal of Advanced Nursing, 43* (6), 569-577.

McGrath, J. P., Hesketh, K. L., Lang, S., & Estabrooks, C. A. (2003). A study of critical thinking and research utilization among nurses. *Western Journal of Nursing Research, 25* (3), 322-337.

McPeck, J. E. (1990). Critical thinking and subject specificity. A reply to Ennis. *Educational Researcher, 4*(2), 10-12.

Morrison, S., & Free, K. W. (2001). Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education*, *40,* 17-24.

National League for Nursing Accreditation Commission, Inc. (2004, revised 2008). *Accreditation manual with interpretive guidelines by program type for post secondary and higher degree programs in nursing.* Retrieved from http://www.nlnac.org/

Paul, R.W. (1995). *Critical Thinking: How to prepare students for rapidly changing world.* Santa Rosa, CA: Foundations for Critical Thinking.

Paul, R. W. (2007). *Critical thinking in every domain of knowledge and belief.* Keynote address at the 27th international conference on critical thinking, Berkeley, CA. Retrieved from http://www.criticalthinking.org

Paul, R. W. (2009, July). *Exploring the relationship between content of thinking in fostering intellectual discipline.* Paper presented at the 29th international conference on critical thinking, Berkeley, CA. Retrieved from http://www.criticalthinking.org

Paul, R. W., & Elder, L. (2002). *Critical thinking: Tools for taking charge of your professional and personal life*. New Jersey: FT Press.

Paul, R. W., & Elder, L. (2006). *Critical thinking: Learn the tools the best thinkers use* (Con. ed.). Pearson, NJ: Prentice Hall.

Paul, R., & Elder, L (2007). International critical thinking essay test. CA: The Foundation for Critical Thinking.

Paul, R., & Nosich, G. M. (1993). A model for the national assessment on higher order thinking. Retrieved from http://www.criticalthinking.org/a-model-nal-assessment-hot.cfm

Paul, R. W, Elder, L., & Bartell, T. (1997). California teacher preparation guide for instruction in critical thinking: Research findings and policy recommendations. State of California: California Commission on Teachers Credentialing.

Romeo, E. M. (2010). Quantitative research on critical thinking and predicting nursing students'NCLEX-RN performance. *Journal of Nursing education, 20*(10), 1-9.

Rubenfeld, M. G., & Scheffer, B. K. (1999). *Critical thinking in nursing: An integrative approach.* (2nd ed.). Philadelphia: Lippincott.

Saucier, B. L., Stevens K. R., & Williams, G. B. (2000). Critical thinking outcome of computer assisted instruction versus nursing process. *Nursing and Health Care Perspectives, 21*, 240-246.

Scheffer, B. K., & Rubenfield, M. G. (2000). A consensus statement on critical thinking in nursing. *Journal of nursing Education, 39*(8), 352-359.

Shin, K., Jung, D., Shin, S., & Kim, M. (2006). Critical thinking disposition and skills of senior nursing students in associate, baccalaureate, and RN to BSN programs. *Journal of Nursing Education, 45*(6), 233-237.

Simpson, E., & Courtney, M. (2002). Critical thinking in nursing education: Literature review. *International Journal of Nursing Practice, 8*(2), 89-98.

Sorensen, H. A. J., & Yankech, L. R. (2008). Precepting in the fast line: Improving critical thinking in new graduate nurses. *The Journal of Continuing Education in Nursing, 39(5),* 208-216.

Spelic, S. S., Parsons, M., Hercinger, M., Andrews, A., Parks, J., & Norris, J. (2001). Evaluation of critical thinking outcomes of a BSN program. *Holistic Nursing Practice, 15*(3), 27-34.

Staib, S. (2003). Teaching and measuring critical thinking. *Journal of Nursing Education, 42*(11), 498-508.

Stein, B., Haynes, A., & Understein, J. (December 6-9, 2003). “Assessing Critical Thinking.” Paper presented at SACS Annual meeting in Nashville, Tennessee. Retrieved from http://iweb.tntech.edu.sti/SACS%20presentations%20paper.pdf

Stone, C. A., Davidson, L. J., Evans, J. L., & Hansen, M. A. (2001). Validity evidence for using a general critical thinking test to measure nursing students’ critical thinking. *Holistic Nursing Practice, 15*(4), 65-74.

Tanner, C. (2007). Connecting the dots: What’s all the buzz about integrative teaching? *Journal of Nursing Education, 46,* 531-532.

Tiwari, A., Lai, P., So, M., & Yuen, K. (2005). A comparison of effect of problem based learning and lecturing on the development of students’ critical thinking. *Medical Education, 40* (6), 547-554.

Twibell, R., Ryan, M., & Hermiz, M. (2005). Faculty perception of critical thinking in student clinical experience. *Journal of Nursing Education, 44*(2), 71-79.

Wagner, T. A. & Harvey, R. J. (2006). Development of a new critical thinking test using item response theory. *Psychological Assessment, 18*(1), 100-105.

Watson, G., & Glaser, E. M. (1980). Watson-Glaser critical thinking appraisal manual. Clevland: Psychological Corp.

Wilgis, M., & McConnell, J. (2008). Concept mapping: An educational strategy to improve graduate nurses’ critical thinking skills during hospital orientation program. *The Journal of Continuing Education in Nursing, 39(3),* 119-126.

Young, D. (1998). The importance of critical thinking (Letter to the editor). *The Journal of Nursing Education, 37*(4), 153.

SECTION 3

MANUSCRIPT 2

CONTENT VALIDATION OF THE CTSAS

Abstract

Critical thinking is an essential skill for nurses to effectively address client care in today's changing health care system. Ensuring the development of critical thinking skills (CTS) and assessing nursing graduates' acquisition of this skill is a mandate of all nursing curricula. No easy and economical assessment instrument is available for students to self-monitor and improve their thinking skills. The aim of this study was to develop a psychometrically sound self-assessment scale for monitoring and improving critical thinking skills of undergraduate nursing students. The design chosen was scale construction. The consensus definition of CTS purported by the American Philosophical Association was adapted to generate a pool of items for each of the six core cognitive skills (Facione, 1990). A total of 196 items were generated for the16 sub skills which form the six core cognitive skills. Experts' ratings on the items were analyzed for item ambiguity, Content Validity Index (I-*CVI)* and Aiken's Content Validity Coefficient (VI_k). One hundred and fifteen (115) items with a range of value $I\text{-}CVI \geq .7$ to .938 and range of VI_k value .69 to .95 significant at $p<.05$ were retained. Some of the items were modified based on the comments provided by the expert judging panel. These 115 items were further tested on undergraduate nursing students from India and Canada in order to establish construct validity. The scope of this paper is to describe the process and results of generating evidence for content validity of the Critical Thinking Self-Assessment Scale (CTSAS).

Key words: critical thinking skills, content validity, construct validity

This manuscript is prepared according to the author's guidelines for the *Journal of Nursing Measurement*.

Content Validation of the Critical Thinking Self-Assessment Scale

Instrument or scale construction is a thriving activity. Measurement scales quantify phenomena of interest and thus shed new light on the phenomena. Theories play a key role in conceptualizing the phenomena. Measuring elusive, intangible phenomena or latent traits from multiple evolving theories has been a challenge to researchers (DeVellis, 2003). In spite of having a range of data gathering sources and data analysis techniques, the information on the process and procedures used in establishing the reliability and validity of many of the scales reported in the literature is unavailable, insufficient or inappropriately reported by the scale development researchers (Aiken, 1985). Expert judges' rating on the degree of match between the items and construct definitions is a crucial phase in the scale construction. However, Messick (1989) remarked that documentation of systematic assessment of item ratings provided by the experts is seldom included in the literature. This paper describes the process and statistical procedures used to establish content validity evidence for the Critical Thinking Self-Assessment Scale (CTSAS).

Theoretical Perspectives

The acquisition and enhancement of critical thinking skills (CTS) are expected outcomes of undergraduate nursing education programs in several countries. The disposition to think critically is a latent trait upon which the cognitive skills can be developed. This development of CTS in students has gained a substantial momentum in nursing education during the last two decades as seen by the changes in learning settings and strategies that foster students' and nurses' CTS (Carter & Rukholm, 2008). Nurse educators and researchers have used a variety of assessment strategies and research methods, both quantitative and qualitative designs, for measuring the use of CTS. However, educators are often challenged with the measurement of

CTS due to lack of assessment instruments and strategies available or those that are available are not easily accessible and economical.

Instruments or scales are a “collection of items combined into a composite score, intended to reveal levels of theoretical variables not readily observable by direct means” (DeVellis, 2003, p. 8-9). In this study, the researcher developed a self-assessment scale for measuring the levels of Critical Thinking Skills (CTS) of undergraduate nursing students. The development of a scale such as the CTSAS assumes the logical positivist view that “these phenomena exist and they can influence behaviour, but they are intangible, it may be appropriate to infer their existence from their behavioural consequences” (DeVellis, p. 9). The study was based on the assumption that a latent variable such as CTS could be easily inferred from an overt behaviour of its user. However, the observations of the behaviour do not indicate the internal processes on the use of cognitive skills that direct the behaviour. For example, according to Bandura’s (1986) cognitive learning theory, the symbolizing cognitive capability helps the learner process, internalize and transform experiences into new information or arrive at inferences and the overt behaviour that can be observed from that processes are the symbols or syntaxes used for communicating those thoughts and inferences.

Development of scientific scales mandates scale developers to provide adequate evidence of scale validity and reliability. According to Messick (1990), content validity is one of the six aspects of the unified concept of construct validity. The unified concept of construct validity subsumes all categories of validity including such as measures of content validity, predictive validity, concurrent validity, discriminant and convergent validity, criterion related validity, and factor structure and provides the evidence about the construct validity in an assessment instrument. Content validity is an important component of construct validity because it provides evidence about "the degree to which the elements of the assessment are relevant to and representative of the targeted construct" (Haynes, Richard, & Kubany, 1995, p. 238).

Content validation (CV) is a multi-method, quantitative, and qualitative process. The purpose is to minimize the potential error variance and increase support for construct validity (Devellis, 2003). Content validity greatly depends on how precisely the construct is defined, the degree to which the experts agree about the domains and facets of the construct, and how well the experts matched the items to the domain definition. Content validity is important when using aggregated scores for finding factor structure because the aggregate variable is a combination of multiple measures. Thus, the components of an aggregate variable should be relevant and representative of the aggregate construct (Haynes, et al., 1995). The critical thinking construct is a latent trait and is composed of a set of cognitive skills that are unobservable variables, which are inferred from behaviour, and hypothesized to explain the covariance between observed results.

Content Validation of the CTSAS

According to DeVellis (2003), construction of a scale involves a series of systematic logical steps. These are: (1) determine what it is you want to measure, (2) generate an item pool, (3) determine the format for measurement, (4) review of the items by expert judges, (5) include validation items, (6) administer items to a development sample, (7) evaluate the items, and (8) optimize the scale. The scope of this paper is to report on the process of generating adequate evidence for content validity of the CTSAS. Evaluating expert judges' ratings remains the key criteria in making judgments regarding the items for the scale. The following section describes each of DeVellis's steps in developing the CTSAS.

Determine What It is You Want to Measure

A clear, in depth review of the theoretical literature related to the phenomenon of concern is valuable in conceptualizing the related constructs and deciding on the boundaries of the construct (DeVellis, 2003; Polit & Beck, 2008). The success of observing true co-variance between the variables or items is dependent on the ability to accurately conceptualize the unobservable construct (Hinkin, 1995), in this case critical thinking skills (CTS). The constructs are synthesized variables. The CTS constructs were adopted from the consensus definition of CTS purported by the American Philosophical Association (Facione, 1990). These cognitive skills and sub-skills of CTS are the result of a lengthy, six round Delphi survey involving a diverse group of experts from various disciplines.

Construct definition. According to the APA consensus definition, critical thinking is seen as a "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which judgment is based" (Facione, 1990, p. 2). Critical thinking (CT) is an essential tool of inquiry. The experts' consensus on critical thinking included its application in a variety of situations. These situations include: in areas of life and learning, with reading and writing, and could be effectively applied in all programs rich with discipline- specific content (Facione, 1990). Practicing professionals, in particular, health care professionals rely on CT to effectively use their knowledge of the field and experience to come to purposeful self-regulatory judgments for dealing with concerns in clinical, leadership, communication, legal, ethical, economic, policy, or strategic design.

Nursing, as a practice profession, has a rich content base for practice and students and practitioners need to think through the content and understand and reflect on its application in patient care situations. Thus, cognitive skills form an essential trait of students for their learning and future practice. The execution of core skills requires a blend of operations of sub skills for their useful application in situations demanding examination and analysis of events and information to arrive at accurate inferences. Critical thinking is “not a linear or step by step process” (Facione, 2007, p.7). The sub skills are arbitrary classifications. When an individual is engaged in intellectual activities these sub skills act interdependently in situations, such as decision making, interpreting a piece of information for use as evidence, or for arriving at solutions to problems. In the case of solving a problem alone or in a group, CT enables individuals to step back and reflect on the quality of their thinking and the judgments made to arrive at solutions. Reflexivity permits people to use their cognitive ability to monitor, correct, and improve their process of reaching a reasonable judgment. Carter and Rukholm (2008) in their study demonstrated development of CTS in students using John’s structured reflection model based on Carper’s patterns of knowing in discipline-specific writing facilitated by teacher student interaction. Integrative thinking and learning was facilitated through reflective writing, scenario testing and Objective Structured Clinical Evaluations (OSCE) (Dickieson, Carter & Walsh, 2008). Austin, Gregory and Chiu (2008) found that reflection-in-action and self-assessment contributed positively to improvement of CTS among pharmacy students. The researcher believes that CT is not discipline specific, however, she recognizes that individual research studies are often discipline specific in nature.

The core constructs of CT are the cognitive skills Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-regulation. These skills are employed interactively in the reflective reasoning process of making a judgment about what to believe and what to do (Facione, 2007). Each of these six skills is further classified with two to three sub-skills, (Facione, 1990). Table 3.1 outlines the classification of core cognitive skills and sub skills. Refer to Appendix G for definitions of each of the core cognitive skills and sub-skills as described by Facione (1990).

Table 3.1

Core Cognitive Skills and Sub-skills

Core Cognitive Skills	Sub-skills
Interpretation:	Categorization, Decoding Significance and Clarifying Meaning.
Analysis:	Examining Ideas, Detecting Arguments and Analyzing Arguments.
Evaluation:	Assessing Claims and Assessing Arguments.
Inference :	Querying Evidence, Conjecturing Alternatives and Drawing Conclusions.
Explanation:	Stating Results, Justifying Procedures and Presenting Arguments
Self-Regulation:	Self- Examination and Self-Correction

Conceptual framework. The core cognitive skills are conceptualized as interrelated and interdependent. Attainment of each core skill is in relation to the others and depends on the development of the others. The stage theory purported by Paul and Elder (2002, 2006) explicates the six stages of the developmental hierarchy of CTS from unreflective thinker, challenged thinker, beginning thinker, practicing thinker, and advanced thinker to a master thinker. The stage theory demonstrates a linear process which simultaneously progresses from stage one to six with further development of core cognitive skills at each stage. The pattern of development of CTS, that is, its elements, traits, and intellectual standards are interdependent and interrelated while the stage development is hierarchical. For example, as a practicing thinker one might want to review a piece of information and test the idea exercising your thoughts, ensuring the standards of accuracy, precision, and the evidential support for a valid reasoning to draw inferences and make decisions.

The conceptual framework in Figure 3.1 shows the core cognitive skills and their sub-skills (Facione, 1990), and their interdependent relationships using a structural pattern and presented in a schematic form. The researcher proposes that future research could assist in assigning criteria or rubrics for assigning scores for levels of CTS to place individuals on the continuum from unreflective thinker to master thinker.

Figure 3.1. Researcher-created Conceptual Framework of Core Cognitive and Sub-skills of CTS

(Source: Concepts adapted from Facione, APA, 1990)

Generating an Item Pool

"Item writing is largely a creative art" (Haladyna, Downing & Rodrigues, 2002, p. 329). Definition of the critical thinking construct and identification of the core cognitive skills and sub skills assisted in developing a pool of items for the scale. Domain sampling of items included random sampling of homogeneous items (Dawis, 1987; DeVellis, 2003; Polit & Beck, 2008) and purposive sampling to achieve diversity (Shadish, Cook & Campbell, 2002), ensuring that the items reflected the construct and purpose of the scale; items were clear and unambiguous, and avoided wordiness, and double barrelled and double negative statements. The researcher was cautious in writing simple, short sentences that were easy to comprehend and yielded quick responses (DeVellis, 2003; Haladyna et al., 2002).

Most items were positively worded although negatively worded items were included. The literature reveals that reverse scored or negatively worded items tend to reduce the validity of responses and introduce systematic error to a scale. Several researchers have shown that reverse scored items result in an artifactual response factor. Previous studies have revealed that item loadings for reverse scored items were often lower than positively scored items that loaded on to the same factor (Hinkin, 1995). Thus, when the initial review of the items by colleagues revealed some confusion with these items, they were excluded.

Using items from a previously developed scale was an option to consider. However, no self-assessment scale for measuring critical thinking could be located in the literature. Using items from existing instruments was not considered because the items were either multiple choice or rubrics used for the purposes of evaluating CTS through qualitative methodology or the purpose of the scale was to measure the effectiveness of curriculum and instruction. The researcher was aware that too few items might affect internal consistency and test-retest reliability and too lengthy a scale might induce response bias. However, initially, an exhaustive list of 196 items was developed for ensuring adequate representation of the constructs and considering redundancy in providing choices for selection (DeVellis, 2003). The number of items for content validation should be at least 50% more than the targeted number for the final scale (DeVellis). Table 3.2 illustrates the distribution of items across the core skills and sub skills.

Table. 3.2

Number of Items for Core Cognitive Skills and Sub-skills

Core Cognitive Skills	Sub-Skills	Number of items
1. Interpretation	1. 1 Categorization	11
	1. 2. Decoding significance	12
	1. 3. Clarifying meaning	13
2. Analysis	2.1. Examining ideas	12
	2.2. Detecting arguments	12
	2.3. Analyzing arguments	12
3. Evaluation	3.1. Assessing claims	12
	3.2. Assessing arguments	12
4. Inference	4.1. Querying evidence	11
	4.2. Conjecturing alternatives	12
	4.3. Drawing conclusions	13
5. Explanation	5.1. Stating results	12
	5.2. Justifying procedures	12
	5.3. Presenting arguments	12
6. Self-Regulation	6.1. Self-examination	14
	6.2. Self-correction	14
Total		196

Determining the Format of the Instrument

Each item in the scale included a stimulus section composed of a declarative statement illustrating an activity within each of the sub-skills one performs as he or she is engaged in the critical thinking process. The response category included a seven-point rating scale indicating frequency of performance. The category descriptions include (0) = never, (1) = rarely, (2) = seldom, (3) = occasionally, (4) = often, (5) = frequently, (6) = Always. A simple seven point rating scale was chosen to ensure high variability that entails data reduction analysis (Dawis, 1987).

Ethical Considerations

The study was reviewed and approved by the University of Saskatchewan (U of S) Behavioural Research Ethics Board (Beh-REB). The experts were recruited for their participation and those who agreed to participate voluntarily were included. Two experts from the advisory committee were included to round out the convenience sample.

Methodology

The main objective of this research was to establish content validity evidence for the CTSAS. Expert judges' rating of items was acknowledged as an important aspect of scale construction in assessing the items for content-relevance and representativeness.

Expert Review of Item Pool

The review began with the researcher appraising whether or not the items captured the intended construct, as well as the wording, language, grammar, and readability of the items (Polit & Beck, 2008). The researcher used her course peer group initially to assess the quality of the items. Many items were modified accordingly including the rejection of negatively worded items. The researcher then provided the operational definitions of the construct and items for experts' evaluation. A pilot testing of the scale on a small sample (20) of the target population was completed to check the clarity, language, and understanding level after the experts' ratings and modification of the scale (Polit and Beck, 2008). A detailed discussion of the expert evaluation and comments from the target population follows in the section on results.

Evaluation of Experts' Ratings

The use of experts' judgments of items to the construct domain is the most frequent and simple way of finding evidence of content validity. The experts' judgment consisted of two essential processes: item content relevance and item content representativeness (Messick, 1989). Item content relevance refers to "the degree to which the content contained within a test item is representative of the "targeted construct" the item is designated to measure" (Dunn, Buffard, & Rogers, 2000, p. 16). The experts' judgment and ratings assessed the content relevance (content validity). Item matching assessed representativeness, that is, experts were provided with a list of items and asked to match the items with the constructs (Dunn et al, 2000; Haynes, Richard & Kubany, 1995; Messick, 1989).

Methods of Assessing Experts' Ratings.

It is important that the instrument developers generate cumulative evidence for content validity as there is no consensus on the best method for assessing content validity of experts' ratings (Hellsten, 2008; Polit & Beck, 2008). Some of these methods include: simple statement of agreement, inter-rater agreement, item ambiguity (range), mean item rating, median item rating, Content Validity Coefficient (VI_k), Content Validity Index (*I-CVI*), Content Validity Ratio, Average Congruency Percentage (*ACP*), Factorial Validity, and Multi-Rater Kappa Coefficient (Hellsten, 2008). Item content relevance can be assessed by examining the degree of ambiguity among experts' ratings for the item by range; the central tendency of experts' ratings by mean, median and mode (the median may be preferred to mean because the mean is affected by the outliers or aberrant ratings); and Content Validity Coefficient (VI_k) and Content Validity Index (*I-CVI*) for each item (Hellsten).

For example, while using the rating scale, the experts' agreement on the items was determined by summarizing the data for the item *CVI* statistic. Based on a five-point scale (0 – 4), the formula for finding the *I-CVI* for each item is “number of raters giving ratings of 3 or 4 divided by the number of experts” (Polit & Beck, 2008, P. 459). A value of 0.80 is considered as an acceptable level. The average Congruency Percentage (*ACP*) is used as an overall measure of the content validity of a scale (Hellsten, 2008). In order to calculate the *ACP*, panel members were asked to first read the specifications for each dimension and then to rate each item according to the item’s congruence with the specifications (Amstrong, Cohen, Eriksen & Cleeland, 2005; Beck & Gable, 2001). The proportion of items rated as 3 or 4 by each expert was converted to a percentage and ACP was calculated as the mean ratings for all the experts in the panel.

Hellsten stated that there is “no gold standard or a priori criterion” to decide content validity (2008, p. 5). Polit and Beck claimed, “validation efforts should be viewed as evidence gathering enterprises, in which the goal is to assemble sufficient evidence” (2008, p. 458). One of the popular methods used to reach an agreement on content relevance is the Content Validity Coefficient (VI_k). “The formula for calculating the VI_k is = $VI_k = S[j(c-1)]$, where VI_k is the validity coefficient, *K* the item, *j* the number of raters, and *c* is the number of rating categories” (Hellsten, 2008, p. 8). A value close to 1 indicates high validity, and the significance can be found from the “right tailed probability table” provided by Aiken (1985, p.134).

For the purpose of content analysis of the CTSAS the researcher used the *median, range, I-CVI, and* $VI_{k..}$ This cumulative evidence influenced the validity decision for retaining, rejecting, or modifying the items. The quantitative indices of content validity were also supplemented by qualitative feedback from experts. The order of decision making was as follows:

1. Right tailed probability value of validity coefficient VI_k (Aiken, 1985, P.134). For the number of rating categories five (0-4), a value of 0.73 was significant at p <0.009 for 14 items (maximum number of items in a subscale), and a value of 0.66 was significant at $p < 0.036$ for 17 experts. For the *I-CVI* Validity Index, 1 is ideal and an "acceptable level for more than 5 experts raters *I-CVI* is *0.78* or more" (Poilt & Beck, 2008, p.483).
2. Median and Range to identify ambiguity of the items.
3. Experts' comments for modifying the items for wording and language.
4. Target population responses in terms of confusion, clarity, understanding, and language for modifying, and retaining an item in the scale.

Scale format

Choosing the number of responses depends on the number of items, the ability of the respondent to discriminate among response choices, and the researcher's goal (Devellis, 2003). It was decided that the CTSAS would have a seven point response continuum of 0 to 6, with each point corresponding to how frequently the skills were demonstrated ranging from a score of 0 "never" to 6 "Always". A seven point scale was used because the larger response option increases the opportunity for variability and provides more information. A neutral option response continuum was included for those respondents who were uncertain of the items or unable to discriminate the category meaningfully to avoid the tendency to not respond to an item (DeVellis, 2003; Oppenheim, 1996). Having a neutral option in the middle protects the end points being affected. The scores for each category of the scale were summated and the summated scores were used for the analysis. The total score possible for the scale will be available after the content validation and final decision regarding the items to be included in the scale.

Analysis

Composition of the Expert Panel

Several psychometricians have suggested that researchers report the characteristics and qualification of the experts selected. The desired composition and qualifications depends on the type of scale, the intended use, and the target population. This researcher followed similar criteria used in the APA Delphi survey for selection of experts for item rating review. Thus, experts holding a terminal degree from different disciplines (e.g., Education, Nursing, Medical Education, Philosophy, Science, Social Sciences, Educational Psychology, Measurement and Evaluation, Kinesiology, Geology, and Theology) were sought for the panel. The researcher also sought nursing experts from Canada, the United States, and India. India was included because it is the researcher's nationality and the scale was and will continue to be tested with Indian student populations.

Twenty three panel members were identified and contacted and 21 agreed to participate. The mailed content validation packages included an invitation letter and consent for participation, the proposed CTSAS scale, a list of construct definitions, the item rating scale with definitions of core and sub-skills and items for each, and a demographic questionnaire. The experts were provided two weeks in May 2010 for completing and returning the package. Email reminders to the experts were used to enhance the rate of return. Seventeen of the 21 experts returned the completed package, giving a response rate of 81% (Appendix H).

The disciplines included were six from Nursing, two each from Medical Education and Theology, and one each from Educational Psychology and Measurement & Evaluation, Educational Philosophy, Geology, Physics, Sociology, Computer Technology, and Kinesiology. The panel was comprised of 42% male and 58% female experts. The majority of the experts selected were from Canada, and 18% of the nursing experts were from the United States and India. Forty-one percent of the experts had received formal training in CTS; 41% of the experts had experience in teaching critical thinking skills; 58.8% had experience in developing and evaluating research instruments, and one of the nursing experts had experience in developing a CT scale for evaluating effectiveness of curriculum and instruction on CTS. Only two of the 17 experts used an instrument for measuring CTS. Table 3.3 presents the background information of the expert panel.

Table. 3.3

Background Information of the Expert Judges' Panel

n=17

Background information of Expert Panel	Male		Female		Total	
	N	%	n	%	N	%
1. Education						
1.1. PhD	7	41.2	9	52.9	16	94.1
1.2. M. Ed (Philosophy)	-		1	5.9	1	5.9
Total	7	41.2	10	58.8	17	100.0
2. Training in Critical Thinking (CT)						
2.1.Yes	3	17.7	4	23.5	7	41.2
2.2. No	4	23.5	6	35.3	10	58.8
Total	7	41.2	10	58.8	17	100.0
3. Experience in teaching CT						
3.1.Yes	4	23.5	3	17.7	7	41.2
3.2. No	3	17.7	7	41.2	10	58.9
Total	7	41.2	10	58.9	17	100.1*
4. Experience in developing & evaluating research instruments						
4.1.Yes	6	35.3	4	23.5	10	58.8
4.2. No	1	5.9	6	35.3	7	41.2
Total	7	41.2	10	58.8	17	100.0
5. Experience in developing CT scales						
5.1. Yes			1	5.9	1	5.9
5.2. No	7	41.2	9	52.9	16	94.1
Total	7	41.2	10	58.8	17	100.0
6. Used any instrument for measuring CT						
6.1. Yes	1	5.9	1	5.9	2	11.8
6.2. No	6	35.3	9	52.9	15	88.2
Total	7	41.2	10	58.8	17	100.0

- Value reached over 100% due to rounding decimals.

Results

The item rating scale had a response range from 0 to 4. The experts were directed to read the definitions and rate the items as 3 or 4 if the item measured the construct definition. They were also asked to indicate if the items they rated 3 or 4 together defined the category definition or to indicate what was missing. The mean, median, range, Content Validity Index (*I-CVI)* and Content Validity Coefficient (VI_k) were computed for each item and the significance of VI_k values were determined from Aiken's (1985) *right tailed probability* table. Planned decision criteria outlined earlier were used in retaining, rejecting, or modifying the items. The qualitative comments of the experts were used to modify many items in particular related to structure and language. Of the original 196 items, 115 items were retained or modified.

The *right-tail probability value of validity coefficient* VI_k (Aiken, 1985, p.134) was found from the table for 17 experts, with a rating category of 5- a value of 0.66 was significant at $p < 0.036$, and for 12 items a value of 0.75 was significant at $p < .009$. The acceptable level of validity index (*I-CVI)* ≥ 0.7 for all retained items are almost closer to VI_k, which is a better scientific value than *I-CVI*, as it takes into account the variations in the score since the value is calculated from the summated scores of raters. The VI_k for the items retained were highly significant. Refer to Table 3.4 for the values and decisions made for all the subscales and the number of items retained in each core and subscales. Appendix I shows the results of the analysis for content validity and the decisions made to retain, reject, or modify an item.

Table 3.4

Number of Items in Core Scales and Subscales Retained, Modified, and Rejected after Experts' Rating According to I-CVI & VI_k Values

N = 17

Core and subscales	No. of items	No. of items Retained	Value of *I-CVI* (Ranges)	Value of VI_k (Ranges)	*P*-Values	Comment
1. Interpretation	36	21				15 rejected
1.1 Categorization	11	7	.75-.938	.766-.875	< .01	3 /7mod*
1.2 Decoding Significance	12	7	.687- 1	.70 - .88	< .01	5/7 mod
1.3. Clarifying Meaning	13	7	.81 - 1	.78 - .95	< .01	1/7 mod
2. Analysis	36	18				18 rejected
2.1 Examining Ideas	12	6	.625- .875	.703-.813	< .05	1/6 mod
2.2 Detecting Arguments	12	6	.75 - .875	.73-.91	< .01	3/6 mod
2.3 Analyzing Arguments	12	6	.81 - .875	.76 -. 86	< .01	1/6 mod
3. Evaluation	24	13				11 rejected
3.1 Assessing claim	12	5	.687- .937	.70 - .89	< .01	2/5 mod
3.2 Assessing Arguments	12	8	.75 - .937	.703-.86	< .01	5/8 mod
4. Inference	36	22				14 rejected
4.1 Querying Evidence	11	7	.75 - .937	.73 - .891	< .01	2/7 mod
4.2 Conjecturing Alternatives	12	7	.625-.937	.73 - .91	< .01	2/7 mod
4.3 Drawing Conclusions	13	8	.625-.937	.69 - .92	<.05/.01	4/8 mod
5. Explanation	36	22				14 rejected
5.1 Stating Results	12	5	.687 - .81	.609 - .75	< .05/.01	2/5 mod
5.2 Justifying Procedures	12	9	.75 – 1	.70 -.88	<.05/.01	3/9 mod
5.3 Presenting Arguments	12	8	.81 - .937	.766-.891	< .01	Good
6. Self- Regulation	28	19				9 rejected
6.1 Self-Examination	14	9	.75 - .937	.69 - .90	< .05/.01	3/9 mod
6.2 Self- Correction	14	10	.75 - .937	.70-.92	< .01	5/10 mod

Aiken's VI_k - *Right tailed probability for five rating category (0-4) values of the Validity Coefficient.* (*mod: modified)

Number of items 11, a value of .77 has a $p < .006$
Number of items 12, a value of .75 has a $p < .009$
Number of items 13, a value of .75 has a $p < .006$
Number of items 14, a value of .73 has a $p < .008$
For 17 experts, a value of 0.66 has a p < 0.036

Item relevance and item ambiguity is decided by the values of range for each item. A value of 2.75 or below is considered a relevant or unambiguous item. The ‘range’ values for some of these items were found to be very high (4 or 5) (acceptable range-value is ≤ 2.75) indicating highly ambiguous items mainly due to the language and words used in framing the statements. For example, Item 1.1.1. When presented with a problem, I try to figure out the ‘*various aspects’* of the problem. The *italicized* words were replaced by *‘content’*. In item, item 1.2.3. I ‘*appreciate’* the meaning….. the word *appreciate* was replaced with *‘clarify*. For details of modification of items refer to Appendix I.

Wherever ambiguity was a concern, as suggested by the experts, even if the items had the significant VI_k and acceptable *I-CVI* values, these items were revised and modified to ensure wordings of the items were precise, unequivocal, and easy to understand**.** Language was a challenge for the researcher in constructing the items that are easy to read and universally comprehensible. The suggestions of the experts for rewriting some of the items were very helpful in making items less confusing. Sixteen out of 17 experts (94%) agreed that the items that they rated either 3 or 4 of the subscales together match the construct definition of each category of the designated cognitive skill.

As illustrated in the Table 3.4, between 50 – 72% of the original items were retained following content analysis, with an overall retention rate of 59%.

Table 3.5

Number of Items Retained for Core Cognitive Skills and Sub-skills

Core Cognitive Skills	Sub Skills	Number of items	Items Retained
1. Interpretation	1. 1 Categorization	11	7
	1. 2. Decoding significance	12	7
	1. 3. Clarifying meaning	13	7
2. Analysis	2.1. Examining ideas	12	6
	2.2. Detecting arguments	12	6
	2.3. Analyzing arguments	12	6
3. Evaluation	3.1. Assessing claims	12	5
	3.2. Assessing arguments	12	8
4. Inference	4.1. Querying evidence	11	7
	4.2. Conjecturing alternatives	12	7
	4.3. Drawing conclusions	13	8
5. Explanation	5.1. Stating results	12	5
	5.2. Justifying procedures	12	9
	5.3. Presenting arguments	12	8
6. Self Regulation	6.1. Self-examination	14	9
	6.2. Self-correction	14	10
	Total	196	115

The revised 115 items (refer to Table 3.5) were submitted for a second review to five experts who had agreed to comment on the items. The second review resulted in almost full agreement among the five experts with no suggested changes and the conclusion was that the retained items possessed a high level of face validity. The revised scale with 115 items was piloted with a group of 20 undergraduate students who were asked to comment on the clarity, confusion, difficulty, and ease of reading of the items. The students estimated the approximate time taken to complete the scale as 35-40 minutes. For *Item 1.2.2 "I observe the facial expression or nonverbal cues people use in a given situation"*, 75% of the students rated the item to be difficult to understand and the sentence was revised by removing the phrase '*nonverbal cues'* as some students suggested. The majority (114) of the other items were rated as easy to read, clear or comprehensible.

The final scale was constructed with 115 scientifically filtered items. Items within each subscale were randomly entered. Each had a seven point response continuum of 0 to 6 and the descriptors of 0 = never, 1= rarely, 2 = occasionally, 3 = usually, 4 = often, 5 = frequently, and 6 = always. All items are positively phrased. The total score ranges from 0 to 690.

Discussion

As noted earlier, several articles reviewed on instrument development revealed inadequate explanations related to the content validation process for developing items for a scale. The California Critical thinking Disposition Inventory and Critical Thinking Skills Test manual explains the number of initial items generated and the final version of the scale with reduced numbers but does not provide the details of the content validation process for reaching the final number of items (Facione & Facione, 2007). Other researchers have provided brief and limited information on the process of validation and used four experts to validate the scale (Rask, Malm, Kristofferzon, Roxberg, et al., 2009; Weis & Schank, 2009).

In terms of researchers who have reported on the processes, one study used two rounds of Delphi technique, more qualitative in nature, seeking consensus of opinions of experts on the items. Delphi techniques are very valuable if used for item matching with the construct definitions (Wilkes, Mohan, Luck & Jackson, 2010). Another research team reported a content validity index for the spiritual distress scale and stated four practitioners in the hospital graded the scale as acceptable (Ku, Kuo, & Yao, 2010). Varas-Diaz and Neilands (2009) reported on development and validation of a culturally appropriate HIV/AID stigma scale and provided limited information on content validation, however, they stated they used seven experts' 100% agreement and reported a resulting scale of 68 items.

A group of researchers reported on content validation of the Chinese version of a heart failure learning needs inventory. Ten experts from three professional categories reviewed the inventory for face validity, content validity, and cultural relevance, and described the changes made in the items (Yu, Chair, Chan, & Liu, 2010). Dunn et al. (1999) reported in detail and addressed issues related to content validation in response to Messick's concerns, and Aiken's affirmation that "accurate measurements of variables depend on the sophistication with which the instruments for measuring them are designed" (Dunn et al., p.15.). Dunn et al. provided insight into systematically planning and development of the CTSAS scale and reporting of the results. Similar information on content validation has been reported for previous clinical scale development studies (McMillan, 2001; Vaartio, Sominen & Puukka, 2009; Zheng, You, Lou, Chang, & Lai et al., 2010).

The evidence for content validation of a scale can be established by several methods. Given that critical thinking is a universal construct used by all academic disciplines, this researcher chose to use the most scientific methods and an adequate number of experts from various fields to ensure the items for the CTSAS were relevant and representative of the construct. In the validation process several statistical measures as well as qualitative data were used to help identify a well developed set of items for each subscale and the researcher was cautious in using a conservative level of significance to retain items. As asserted by Messick (1995), a clear definition of the construct and the matching of items by the experts' ratings and appropriate statistical evidence ensures scientific rigour and allows for future researchers in re-examining and supporting or refuting the results.

Conclusion

This paper described the process and provided evidence to support the decisions made to retain or modify items for the CTSAS. The content validation process for the CTSAS aided in deriving 115 well written items for measuring each subscale. These items were the result of careful experts' review and ratings. The analysis of experts' ratings and use of statistical techniques along with qualitative comments and suggestions by experts assisted in modifying, rejecting, and retaining well crafted items with high level of content validity for the final scale. Most of the items were rated by the representatives of the intended target population as easy to read, clear and comprehensible. In the next stage of the study, the scale will be administered to a selected sample from India and Canada for further analysis using data reduction techniques in order to establish the construct validity of the scale.

References

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological measurement, 45*, 131-142.

Amstrong, T.S., Cohen, M.Z., Eriksen, L., & Cleeland, C. (2005). Content validity of self-report measurement instruments: An illustration from the development of the brain tumour module of M.D. Anderson Symptom Inventory. *Oncology Nursing Forum, 32*, 669-676.

Austin, Z., Gregory, P. A. M., & Chiu, S. (2008). Use of reflection-in-action and self- assessment to promote critical thinking among pharmacy students. *American Journal of Pharmaceutical Education, 72(3),* 1-8.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood: Prentice Hall, Inc.

Beck, C. T., & Gable, R. K. (2001). Ensuring validity: An illustration of the process. *Journal of Nursing Measurement, 9,* 201-205.

Carter, L. M., & Rukholm, E. (2008). A study of critical thinking, teacher-student interaction, and discipline-specific writing in an online educational setting for registered nurses. *The Journal of Continuing Education in Nursing, 39*(3), 133-138.

Dawis, R. V. (1987). Scale construction. *Journal of Counselling Psychology, 34*(4), 481-489.

Dickieson, P., Carter, L. M., & Walsh, M. (2008). Integrative thinking and learning in undergraduate nursing education: Three strategies. *International Journal of Nursing Education Scholarship. 5*(1), 1-13.

DeVellis, R. F. (2003). *Scale development: Theory and application.* Thousand Oaks, CA: Sage.

Dunn, J. G. H., Bouffard, M., & Rogers, W. T. (2000). Assessing item content-relevance in sport psychology scale-construction research: Issues and recommendations. *Measurement in Physical Education and Exercise Science, 3*(1), 15-36.

Facione, P. A. (1990). Critical thinking: *A statement of expert consensus for purposes of educational assessment and instruction: Executive summary, The Delphi report*. American Philosophical Association, California Academic Press. Retrieved from http://www.isigntassessment.com/article.html

Facione, N. C., & Facione, P. A. (2007). *The California Critical Thinking disposition inventory: A test of critical thinking disposition. Test Manual*, The California Academic Press, USA.

Facione, P. A. (2007). *Critical Thinking what it is and why it counts.* Update 2007, retrieved from http://www.insightassessment.com/articles.html

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple choice item writing guidelines for classroom assessment. *Applied Measurement in Evaluation*, *15*(3), 309-334.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment. 7*(3), 238-247.

Hellsten, L. M. (2008, March*). Accumulating content validity evidence: Assessing expert panel ratings of item relevance and representativeness.* Paper presented at the National Council on Measurement in Evaluation.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*(5), 967-988.

Ku, Y. L., Kuo, S. M., & Yao, C. Y. (2010). Establishing validity of a spiritual distress scale for cancer patients hospitalized in southern Taiwan. *International Journal of Palliative Nursing, 16(3),* pp. 133-137.

McMillan, J. H. (2001). Secondary teachers' class room assessment and grading practices. *Educational Measurement: Issues and Practices*. Spring, pp.20-32.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1990). Validity of test interpretation and use. (Report No. ETS-RR-90-11) Princeton, N.J: Educational Testing Service.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50* (9), 741-749.

Oppenheim, A. N. (1996). *Questionnaire design, interviewing and attitude measurement*. London: Printer Publishers.

Paul, R. W., & Elder, L. (2002). *Critical thinking: Tools for taking charge of your professional and personal life*. New Jersey: FT Press.

Paul, R. W., & Elder, L. (2006). *Critical thinking: Learn the tools the best thinkers use* (Con. ed.). Pearson, NJ: Prentice Hall.

Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

Rask, M., Malm, D., Kristofferzon, M. L., Roxberg, A., et.al., (2009). Validity and reliability of a Swedish version of the relationship assessment scale (RAS): A pilot study. *Canadian Journal of Cardiovascular Nursing, 20(1)*, pp.16-21.

Shadish, W. R., Cook, T.D., & Campbell, D. T. (2002). *Experimental and quasi-experimental design for generalized causal inference.* Boston: Houghton Mifflin Company.

Vaartio, H., Sominen, H. L., & Puukka, P. (2009). Measuring nursing advocacy in procedural pain care –Development and validation of an instrument. *Pain Management Nursing, 10 (4),* pp206-219.

Varas-Diaz, N., & Neilands, T. B. (2009). Development and validation of a culturally appropriate HIV/AIDS stigma scale for Puerto Rican health professionals in training. *AIDS Care, 21(10),* pp.1259-1270.

Weis, D. & Schank, M. J. (2009). Development and psychometric evaluation of the nurses' professional values scale-revised. *Journal of Nursing Measurement, 17(3),* pp. 221-231.

Wilkes, L., Mohan, S., Luck, L., & Jackson, D. (2010). Development of a violence tool in emergency hospital setting. *Nurse Researcher, 17(4),* pp.70-82.

Yu, M., Chair, S. Y., Chan, C. W. H. & Liu, M. (2010, July/August). Testing the psychometric properties of the Chinese version of the heart failure learning needs inventory. *Heart & Lung, 39*(4) 262-274. Retrieved from http://www.heartandlung.org

Zheng, J., You, L. M., Lou, T. Q., Chen, N. C. & Lai, D. U., et al., (2010). Development and psychometric evaluation of the dialysis patient-perceived Exercise Benefit and Barriers Scale. *International Journal of Nursing Studies, 47,* 166-180.

SECTION 4

CONSTRUCT VALIDITY

ANALYSIS, RESULTS AND INTERPRETATION

Construct Validity of the Critical Thinking Self Assessment Scale

Analysis, Results and Interpretation

Critical thinking skills (CTS) are essential to human life for effectively addressing the challenges of a rapidly changing society. Developing and assessing CTS among young undergraduate students is a role of education. Measuring the acquisition and development of this latent trait is the responsibility of the educators. Educators often are challenged with assessment of CTS due to lack of accurate and scientific devices to measure the presence and level of CTS skills. The purpose of this developmental research was to develop and evaluate the preliminary psychometrics properties of the Critical Thinking Self Assessment scale (CTSAS) for undergraduate nursing students. The study used Cronbach's Generalizability theory in developing, testing and evaluating the psychometrics of the CTSAS (Brennan, 2006). The psychometric evaluation was based on Messick's Unitary concept of construct validity framework (Messick, 1990). The Unitary concept of construct validity framework includes content validity, structural validity, reliability, convergent validity and discriminant validity, criterion validity and consequential validity: this study was concerned with the first four elements of the framework. Content validation of CTSAS had already been achieved using a panel of seventeen experts, with an outcome of reducing 196 items to 115 items distributed across six core cognitive skills with 16 subscales (Section 3, Manuscript 2) (Appendix T).

Objectives of the Study

The main objectives of the study were: (i) To establish evidence for construct validity (structural validity) of the CTSAS; (ii) to establish reliability (internal consistency and stability) of the CTSAS; and (iii) to establish convergent validity of the CTSAS. This section of the dissertation presents the process and procedures involved in testing and evaluating the 115 item

CTSAS for its reliability, structural validity and convergent validity. A series of statistical analyses were performed to evaluate the items in the scale.

Testing the Scale on a Developmental Sample

In preparation for the next phase of testing, the scale format included a title page, a brief description of the CTSAS, instructions for the participants for responding to the scale in the opscan (answer) sheets, and the actual questions. All items were positively phrased and were numbered from 1-115 marked section A without revealing the names of the different skills. The scale was scored over a seven point (0-6) response category indicating how often the participants performed these skills. The response category description included: 0-never, 1-rarely, 2-occassionally, 3-usually, 4-often, 5-frequently, and 6-always. The CTSAS items were followed by the 18 item Need for Cognition Short Form (NCS - SF) scale as section B, numbered 116 to 133, which provided data for convergent validity testing. The NCS scale was placed in the last section of the questionnaire to avoid introducing a context that would influence the core cognitive skills. A composite score was assigned to each of the 16 subscales and the core cognitive scales to measure critical thinking skills. The number of items and the possible scores for each of these subscales and core scales are presented in Table 4.1. The total possible score in the scale ranged from 0 to 690. Participants' demographic information was also collected on the answer (opscan) sheet (Appendices T – T 3).

Table 4.1.

Number of Items for Core Cognitive Skills and Sub-skills

Core Cognitive Skills	Sub Skills	Number of items	Maximum Scores
1. Interpretation	1. 1 Categorization	7	42
	1. 2. Decoding significance	7*	42
	1. 3. Clarifying meaning	7	42
2. Analysis	2.1. Examining ideas	6	36
	2.2. Detecting arguments	6	36
	2.3. Analyzing arguments	6	36
3. Evaluation	3.1. Assessing claims	5	30
	3.2. Assessing arguments	8	48
4. Inference	4.1. Querying evidence	7	42
	4.2. Conjecturing alternatives	7	42
	4.3. Drawing conclusions	8	48
5. Explanation	5.1. Stating results	5	30
	5.2. Justifying procedures	9	54
	5.3. Presenting arguments	8	48
6. Self-Regulation	6.1. Self-examination	9	54
	6.2. Self-correction	10**	60
	Total	115	690

*one item excluded from the analysis after data screening.
** three items excluded from the analysis after data screening

Sample characteristics. The participants were undergraduate nursing students from three sites of the College of Nursing, University of Saskatchewan, Canada and BSc. Nursing students from five colleges of Mahatma Gandhi University (recognized by University of Grant Commission, India), Kerala State, India. Of the 950 students enrolled in the five Indian colleges, 887 participated in the survey with response rate of 94 %. In Canada, the researcher targeted 353 students who were enrolled in years III and IV of the program, and 144 (41%) students volunteered to complete the scale.

Participants were provided with a packet that included the invitation letter, consent form and the question booklet along with an opscan sheet for the responses (Appendixes T – T 3). Participants completed the questionnaire in the classroom within 35 - 45 minutes. The test was repeated after two weeks in three Indian colleges with 251 year III and year IV students. Participants were offered a chance to enter a draw for a $50 gift certificate from U of S bookstore for each of the three sites in Canada or local medical bookstores close to each of the colleges in India.

The researcher personally administered the questionnaire and provided necessary instructions for completing the scale. The background information on the participants is presented in the Table 4.2. In India, the entire target population, all the four years of students in the program, of the five colleges participated, whereas in Canada, only Years III and IV of the University of Saskatchewan students were targeted. Year I and Year II students were not considered as they were admitted to Saskatchewan Institute of Applied Science and Technology (SIAST) and the students are not in a University learning environment. Including these students would have added another intervening variable to the study – that of learning environment. For the Indian population, the response rate ensured the sample was representative of the population. Demographics for the Canadian population were unavailable to compare with the sample to ensure representativeness. Hence, a comparison between the two sample characteristics could not be established. However, the majority (95%) of the participants in both groups were female. Most (95%) of the Indian participants were in the age group of 18 to 24 years and 65% of the Canadian participants were in the age group of 18-24 years.

Table 4.2.

Sample Characteristics

Sample Characteristics	Canada (N=144)		India (N=887)		Combined (N=1031)	
Gender	n	%	n	%	N	%
Male	6	4.2	38	4.3	44	4.3
Female	138	95.8	849	95.7	987	95.7
Total	144	100	887	100	1031	100
Age						
18-20 yrs	14	9.7	532	60.0	546	53.0
21-24 yrs	79	54.9	352	39.7	431	41.8
25-28 yrs	30	20.8	3	0.3	33	3.2
29 or more	21	14.6	0	0	21	2.0
Total	144	100	887	100	1031	100
Year of study						
Year I	0		241	27.2	241	23.3
Year II	0		236	26.6	236	23.0
Year III	85	59.0	171	19.3	256	24.8
Year IV	59	41.0	239	26.9	298	28.9
Total	144	100	887	100	1031	100

Step 1: Exploratory Data Analyses

Figure 4.1. Plan of Exploratory Data Analysis

Step 1: Data Distribution

1.1 Find missing data for each subscale
1.2 Cleaning data / case wise deletion
1.3 Distribution & skewness /kurtosis
1.4 Sensitivity analysis

→

- Check for accuracy and integrity of data entry.
- Cleaning data is done to determine any errors in data entry.
- Identify the shape /normality distribution examine for skewness and kurtosis, outliers histograms, Q-Q plots, Box plots and flag Items that do not meet normality.
- The purpose of sensitivity analysis is to check if inclusion or exclusion of items influences the distribution. For example, including and excluding items that are skewed in analysis.

Data Distribution

Results and interpretation of EDA. Data were entered manually and when completed, every 10th case was re-examined for errors and corrected as required. Data screening with exploratory data analysis in SPSS (PASW Statistics Grad Pack 18 for Windows) was performed to check for accuracy of data entry and missing values. A few more errors in data entry were detected and corrected. Four cases were identified from the probability plots and were rectified by revisiting the original data. A total of 13 cases were identified with missing values in the Indian sample. These cases were excluded from the analysis within the SPSS program because the missing values constraints the size of correlations in R matrix. Most multivariate statistics such as Exploratory Factor Analysis calculate the eigenvalues and eigenvectors from the correlation matrix. With missing data correlation matrix, the eigenvalues sometimes become negative. “The positive eigenvalues are inflated by the size of negative eigenvalues resulting in inflation of variance" and the outcome could be a distorted statistical result (Tabachnic & Fidell, 2007, p 70). The Canadian sample had no missing values for any items.

The normality of data was examined for each item looking at mean, standard deviation, skewness, kurtosis, the histograms, stem and leaf, and box plots. Four items (item10, item110, 113, and 114) were either skewed or kurtotic or found to have outliers and were removed from further analysis. Refer to Table 4.3 for the values. Data analyses were repeated excluding these four items and yielded better distributions for those scales. Results of the analyses of each of these subscales are presented in Appendix J. Tables 4.4 and 4.5.

Table 4.3.

Non- normal Distribution of Data, Skew and Kurt Values of Items Deleted

Core scales /Sub scales /Items	Skew		Kurt		Outliers		Decision
	India	Canada	India	Canada	Ind	CA	
CS 1-SS 2 (*Decoding significance*) Item 10 I observe the facial expression people use in a given situation	-1.646	-1.056	1.916	0.491	13	--	Item excluded for EFA
CS 6-SS 16 (*self –correction)* Item 110 I respect others' point of view even if they contradict mine	-.801	-1.046	-.185	-.364		-	Excluded
Item 113 I am aware of my strengths and weaknesses	-1.168	-.635	-.468	-.364	-	-	Excluded
Item 114 I am aware of my values and beliefs and control its undue influence on my thinking.	-.980	-.766	-.357	-.446	5	6	Excluded

The remaining 111 items met the normal probability assumptions. (Refer to Appendix K for Distribution of Normalcy for the individual items).

Exploratory data analyses of the composite scores of each of the core scales and subscales were conducted to examine accuracy, normality and outliers. Analysing the composite scores using the descriptive statistics mean (M), median (Mdn), Standard deviation (SD), skewness (skew), kurtosis (kurt) and normal probability plots revealed another six (6) outliers distributed across the CS Interpretation (1), Inference (5), Explanation (1), and Self-Regulation (1). Two cases had outliers on more than one scale. Thus, a total of 19 cases were excluded from further statistical analysis of various scales.

Exploratory data screening aided in identifying and deleting four items (10, 110, 113 and 114) which probably relate to the socio cultural aspect of thinking skills. The researcher is unsure whether or not these items were affected by differing culturally appropriate behaviour acceptable to the two samples studied. These findings seem similar to the statements made by Paul and Elder (2006), that most of our thinking is either egocentric or sociocentric in nature. Bandura, (1986) also has claimed that personal and social factors influence the development of cognitive skills. These four items had inconsistent results among total Indian sample and the two (groups) subsamples and the Canadian sample. The data analysis was repeated excluding these items and yielded better distributions for those scales.

Step 2: Reliability Analysis of the CTSAS

Figure 4.2. Plan of Reliability Analysis

Step2: Internal consistency reliability

1. Internal consistency : Cronbach's α
   1.1. For each sub scale
   1.2. For each Category
2. Internal consistency with Indian and Canadian samples
3. Stability Reliability: Test-retest (Each subscale and each category With (Yr III & IV- Indian sample)

→

- A Cronbach's α for the total scale will not be meaningful
- Simple correlation.
- A comparison of the Cronbach's α on Indian and Canadian sample would probably add strength to validity of the scale.
- Internal consistency reliability (Cronbach's α) for year I & II combined and year III & IV combined might explain variability among cohort on the CTSAS.

Internal consistency Reliability

The CTSAS with the 111 items retained after EDA were analysed to assess the reliability of the scale. Reliability (internal consistency) analysis of the six (6) Core Scales (CS) and 16 Sub-Scales (SS) was performed and the reliability values (Cronbach's α) are reported in this section with interpretation.

Results and interpretation. The overall Cronbach's α for each of the CS and SS are presented in Table 4.6 with the missing cases in each scale for the Indian sample. There were no missing values in the Canadian sample. The overall reliability of the core scales was within the range of acceptable levels (0.7 to 0.8), as indicated by Kline (2005). However, for four of the subscales the Cronbach's α values were below 0.7, that is two SS from CS 1 Interpretation, one in CS 2 Analysis, and one in CS 3 Evaluation. Table 4.6.1 shows the results of the reliability analysis for *decoding significance* [1] (SS 2) for the Indian sample.

[1]Core scales are indicated in the text with regular font and subscales are with *italic* font.

Table 4.6.

Cronbach α - Reliability Values after EDA Screening and Prior to EFA

Scale Category	# Items	India (N=868) Missing cases	India (N=868) r	Canada (N=144) r
CS1 Interpretation	20	5	.834	.890
SS 1 Categorization	7	1	.766	.826
SS 2 Decoding Significance	6	2	**.595**	.782
SS 3 Clarifying Meaning	7	4	**.667**	.829
CS2 Analysis	18	7	.857	.917
SS 4 Examining Ideas	6	4	.713	.811
SS 5 Detecting Arguments	6	4	**.666**	.825
SS 6 Analysing Arguments	6	3	.721	.855
CS 3 Evaluation	13/12*	5	.817/.822**	.836
SS 7 Assessing Claim	5	1	**.685**	.807
SS 8 Assessing Arguments	8/7*	5	.733/.746**	.707/.808**
CS 4 Inference	22	12	.875	.927
SS 9 Querying Evidence	7	5	.736	.822
SS 10 Conjecturing Alternatives	7	6	.749	.844
SS 11 Drawing Conclusions	9/8*	5	.745/.748**	.845
CS 5 Explanation	22	10	.902	.931
SS 12 Stating Results	5	0	.740	.790
SS 13 Justifying Procedures	9	4	.822	.875
SS 14 Presenting Arguments	8	8	.779	.897
CS 6 Self-regulation	16	5	.887	.932
SS 15 Self- examination	9	2	.818	.900
SS 16 Self- correction	7	4	.782	.842

*item deleted;
**values after deletion of items.

For example, the overall Cronbach α for the SS *decoding significance* was 0.595. Although the item-total correlation for two items 8 and 9 was < 0.3 (r = 0.256, r = 0.280), the alpha value (r =0.583, r = 0.570) for these items was closer to the overall α. The concern here is when the alpha for specific items is greater than the overall alpha value for that scale. In this case, these values were not greater than the overall α (0.595) for the scale and there was no change in the overall values when tested with exclusion of these items. Hence, these items were retained for further analysis. The reliability values for these items for the Canadian sample were very good (α = 0.890 for CS and 0.782- 0.829 for SS).

Table 4.6.1

SPSS Output for Subscale 2 Decoding Significance

Item-Total Statistics (India)

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
8. I figure out a person's purpose in asking a question.	19.02	26.555	**.256**	.118	**.583**
9. I clarify the meaning of an individual's gesture in a given situation.	18.77	27.262	**.280**	.119	**.570**
11. I try to identify the social importance of the information presented in the texts.	19.31	24.985	.387	.182	.525
12. I look for various relationships among concepts to understand the meaning.	19.08	25.482	.375	.157	.531
13. I examine the values rooted in the information presented.	19.02	26.068	.369	.164	.534
14. I identify the author's views and intentions in the issue presented.	19.57	25.812	.317	.142	.555

However, for *decoding significance* (SS 2) the value of $r = 0.782$ was low compared to other scales with values > than 0.8 and within an acceptable level (Table 4.6).

In the Indian sample, similar findings were noted in Analysis (CS 2). In this CS, the *analyzing arguments* (SS 5 - 6 items scale), item 28 and *detecting arguments* (SS 5) item 33 showed item-total correlation < 0.3 (0.286 and 0.232) and an overall α value for the SS = 0.666. The analysis was repeated with the exclusion of these two items, which resulted in an improved reliability for this subscale with α = 0.690. There were two items in Evaluation (CS 3) and Inference (CS 4) that had an inter-item correlation < 0.3. The analysis was repeated with exclusion of these items which resulted in improved reliability. A decision was made to run Exploratory Factor Analysis with both inclusion and exclusion of these items. For details of the description of the scale and the inter-item correlation values please refer to Table 4.6.2. in Appendix L.

The analysis of all the remaining core scales and sub scales showed very good reliability values. In particular, the Canadian sample showed excellent results on the reliability of the various scales. The reliability was reanalyzed with the retained items following EFA and is reported later in this document.

Stability Reliability: Test-Retest reliability

A valid instrument must also be reliable. Test-retest reliability measures the stability of an instrument over time. Other things being equal, individuals should get the same score on a questionnaire if they complete it at two points in time (Field, 2009). The CTSAS was administered to 251 Year III and Year IV students in the Indian sample from three colleges who were willing to participate a second time. The interval difference between the first and second tests was two weeks. Exploratory data screening revealed negligible variability.

Results and interpretation. The data met the assumptions of normality. The scores on core and sub scales were summated separately and subjected to Pearson's product-moment correlation (*r*) and Spearman's (*rho*) correlation coefficient. All core-scales and subscales' Pearson correlation coefficient of *r* and Spearman *rho* were significant at $p \leq 0.01$ indicating a high correlation between the first and the second tests. The results are presented in Table 4.7 of Appendix M. The *r* value for SS 11 *justifying procedures* had a very low correlation, although significant, compared to other scales and the scale was found to be significant possibly due to the large sample size.

The coefficient of determination R^2 for the Pearson *r* shows the amount of variability in one variable that is shared by the other. In this case, the percentage of shared variability reveals that maximum variability explained with repeated test was 27.77% which means more than 72.33% shared variance was accounted by other factors. The high correlation, thus, indicates possible similar results with repeated testing. However, test-retest reliability at this stage would not provide much information about the stability of the scale as the structure and length of the scale changed after Exploratory Factor Analysis and Confirmatory Factor Analysis.

Step 3: Exploratory Factor Analysis (Internal Validation)

Factor analysis helps the researcher understand the underlying theoretical structure and the factors that are included within the structure. It is useful in developing and assessing theories. EFA assists in answering the questions: (i) What is the structure of critical thinking skill (CTS) construct? and (ii)To what extent does the factor structure reveal pre-designated critical thinking constructs?

The E FA is based on the assumption that the measured responses are based on the underlying dimensions. Principal Component Analysis (PCA) was used for simple data reduction in order to identify the underlying dimensions. PCA is the linear combinations of the measurements and thus contains both common and unique variances and extracts maximum variance from the observed data set along with residual variance.

Preliminary Exploratory Factor Analysis (EFA) was conducted for testing assumptions and sampling adequacy. Appendix N Figure 4.3 gives the criteria set for the testing assumptions of EFA, determining factors, and retaining of items within factors.

Plan of Data Analysis

Exploratory Factor Analysis (EFA) was completed with both orthogonal (varimax) and oblimin rotation

1. Each of the Core scales (6): (Whole sample-India)
2. Each of the Core scales (6): Year I & II combined (India)
3. Each of the Core scales (6): Year III & IV combined (India)

Results and Interpretation

The total Indian sample size, after cleaning data, was 868 (19 cases either missing values or outliers were excluded) across the sub scales for conducting EFA. The sample included students from all the four years of the undergraduate programs. The Canadian sample, comprised of 144 participants, was subjected to a trial run of EFA to see if there were any similarities or differences in the pattern of item loadings. Principal components extraction with varimax and oblimin rotation was performed through SPSS (PASW Statistics Grad Pack 18) for the six (6) core scales independently for the total Indian sample and for the two subsamples. The results of the analysis and interpretation for each scale are presented in the following sections.

Core Scale 1 Interpretation. Principal Component Analysis (PCA) was performed for the 20 items in the Core Scale (CS 1) Interpretation with orthogonal (varimax) rotation. As seen from Table 4.8 the Kaiser–Mayer-Olikin (KMO) measure verified the sampling adequacy for the analysis, KMO = 0.889 (great according to Field, 2009). All KMO values for individual items were > 0.664 which is above the acceptable limit of 0.5 (Field, 2009). Bartlett's Test for sphericity χ^2 (*df* 190) = 3244.299, $p < .001$, indicated that correlations between items were sufficiently large for PCA and the correlation matrix revealed factorability of R with an $\alpha <$ .001 level, although a few of the inter-item correlations produced < 0.3 values which probably were attributable to the large sample size as "larger sample sizes tend to produce smaller correlations" (Tabachnick & Fidell, 2007, p. 614). An initial analysis was run to obtain eigenvalues for each component. Five factors extracted with Kaiser criterion of ≥ 1 in combination explained 48.59% of the variance. Refer to Table 4.8 for the values of assumptions tests and factor extraction.

Table 4.8

SPSS output: Results of Assumption Test for Core Cognitive Scale 1 Interpretation (Indian Sample N = 882)

Test performed	Values Obtained	Probability	Comments
No. of Factors	Varimax (V) -5		Oblimin rotation provided a
extracted with	Oblimin (O) -3		better solution for the CS
eigenvalues above 1	V = 48.591%		interpretation.
Variance	O =38.124%		
explainedKMOBartlett’	.889	$p < .001$	Factor 1 loaded with 7 items
s Test of Sphericity	3244.299*		Factor 2 loaded with 4 items
Range of SMC	*df.* (190)		Factor 3 loaded with 6 items
values(Communalities)	*V= 0.378 - 0.578		
	*O = 0.237 - 0.501.		

*V : Varimax; *O : Oblimin

Figure 4.4 Scree plot showing point of inflexions - CS 1 Interpretation

The scree plot in Figure 4.4 was rather confusing with two points of inflexion. The five factor rotated (varimax) component matrix did not produce a well defined factor solution with a range of communality between 0.378 and 0.578 for items after extraction. According to the set cut off of value 0.45 for inclusion of variables, seven variables loaded on to Factor 1, five loaded into F2, three loaded in F3, two loaded in F4 and two loaded in F5, which was not an interpretable solution. All items loaded independently to each factor and only one out of 20 items did not load onto any factors. However, the unsymmetrical transformation matrix for factor correlation and the scatter plot clusters indicated the scope of more homogenous factor loadings with oblimin rotation.

An oblimin rotation was tried by fixing a three factor solution (Figure 4.4 scree plot first point of inflexion), and the items that had a loading below the value of 0.45 were suppressed. A better interpretable solution was obtained with the total variance of 38.124% for the three factor solution, and the communalities after extraction ranged between 0.237 and 0.501. Defining the three factors became more clear and closer to the pre-designated construct of subscales. Out of the 20 items in this core scale, seven items loaded onto Factor 1 labelled *categorization* (SS 1), four items loaded in Factor 2 labelled *clarifying meaning* (SS 3), and six items loaded onto Factor 3 labelled as *decoding significance* (SS 2). Three items were below the cut off level for loading in any of the factors. All item loadings achieved independent loadings with more than adequate weights in each factor. The loadings of items on factors, communalities and percent of variance for total Indian sample are shown in Table 4.9.

The PCA for Interpretation (CS 1) performed separately for sub samples 1 (Yr I & Yr II) and 2 (Yr III & Yr IV) produced closer to these results with minor variations in Factor 2 and Factor 3. Interestingly, running PCA with 144 Canadian participants, six (6) items each loaded

Table 4.9

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 1 Interpretation with Inclusion and Exclusion of Item 8 - Indian Sample (N=882)*

No	Items	Oblimin			Oblimin			Com h^2
		1 Cat *	2 DS*	3 CM*	1 Cat*	2 CM*	3 DS	
5.	I break the complex ideas into manageable sub-ideas.	.736			.704			.526
2	I sort the information into sub sets.	.717			.641			.536
3	I classify data using a framework.	.652			.711			.512
1	When presented with a problem, first I try to figure out the content of the problem.	.623			.562			.488
4	I break down problem into discrete parts.	.570			.530			.401
6	I categorize similar and related information into groups	.552			.510			.451
7	I classify whole information into specifics.	.530			.532			.424
20	I figure out the meaning of another's point of view.		.695			.750		.533
21	I seek clarification of the meanings of another's opinion or points of view.		.645			.642		.567
8	I figure out a person's purpose in asking a question.		.585		excluded			.624
17	I clarify my thoughts by explaining to someone else.		.521			.529		.564
9	I clarify the meaning of an individual's gesture in a given situation.		.490			--		.321
15	I restate another person's statements to clarify the meaning.	-	-	-		.551		.578
10	I try to identify the social importance of the information presented in the texts.			-.732			.633	.496
14	I identify the author's views and intentions in the issue presented.			-.705			.729	.501
13	I examine the values rooted in the information presented.			-.624			.543	.435
11	I look for various relationships among concepts to understand the meaning.			-.462			.518	.383
16	I figure out an example which explains the concept /opinion.	-	-	-			.486	.410
19	I look for analogies of the words and concepts to clarify meaning.	-	-	-			.481	.403
18	I try to differentiate between opinions and ideas to remove ambiguity.	-	-	-			-	.509
	Percent of Variance : 38.124%							
	Reliability (Core scale) α = 0.826				0.766	0.612	0.678	

*Extraction Method: Principal Component Analysis. Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 6 iterations.
*Cat –Categorization (7items); * DS – Decoding Significance (6 items); * CM – Clarifying Meaning (4 items).
Comu h^2: Communalities. Better solution with exclusion of items 8. (**17 / 20 items retained**).

independently on to three factors and were very close to the pre-designated constructs. With this sample, item loadings revealed more homogenous loading with all the three factors of the core cognitive skill interpretation. Loading weights were also found to be remarkably higher for some items than in the Indian sample total as well as the two subsamples separately analysed (refer to Table 4.9.1 in Appendix O). The sampling adequacy was tested with a value of KMO = 0.863 and a significant Bartlett's Test for sphericity $p < .001$, showing sufficiently large inter-item correlation. The rule of thumb for sample size to run EFA is five to ten participants per variable (Field, 2009), which was fairly satisfactory with the Canadian sample when the core scales were separately analysed. The variance explained by these factors was 51.01% with communalities ranging from 0.327 to 0.641, which were much larger than for the Indian sample. The total Indian sample, and the two subsamples and the Canadian sample revealed similarities in factor loadings to a great extent.

In this CS Interpretation, item 8 "*I figure out a person's purpose in asking a question*" showed a reliability value < 0.3 and the item alpha (0.583) very close to the overall alpha (0.595) for the sub-scale. Hence, the EFA was repeated with the exclusion of this item and yielded loadings with increased weights for the items 3, and 7, in F1, 17 and 20 in F2 and items 11 and 14 in F3. Loading weights decreased for items 1, 2, 4 and 5 in F1, 21 in F2, and 10 & 13 in F3. There were no changes found in the variables to factors. However, three items 15, 16 and 19 that did not reach the cut-off point in first testing loaded on to F2 and F3 (Table 4.9). A three factor solution was more clear with the Indian sample with the exclusion of item 8. However, the items loading with Canadian sample revealed a better consistency with the construct definition. Item 8 that was excluded from Indian sample showed a loading weight 0.615 onto the designated construct with the Canadian sample.

Thus, The CS 1 Interpretation suggested retaining 17 of the 20 items. Results of the EFA for the Canadian sample (Table 4.9.1) can be found in Appendix O.

Core Scale 2 Analysis. Principal Component Analysis was performed on the 18 items with the Indian sample total and separately for subsamples 1 and 2 and the Canadian sample. Similar to CS 1 all the assumptions were tested and results found to be within acceptable levels. (Appendix O 1 for a comparison of values of all scales for KMO, Bartlett's test, factor extraction both varimax and oblimin solutions and variance explained for the entire sample- India total, India subsamples 1 and 2, and Canada sample). For CS 2, PCA extracted three factors and varimax failed to rotate in 30 iterations. Oblimin rotation produced a three factor solution and accounted for 42.555% of the variance. The communalities and pattern matrix with factor loadings from the oblimin rotation is presented in the Appendix O 2, Table 4.10.

For the India total, a three factor solution closer to the pre-designated two sub-scales was found with Factor 1 loading with 8 items labelled *analyzing arguments* (SS 6), Factor 2 loading with 5 items labelled as *examining ideas* (SS 4) and Factor 3 with one item *detecting arguments* (SS 5). The inter-item correlation for item 33 (loaded in Factor 3) in reliability revealed < 0.3 and analysis with exclusion of item 33 did not improve the solution, hence, this item was retained as it had a high loading weight (Table 4.10.1, Appendix O 3). However, this item was not included in the reliability analysis or Confirmatory Factor Analysis as it was a single item loaded in one factor. The items that loaded on Factors 1 and 2 were found to be more homogeneous for the subscales. Fourteen of the 18 items were retained. Four of the items from the original subscale *detecting arguments* loaded into *analyzing arguments*, and it is possible that these items were not appropriately worded for discrimination between the two subscales.

Principal Component Analysis was performed with the Canadian data for the CS 2 Analysis (Appendix O 4 - Table 4.10.2) revealed interesting results. Factor extraction with oblimin rotation accounted for 57.367% variance and communalities ranged from 0.415 to 0.791. Seventeen out of 18 items independently loaded on to Factor 1 (4 items), Factor 2 (8 items) and Factor 3 (5 items). One variable loaded on to two factors and one resulted in zero loading. However, the loading patterns were different, the loading weights were far greater than the Indian sample and more or less closer to the designated construct.

Core Scale 3 Evaluation. Principal Component Analysis with varimax and oblimin rotation extracted two factors with a variance of 40.22% and the variables to factors maintained homogeneity. The loadings were quite close to the predefined dimensions of the sub scales. Factor 1 loaded with 7 items labelled *assessing claims* (SS 7) and Factor 2 with three items labelled *assessing arguments* (SS 8). Ten of the 13 items could be retained after reliability check. For item 46 "*I rarely examine the flaws in an argument*" the inter item correlation was < 0.3, hence, analysis was performed with inclusion and exclusion of this item which resulted in negligible changes. It was decided to keep this item for its theoretical importance. Analysis with the Canadian sample found independent loadings of 7 items into Factor 1, three items into Factor 2 and three items loaded on to both the factors. Most of the items had good loading weights. However, the pattern of loading differed from that for the Indian sample. Factor loading for the Indian sample with inclusion and exclusion of item 46 and a comparison of items loadings with Canadian sample is presented in Table 4.11 (Appendix O 5).

Core Scale 4 Inference. Principal Component Analysis with the Indian samples using oblimin rotation extracted three factors and accounted for a variance of 40.23%. Factor 1 accounted for 28.925%, Factor 2 and 3 with 6.006 % and 5.300% respectively. Factor loadings and communalities for a comparison with Canada are presented in Table 4.12 in Appendix O 6. Factor 1 loaded with 10 items and was labelled *drawing conclusions* (SS 11), Factor 2 with 7 items was labelled as *querying evidence* (SS 9) and Factor 3 loaded only one item. All items loaded independently to each factor and four items showed zero loading. All items that loaded on to Factor 2 had very close proximity to the pre designated subscale (SS 9). Eighteen out of 22 items revealed good loading weights. The decisions on retaining items depend greatly on reliability testing. The subscale *conjecturing alternatives* (SS 10) consisted of seven items where five items got loaded into *drawing conclusions* (Factor 1). It was evident from the values of content validity analysis most of these items revealed a range value between 3 and 5 which indicated item ambiguity although these items showed high values of *I-CVI* and Aiken's VI_k. There is evidence of mixed understanding in this sub scale. This subscale did not emerge as a Factor with the item loadings in EFA. For this scale, Canadian sample loadings revealed almost similar results except that Factor 3 loaded with two items.

Core Scale 5 Explanation. Principal Component Analysis with varimax rotation extracted four factors (fourth factor with two items). The fourth factor was not an interpretable solution and did not have an adequate number of variables (minimum number of variables to factor is 4:1) for a factor. Hence, oblimin with 3 factor solution was tried and was useful and interpretable. Table 4.13 (Appendix O 7) shows the factor loading. In this core scale, Factor 1 loaded with items from both the subscales *justifying procedures* and *stating results* which were appropriate to label as *stating results* (SS 12). However, three of the items from *justifying procedures* loaded into Factor 2 with large regressions weights and were labelled accordingly as *justifying procedures* (SS 13). All six items that independently loaded on to Factor 3 were from the pre-designated subscale *presenting arguments* (SS14). There was no overlap in the item loadings both in the Indian samples. For this scale, the Canadian sample had more appropriate items loadings to factors. That is, 8 items loaded into Factor 1 with an average loading weight of 0.642, 6 items into Factor 2 with an average loading of 0.714 and 5 items into Factor 3 with an average loading weight of 0.615. However, the pattern of loadings demonstrated a difference between the samples. Eighteen of the 22 items are useful for retention.

Core Scale 6 Self-regulation. Principal Component Analysis with varimax rotation extracted two factors. Factor 1 accounted for 38.044% of the variance and Factor 2 with 6.257% which together explained 44.301% of the variance. The scatter plot revealed the possibility of a more homogeneous convergence by oblimin rotation and generated a useful solution for interpretation. As seen from scatter plots the oblique rotation produced clear distance and linearity of the variables (Figures 4.5 and 4.6 in Appendix O 8). Ten items loaded on to Factor 1 labelled *self-examination* (SS 15) and five loaded on to Factor 2 labelled *self-correction* (SS 16). Fifteen of 16 items could be retained.

Most of the items loaded with large loading weights and also loaded independently. The average loading weight in F1 (varimax) was 0.615 compared to (oblimin) 0.645, for 10 items. The average loading weights in F2 for five items was 0.638 (oblimin) compared to 0.681 (varimax) for three items. The oblimin solution seems to have provided a more accurate measure of variables to the construct. The oblimin solution was used for interpretation and reliability analysis. EFA for CS 6 analyzed with Canadian sample had seven items each loaded into two factors. However, the loading pattern demonstrated a difference between the two samples. Table 4.14 in Appendix O 9 gives a comparison of for the factor loadings for this scale.

Overall findings of EFA. To sum up, a total of 111 items across the six scales were subjected to EFA and interpreted separately. Principal Component Analysis for data reduction was performed for the six core cognitive scales of CTS construct and revealed encouraging results. Two subscales ended with a single item and they were excluded from further analysis. Although EFA was completed for the total Indian sample and subsamples and the Canadian sample, only the results for the total Indian sample were used for retaining the items and for further reliability analyses. The final number of retained items was 90 and can be found in Table 4.15.

Table 4.15

*Number of Items Retained for Core Cognitive Skills and Sub-skills after EFA**

Core Cognitive Skills	Sub Skills	Number of Items for EFA	Items Retained after EFA
1. Interpretation	1. 1 Categorization	7	7
	1. 2. Decoding significance	6	6
	1. 3. Clarifying meaning	7	4
2. Analysis	2.1. Examining ideas	6	5
	2.2. Detecting arguments	6	1**
	2.3. Analyzing arguments	6	8
3. Evaluation	3.1. Assessing claims	5	7
	3.2. Assessing arguments	8	3
4. Inference	4.1. Querying evidence	7	7
	4.2. Conjecturing alternatives	7	1**
	4.3. Drawing conclusions	8	10
5. Explanation	5.1. Stating results	5	9
	5.2. Justifying procedures	9	3
	5.3. Presenting arguments	8	6
6. Self-Regulation	6.1. Self-examination	9	10
	6.2. Self-correction	7	05
	Total	111	90

*EFA retained 90 items. **Items not included for reliability and CFA.

Reliability of the Scales after Exploratory Factor Analysis

The reliability analyses for the retained 90 items after EFA in each of the core and sub scales were computed and presented in Table 4.16 (Appendix O 10). All six core cognitive scales had acceptable levels of reliability, that is, greater than 0.7. Of the fourteen subscales, 10 subscales had reliability greater than 0.7 (range 0.702 to 0.876) and four sub scales had a value below 0.7 (range 0.606 to 0.678). The reliability of the total 90 item scale CTSAS for the Indian sample obtained a Cronbach's α value of 0.960 and the α value for the Canadian sample was 0.975. Thus, preliminary psychometrics of the CTSAS were encouraging. The researcher was hopeful of deriving further insight into the construct validity of the scale with the results of Confirmatory Factor Analysis.

Step 4: Confirmatory Factor Analysis (External Validation)

Confirmatory Factor Analysis (CFA) with the Canadian sample (N =144) was performed to examine the goodness of fit indices for the variables to the factors in order to support the construct definition, in other words to establish construct validity. Confirmatory Factor Analysis is useful in testing hypotheses of relationships between observed variables and their underlying constructs. In the CFA analysis, Structural Equation Modeling (SEM) with Maximum Likelihood (ML) estimation was used to determine the adequacy of model fit to the data. For conducting the CFA the researcher used Analysis of Moment Structure (AMOS 18), which is provided with PASW Statistics Grad Pack 18 (SPSS). The criteria used for assessing model fit for the data are presented in Figure 4.7 (Appendix P). The six core scales (90 items) were separately subjected to CFA with the Canadian sample.

Null Hypothesis. $H_0 = \mathcal{E} \geq 0.05$ **(**Hypothesis of "not-close fit" in the population. The assumption is that the specified model holds in the population). The "Maximum Likelihood Factor extraction estimates population values for factor loadings by calculating loadings that maximize probability of sampling the observed correlation matrix from a population" (Tabachnick & Fidell, 2007, p.63).

The hypothesised models with the results are presented in Figures 4.8.1 to 4.8.6. Figures 4.8.1 to 4.8.3 are found adjacent to the appropriate text and Figures 4.8.4 to 4.8.6 are placed in the Appendices P 1 to P 3. Loadings of variables (items) with EFA from Indian sample were closer to the 'a priori' constructs, which were fixed to develop the path diagram in AMOS graphics. Data from the Canadian sample were screened for the assumptions of CFA which were similar to those of the EFA. There were no missing data. Normality of observed variables was assessed through examination of histogram and descriptive statistics. None of the observed variables was skewed or kurtotic. The data were fairly correlated and no multicollinearity was observed from the correlation matrix. The data satisfied the assumptions for running the CFA except for sample size and probability sampling, as the minimal optimum sample size required for achieving a good model fit would be 200. Assumptions related to sample size and sampling are discussed in great detail with the results and interpretation of CFA.

All six model path diagrams were developed with AMOS graphics and the analyses were run to estimate the chi-square values, Comparative Fit Index (CFI), and Root Mean Square Error Approximation (RMSEA) with lower and higher limits of the 90% confidence interval. The criteria used for assessing model fit are presented in Figure 4.7-Appendix P. A series of CFA was conducted using the AMOS program. All models were fit using Maximum Likelihood parameter estimation. The fit indices for each of the six core scales with their latent variables are presented in Table 4. 17. The model fit for each of the scales is described.

Results of the Confirmatory Factor Analysis

The hypothesized model of "not-close fit" is $H_0 = \mathcal{E} \geq .05$, that is, a sufficiently low value of chi-square (non-significant) results in rejection of the model, which therefore supports the alternate hypothesis (MacCallum, Browne, & Sugawara, 1996). As seen from the Table 4.17, the chi-square values for all six scales were significant. The chi-square values for CS 1 Interpretation χ^2(116, N = 144) = 198.630, $p < .01$, robust CFI = .89, RMSEA = .071; CS 3 Evaluation χ^2 (64, N = 144) = 50.076 $p < .01$, robust CFI = .964, RMSEA = .058; and the CS 4 Inference χ^2 (118, N = 144) = 208.493, $p < .01$, robust CFI = .902, RMSEA = .073, were close fit models.

A non-significant χ^2 makes a good fit model and could result in rejection of the null hypothesis. In this case, as stated by Bentler (1995), most often with large sample size the χ^2 is significant and with a small sample the assumptions of the χ^2 test reveal an inaccurate probability

Table 4.17

CFA - Model Fit Indices for the Six Core Cognitive Skills

Core cognitive skill Models	Index of Fit						
	Ch-Square (χ^2)	χ^2/df ratio	CFI***	RMSEA***	90% Conf. Int. L	H	Fit/ No Fit
CS1Interpretation	**198.630 *df* 116	**1.712**	**0.890**	**0.071**	.054	- .087	Fit
CS2 Analysis	**196.647 *df* 64	3.088	0.837	0.121	.102	- .140	No Fit
CS3 Evaluation	*50.076 *df* 34	**1.47**	**0.964**	**0.058**	.01	- .090	Fit
CS 4 Inference	**208.493 *df* 118	**1.767**	**0.902**	**0.073**	.057	- .089	Fit
CS 5 Explanation	**383.869 *df* 132	2.908	0.819	0.116	.102	- .129	No Fit
CS 6 Self-Regulation	**251.815 *df* 89	2.829	.860	0.113	.097	- .130	No Fit

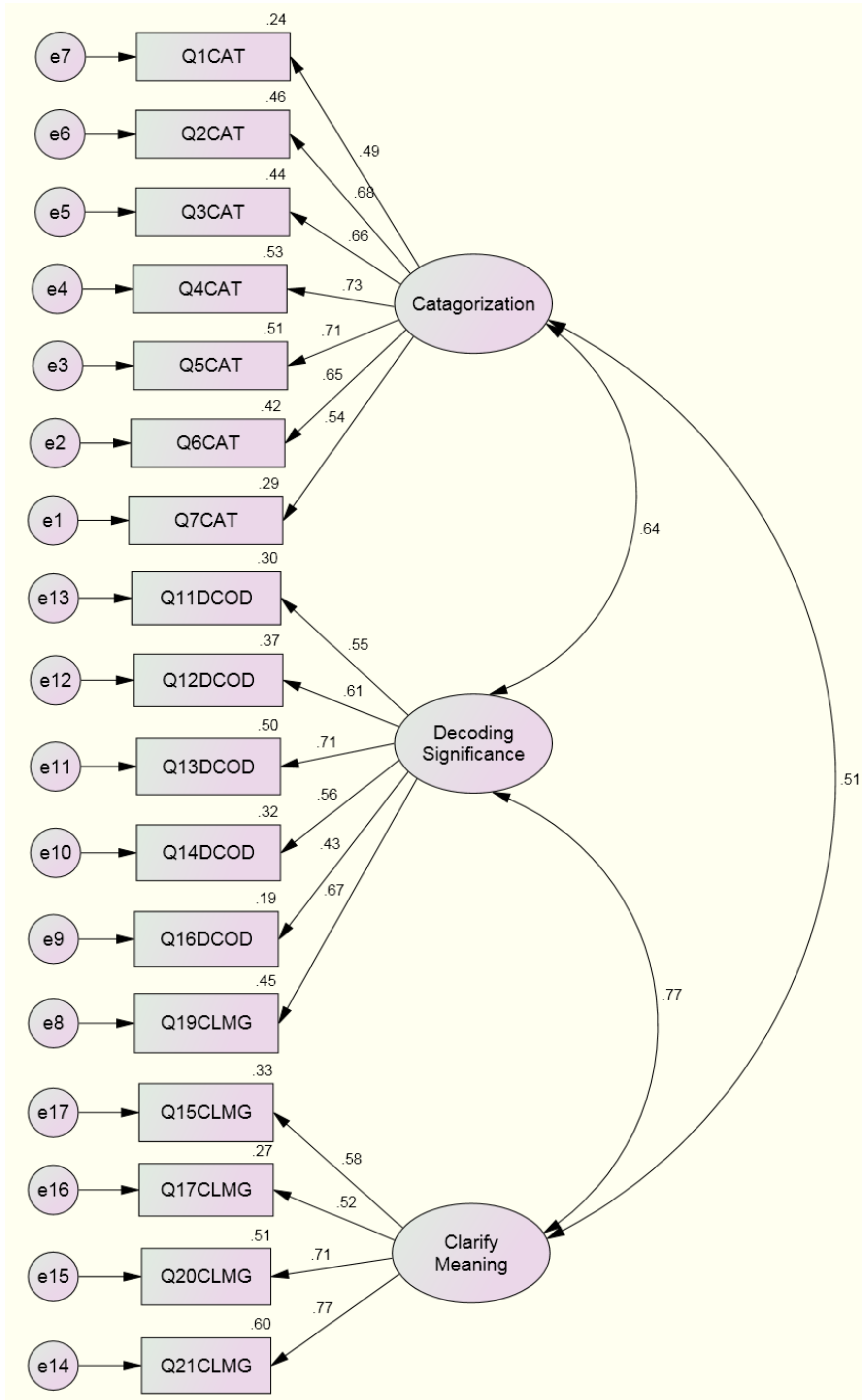
***CFI-Comparative Fit Index, ***RMSEA-Root Mean Square Error Approximation.
** Significant p= <.000 *Sig p= <.05

levels. Hence, "a rough "rule of thumb" directly related to the χ^2 value is that a good fitting model may be indicated when the ratio of the χ^2 to degrees of freedom is less than 2" (Tabachnic &b Fidell, 2007, p. 715). In the case of these three scales - Interpretation (χ^2/df = 1.712), Evaluation (χ^2/df = 1.47), and Inference (χ^2/df = 1.767) the ratio is less than 2 and strongly supports that the three scales are close fit models. (Figure 4.8.1 to 4.8.3 shows the graphical presentation of the good fit models).

The results indicated that the evaluation scale model was a very good fit and the other two scales also had good fit indices although the CFI values did not reach the acceptable value of 0.95. Further, examination of the parameter estimation strongly supports the model fit as evidenced from data presented in Figure 4.8.1, 4.8.2, and 4.8.3, as strong relationships emerged among the variables and factors indicating loadings of homogeneous variables on those subscales. There exists a relationship among factors (subscales), which is neither too high nor too low indicating the subscales are interrelated. Thus, the results indicated strong support for the construct definition. Figure 4.8.1, 4.8.2, and 4.8.3 show the item loading with the regression weights. Thus, the EFA structure from the sample was consistent with the results of CFA for the three scales and it is plausible that one could replicate a similar structure in a comparable population.

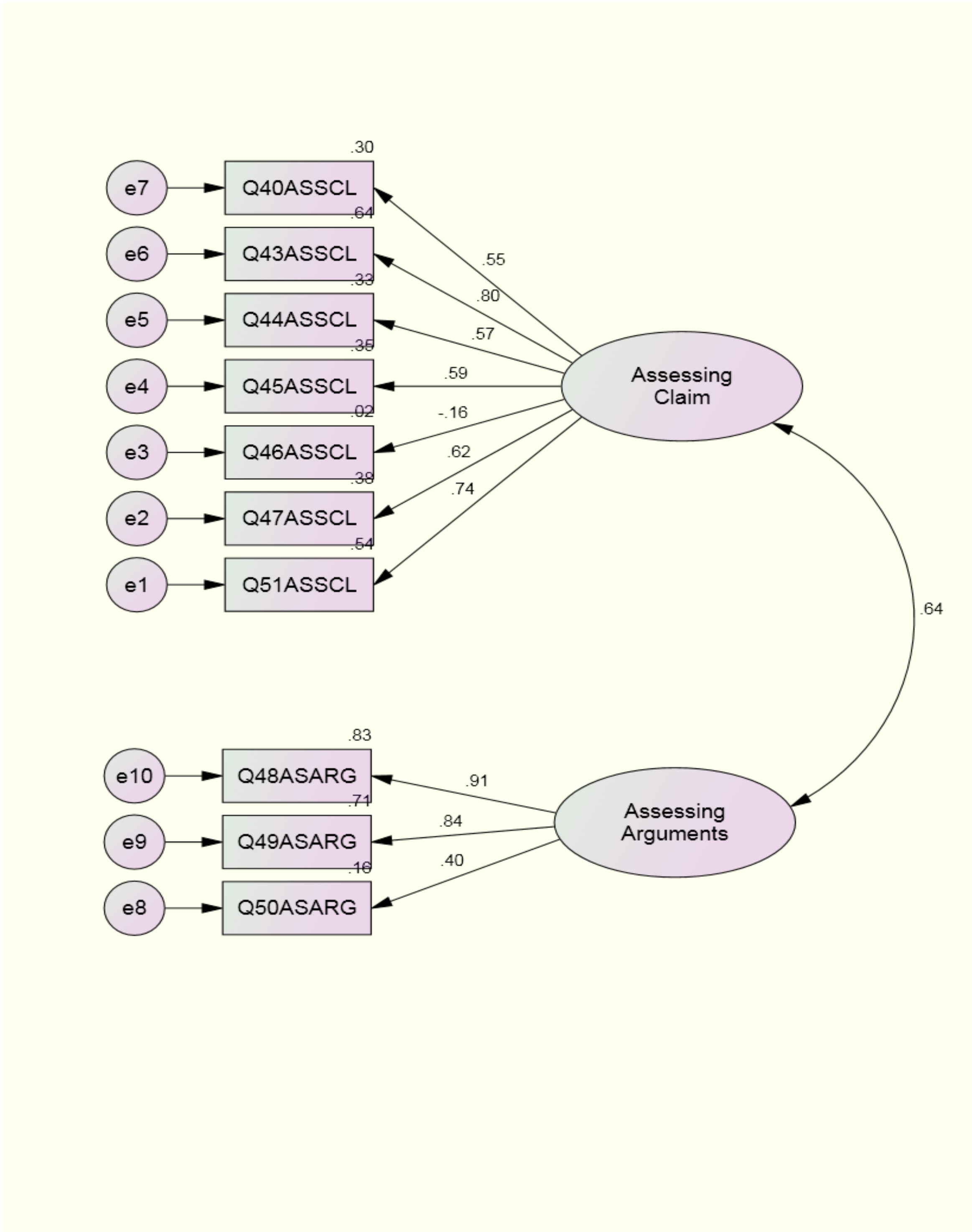
The remaining three scales CS 2 Analysis χ^2 (64, N = 144) = 197.647, p < .01, robust CFI = 0.837, RMSEA = 0.121; CS 5 Explanation χ^2 (132, N = 144) = 383.869, p < .001, robust CFI = 0.819, RMSEA = 0.116; and CS 6 Self-regulation χ^2 (89, N = 144) = 251.815, p < .001, robust CFI = 0.860, RMSEA = 0.113 demonstrated poor fit models with large chi-sqauare values, large error variances as indicated by the values of RMSEA, and closer to the acceptable levels of Comparative Fit Indices. The χ^2 to *df* ratios were more than 2 as seen in Table 4.17.

Figure 4.8.1 AMOS Results of Confirmatory Factor Analysis Model for CSI: Interpretation



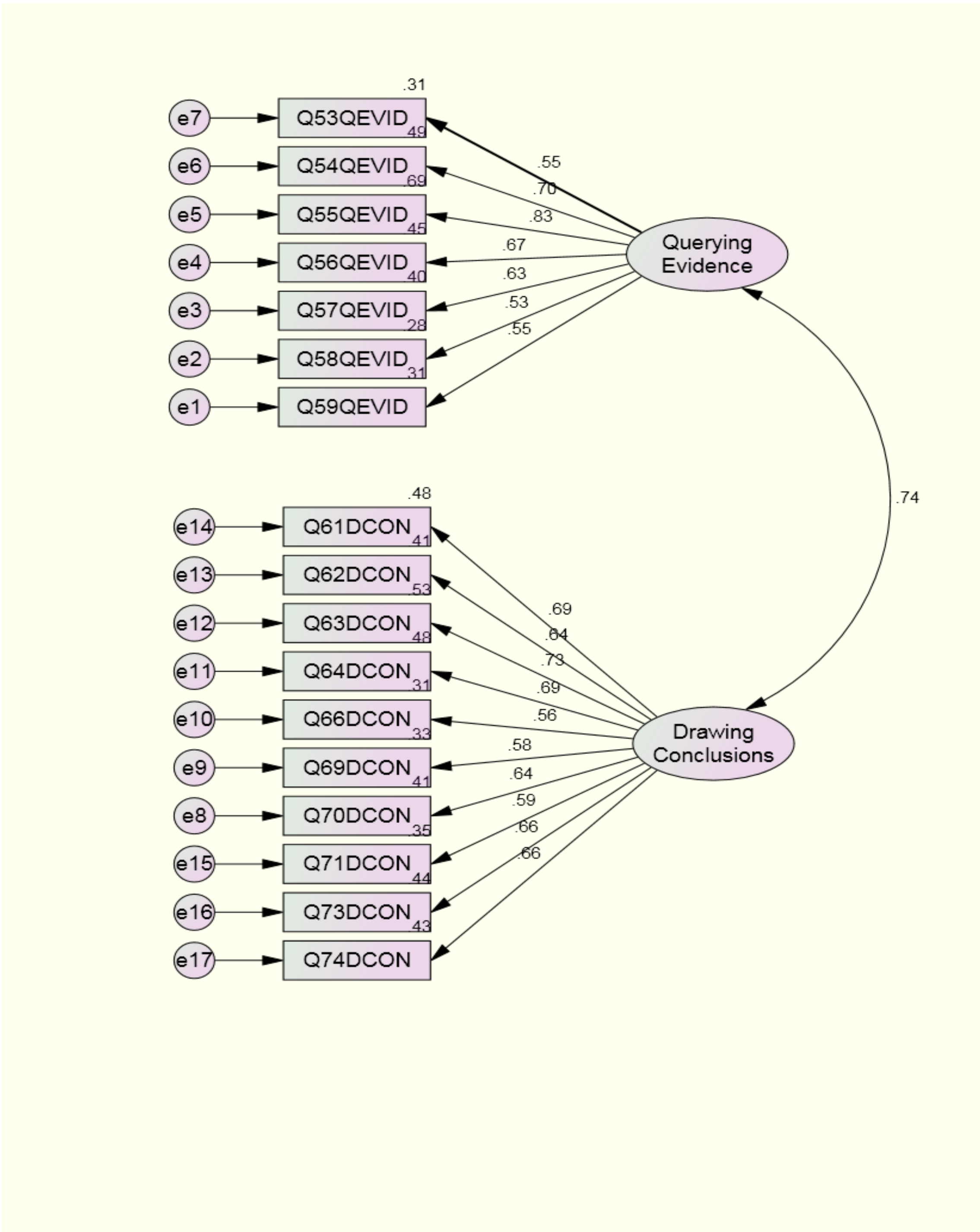
Note: CS 1 Interpretation turned out to be a good fit model.

Figure 4.8.2 AMOS Results of Confirmatory Factor Analysis Model for CS 3: Evaluation



Note: CS 3-Evaluation revealed best fit model.

Figure 4.8.3 AMOS Results of Confirmatory Factor Analysis Model for CS 4: Inference



Note: CS 4 Inference was a good fit model. One of the subscales failed to emerge with EFA.

As shown in Figures 4.8.4, 4.8.5 and 4.8.6, (Appendices P 1 to P 3) the variables for these scales were strongly correlated and had heavy regression weights. These values indicated that the variables were homogeneous and closer to EFA factor structure and therefore were more or less consistent with the construct definiton. The variables seem to be a probable measure of the construct yet demonstrated a poor model fit which may be attributable to inadequate sample size or lack of probability sampling and the obtained large error variance ($1\text{-}R^2$). For example, for the variable 89 (*I write essays with adequate arguments supported with reasons for a given policy or a situation*) in the SS *Justifying Procedures* (Figure 4.8.5 in Appendix P 2), the value of $1\text{- }R^2$ (1 - 0.08) the unexplained variance was very high (0.92). Similarly, occasional high error variances were observed for other factors and sub scales.

Interpretation

Psychometricians purport that for a true model, the χ^2 has an expected value equal to the degrees of freedom and does not vary with sample size (Marsh, Balla & McDonald, 1988). The authors further asserted that the χ^2 can be made small by reducing the sample size. Because of this influence a poor fit model with small sample size might result in a non-significant χ^2, where as a good fit model with large sample size may become statistically significant. Thus, “testing models with large sample is always desirable” (Hoelter, 1983, p. 328). In particular, a large sample size is recommended for CFA as the χ^2 provides a test where the residual difference between the sample (S) and population (Ɛ) converge in probability to zero as the sample size approaches infinity (MacCallum, Browne, & Sugawara, 1996; Tucker & Lewis, 1973). The minimum sample size thus suggested for CFA is 200. Increasing the parameters necessarily results in better (lower) chi-square values (MacCallum, et al.). Though there are varied opinions and evidence regarding sample size and there are programs that can be used to conduct small sample CFA, the present study with a sample of 144demonstrated path diagrams with more than adequate regression weights and high correlations between variables and factors, and reasonable correlations among factors. The model fit obtained and the regression weights of the variables showed an almost consistent factor structure with the EFA factor structure in factor loadings. A minimum sample of 200 would have yielded better results.

Three scales demonstrated poor model fit whilst showing the fit indices values closer to acceptable levels. However, the item loadings revealed large regression weights, in some cases better than EFA loading weights. For example, in the CS 2 Analysis, the CFA loadings were much higher than the EFA loadings. In the EFA, a factor emerged with one item and the other four items from this pre-designated scale loaded on to second factor namely "*analyzing arguments*". Thus, it is probable to conclude that these items did not discriminate well and are possibly not the appropriate measure of the construct. This warrants future scrutiny of these items.

The EFA loading in CS 2 for Canadian sample had four items loaded onto Factor 1 with an average loading weight of 0.747, eight items loaded onto Factor 2 with an average loading weight 0.611 and only one item loaded onto Factor 3 with a loading weight of 0.549. The two SS emerged with ≥ than adequate number of variables and fairly average loading weights. These findings seem to ponder whether or not the factor structure revealed in EFA was a rightly specified (identified) model or a mis-specified model. To check this one could try an alternate model by fixing these item loadings from EFA with the Canadian sample in a path diagram for testing the goodness of fit. However, this would require either a fresh sample or the availability of a large Canadian sample and random sampling, neither of which were available.

There are variables that resulted in low loading weights in the CFA although all the items loaded in the EFA were > 0.45. For example, in the CS Interpretation, the item 16 the loading weight was 0.43, item 46 in CS Evaluation was -0.16, and the error variance for this variable was 1- R^2 (1- 0.02= 0.98),which is very high, and for item 50 the loading weight was 0.40. Another possibility is that some of these may be mis-specified models requiring further evaluation. Similarly, scales that were identified as the poor model fit had very high standardized loading weights as high as 0.91. However, some of these subscales had only three variables. This phenomenon to some extent explains the reason for a large chi-square value as the reduced number of parameters in the model provides the large value.

For example, the CS 5 Explanation, although it emerged with three factors in the Indian sample, Factors 1 and 3 had 9 and 5 items respectively and Factor 2 loaded with three items, and their average loading weights were Factor 3 with 0.707, Factors 1 and 2 with 0.579 and 0.642 respectively indicating strong correlations with variables. The Canadian sample for this CS showed a better pattern of loading, that is, 8 items loaded into Factor 1 with an average loading of 0.614, and 6 and 5 items loaded into Factor 2 and 3 with average loading weights of 0.714 and 0.615 respectively. However, with large loading weights and number of items to factors the model could not achieve goodness of fit index. There is a need for further testing of the model with the alternate model that derived from the EFA structure with a new Canadian sample. Thus, the poor fit model result of CFA may be the consequences of the unmet assumption related to probability sampling and sampling adequacy for performing CFA.

However, in the subscale inter-factor correlations, all factors demonstrated a correlation above 0.6 except for one scale. A possible explanation is that the correlation between SS *justifying procedures* and SS *presenting arguments* was 0.37. This value might be appropriate as these items were more or less independent in nature since they are entirely different tasks.

It is reasonable to conclude that this study is a promising preliminary evaluation of psychometrics properties of the CTSAS. The EFA structure and the loadings were very appealing. The results of the CFA further support the construct's structural validity for the three construct as H_0 of "not-close fit" = $\varepsilon \geq 0.05$ was rejected in favour of the alternate hypothesis that the three scales are consistent and a close fit to the defined construct. The results also support the assumption that these specified models prevail in the population and can be replicated in similar samples. Further, these results support the construct definitions purported by APA (1990).

Step 5: Convergent Validity (External Validation)

Convergent validity refers to establishing relationships between related constructs or assessing two groups that would be expected to perform similarly on the measure (Hinkin, 1995). There is no gold standard measure that could be used as a comparable instrument for measuring critical thinking to realistically test convergent validity. The Need for Cognition Scale (NCS-SF) developed by Cacioppo and Petty (1982) was used to test the convergent validity of the CTSAS. This is a closely related scale to the CTSAS as the need for cognition is a prerequisite for anyone who has aspirations to develop and practice CTS. For the purpose of this study, the convergent validity was tested only for the Canadian sample.

Description of the Need for Cognition Scale

The Need for Cognition Scale-Short Form (NCS-SF) is an 18 item instrument for measuring an "individual's tendency to engage in and enjoy thinking" (Cacioppo, Petty, & Kao, 1984, p.116). The original scale had 43 items. Four studies with diverse cultural groups (Anglos, Hispanics, and Spanish) such as working groups, university faculty, students etc., formed the samples for the refinement of the scale. A series of EFAs and CFAs conducted with different samples resulted in 18 items. The scale had nine negatively worded statements and nine positively worded statements. The negatively worded statements were reverse scored and employed a 5 point Likert-Type scale ranging from "*extremely unlike me, unlike me, neutral, like me to extremely like me" (*Culhane, Morera, & Watson, 2006, p. 57). Appendix Q shows permission to use the NCS-SF scale.

This scale has been tested on undergraduate students across United States and Canada. The scale was tested with several models with a variety of samples in different languages and cultures. All results yielded acceptable levels of goodness of fit indices and acceptable levels of reliability (Culhane, Morera & Watson, 2006). Findings by Cacioppo, et al. (1984); Culhane, Moreara and Hosch (2004); and Sanders, Gass, Wieserman and Bruschke (1992) estimated coefficient alphas of 0.90, 0.86, 0.88, respectively, suggested high internal consistency for the measure. For this scale, the Canadian sample of students in the previous studies had a coefficient alpha in two groups 0.65 and 0.78 respectively (Cacioppo, Petty, Feinstein, & Jarvis 1996).

This scale was placed at the last section of the CTSAS questionnaire and students were alerted to pay attention to the 5 point scale as the first part of the CTSAS was a 7 point response continuum. The reliability analysis for NCS –SF in this study yielded a better $\alpha = 0.831$ in the Canadian sample than in previous reported studies and a 0.704 in the Indian sample.

Results and Interpretation

Bivariate Pearson correlation (*r*) and Spearman Brown (*rho*) performed for the retained 90 items in the CTSAS and NCS-SF (18 items) with Canadian sample of 144 revealed both significant and non-significant results. A comparison of these results is presented in Table 4.18 (Appendix R). The results of the correlation statistics with two methods seem similar. The six core cognitive scales except Explanation (CS 5) had a significant correlation with NCS-SF. The CS 5 and the subscales did not show acceptable levels of correlation with the NCS-SF. Similarly, two subscales from CS Interpretation were found to be non-significant. However, this is a preliminary testing, and when refined, the researcher would expect stronger convergence with the CTSAS. No discriminant validity was tested at this stage.

SECTION 5

SUMMARY, DISCUSSION, LIMITATIONS,

NEXT STEPS,

IMPLICATIONS AND CONCLUSIONS

Summary and Discussion

The study focused on establishing some aspects of Messick's unitary concepts of construct validity framework such as content validity, reliability, structural validity and convergent validity for the CTSAS. Criterion validity, discriminant validity and consequential validity were not considered within the scope of this preliminary evaluation.

The development of the CTSAS was a two stage process of collecting evidence for validity of the scale on the basis of Messick's (1990) unitary concepts of validity. The scale was initially developed through a process of extensive literature review to identify the construct on which to base the scale. Item writing for the scales related to the constructs was initiated during the researcher's doctoral course work. The researcher developed a total of 196 items, peer reviewed and modified, for the six core scales, which included 16 subscales. The first stage involved establishing content validity of CTSAS. The second stage established evidence for the construct validity of the scale. Thus, the scales were examined for psychometric characteristics employing strong statistical techniques to ensure statistical conclusion validity.

Psychometric Evaluation of the CTSAS

Content validity. Content validity of the scale included content relevance that the items reflect the content domain (DeVellis, 2003). The main purpose of ensuring content validity is to minimize potential error variance and increase support for construct validity (Dunn, Bouffard & Rogers, 2000; Hayens, Richard & Kubany, 1995; Messick, 1989). The most convincing approach for assessing content validity is by expert panel ratings. The content validation was a rigorous process using 17 experts from multi-disciplinary fields which resulted in reducing the original 196 items to 115 using stringent statistical criteria for retaining items in the scale. The main criteria used were Aiken's (1985) Content Validity Coefficient (VI_k) and Content Validity

Index (*I-CVI*) (Polit & Beck, 2008). According to Aiken's right tailed probability table, the items that had a VI_k value significant at $p < .05$ and items that showed a value *I-CVI* > 0.75 were retained and some of the ambiguous items that showed a range value > 3 were modified (acceptable level ≤ 2.75). After content validation, 115 items across six core cognitive scales, which were further classified into16 sub scales, were retained in the CTSAS.

Structural validity. For the purpose of construct (structural) validity, the scale was tested using two groups – one sample from Canada and the other from India. The scores obtained were subjected to exploratory data screening which resulted in exclusion of four items from the scale retaining 111 items, and 19 cases from the Indian sample (887 to 868). There were no missing values and no additional skewed or kurtotic items for the Canadian sample. The six core cognitive scales demonstrated acceptable levels of reliability (Cronbach's α) although some of the subscales revealed low reliability. When compared with the reliability values of core scales of the CCTST (Facione, Facione, Blohm & Gittens, 2008) which is based on the APA definition of CTS, the CTSAS revealed better reliability for the core scales.

Data reduction strategies (EFA and CFA) . Exploratory Factor Analysis is a useful analytic method that can determine, empirically, how many constructs or latent variables or factors underlie a set of items (DeVellis. 2003). Exploratory Factor Analysis was performed for the six core cognitive scales separately, using both varimax and oblimin rotations with an absolute value set at 0.45 to suppress variables (items) for retention. The EFA further reduced the 111 items to 90 items across the subscales and demonstrated strong correlations of variables (items) to factors (subscales). These 90 items were fixed in six path diagrams to their 'a priori' latent constructs for performing Confirmatory Factor Analysis (CFA). The variance accounted by each scale included CS1- 38.124%, CS 2 - 42.55%, CS 3- 40.22%, CS 4 - 40.23%, CS 5 -

45.44%, and CS 6 - 44.30% respectively. Most of the regression weights for item loadings were above 0.510 except for seven variables across five core scales which had loading weights of 0.481and 0.486 (CS 1), 0.468 (CS 2) 0.454 (CS 3) 0.493 and 0.471 (CS 4), and 0.474 (CS 6).

The majority of the item loadings for the factors demonstrated homogeneity and consistency with the pre-designated constructs except for two of the core cognitive scales Analysis and Inference. In core scale Analysis (CS 2), one of the subscales (*detecting arguments*) failed to emerge as a factor and four of the items of this pre-designated subscale loaded into *analyzing arguments*. These items probably had similar descriptions and hence, did not discriminate between the two subscales. However, the CFA analysis revealed a poor model fit with a highly significant chi-square (χ^2) and χ^2 to degree of freedom ratio of 3.088. These results indicate either that the scale was a mis-specified model or sampling inadequacy or lack of random sampling, which resulted in high error variance (RMSEA 0.121). However, the EFA conducted with Canadian sample for CS 2 resulted in three distinct factors with large loading weights. Thus, it is crucial to examine further the structure obtained in EFA to be tested for goodness of fit with an alternate model in a different sample to ensure that the model was not a mis-specified one.

Similarly, in the core scale Inference (CS4), one of the subscales *conjecturing alternatives* did not emerge as a factor and five of the seven variables (items) from this subscale loaded on to "*drawing conclusions*". These items need revisiting, restructuring of the contents and re-testing to decide whether to retain the subscale '*conjecturing alternatives'* as part of the construct Inference or whether the items retained through EFA define more accurately the construct with exclusion of this subscale. Interestingly, the CFA conducted for this core scale revealed a good model fit (χ^2 = 208.493, *df* = 118, χ^2/ *df* = 1.767, *CFI* = 0.902, RMSEA = 0.073) with two subscales, showing structural fidelity.

The CFA performed for two other scales (CS1 Interpretation and CS 3 Evaluation) demonstrated homogeneity as most of the variables that loaded on to these subscales were consistent with the pre-designated construct or conceptual framework. These two scales revealed good model fit along with significant chi-square values and χ^2/ *df* ratios with acceptable values (1.712 and 1.47, respectively). The *CFI* for these scales were robust (0.890 and 0. 964), and the RMSEA was 0.071 and 0.058, respectively. For the remaining two core scales (CS 5 Explanation and CS 6 Self-regulation), the EFA demonstrated factor structures and item loadings almost similar to the pre-designated scales. The CFA revealed a poor model fit with large chi- square values and large error variances explained by the RMSEA and the *CFI* did not reach acceptable levels.

To sum up, the null hypothesis 'not-close to fit' with the population $H_0 = \mathcal{E} \geq 0.05$ for the three core scales (CS1, CS3 and CS4) was rejected in support of the alternate hypothesis and it may be concluded that fit of the model with the population was close (i.e., $\mathcal{E} \leq 0.05$). The fit indices for these three scales support the structural validity that the EFA structure emerged with the Indian sample is consistent with the results of the CFA performed with the Canadian sample and strongly indicates the structural fidelity of these three core cognitive scales. The results also support that the APA definition of these constructs is valid and such results can be replicated in comparable populations.

In terms of the other three core scales (Analysis, Explanation, and Self-regulation) while they demonstrated good regression weights showing strong relationships of variables to factors and factor correlations, the error variances demonstrated by RMSEA were very high for a few of the variables (items) along with the estimation of large values of chi-square (χ^2). The probable reasons for such a finding might be: (a) the assumptions of sampling adequacy for conducting the CFA were not satisfactorily met, (b) random sampling of subjects were not possible as participation was voluntary and the researcher had to include all of the available Canadian sample in order to obtain the required sample size, thus, the sampling error could have caused the high error variance, and (c) there were fewer parameters in the model or the EFA factor structure could be a mis-specification model. The findings of MacCallum, Browne, and Sugawara, (1996) purported that a lack of model fit in the sample arises from sampling error or small sample size. As seen from the yielded values of the *CFI* which were closer to acceptable level (0.95) for these three scales (CS 2- 0.827, CS 5 - 0.819 and CS 6 0 - .860), unmet sampling assumptions could be a reason for the large error variance and the resultant poor fit models.

For example, the trial run of EFA with Canadian sample derived three distinct factors with CS 2 Analysis and CS 5 Explanation with an adequate number of items and large average loading weights and items loadings were much more consistent with the pre designated construct. An alternate model would have been fixed with the path diagrams to test the model fit with these homogenous items derived from Canadian sample upon availability of large sample size or an entirely new sample. Thus, these findings indicate the need for further evaluation of the scale to clarify alternate explanations to the obtained results. Some of the extraneous variables' (age, cohort, language) influences on CTS may be examined for more valid explanations.

As seen from the demographic variable age, 35% of the Canadian sample was above 24 years as against 5% in Indian sample. The researcher at this point is unsure if age has any influence in the development of CTS which may have accounted for the obtained response resulting in differences in item loadings between the two samples. One of the assumptions is that CTS are developed over time as the individual goes through a program of study. The Canadian sample was comprised of only Year III and Year IV students compared tostudents from all four years in the Indian sample, which could explain the responses obtained. The interaction effects of these variables (age, gender and cohort) on CTS were not within the scope of this study. It is definitely an area need to be explored before the scale is said to be ready to use.

Reliability. The reliability analysis of the retained items from the EFA for the Indian sample for all the core scales ranged between 0.796 and 0.896. Three subscales had a reliability value below 0.70 (a value of 0.612 for SS 2 *decoding significance,* 0.647 for SS 8 *assessing arguments,* and 0.606 for SS 16 *self- correction*). High internal consistency reliability values (above 0.80) were obtained for several core scales and subscales, which could result in a reduction in the length of the scale. Such a decision should be made following an examination of communalities after extraction for each item. An item with a communality (h^2) value < 0.3 could be considered for deletion in order to shorten the scale. It is interesting to note that all of the items that loaded on the three subscales of Interpretation had communalities values > 0.3 and the SS *decoding significance* with a low reliability index also had communality values ranging from 0.383 to 0.501.

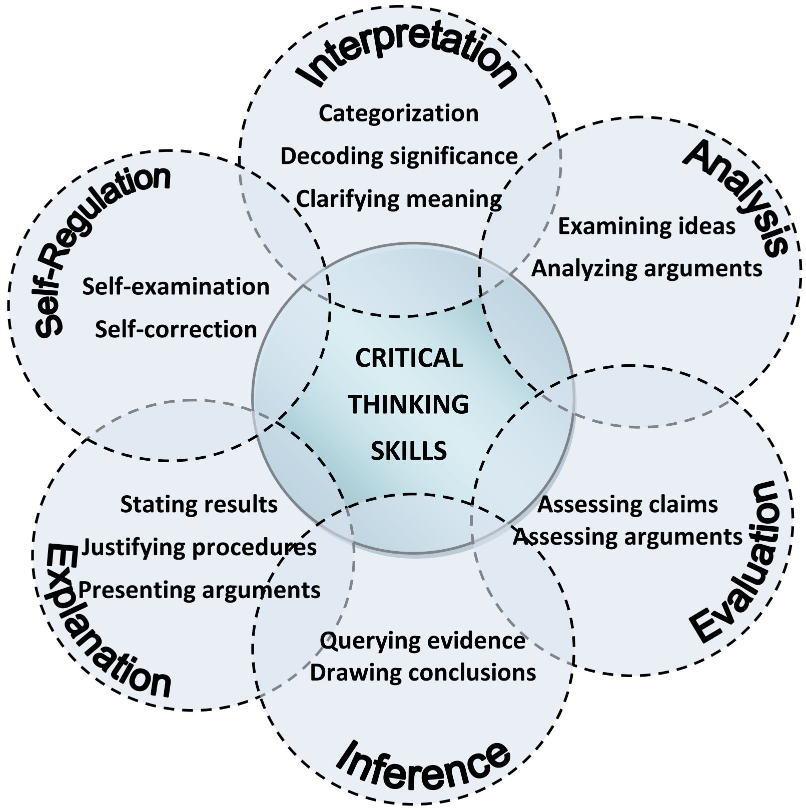
Subscale 8 *assessing arguments* in CS Evaluation revealed a range of communalities values between 0.398 and 0.619, and SS *self-correction* had communalities values ranging from 0.375 to 0.542. These values indicate that all the scales above had more than acceptable levels of communalities and deleting items may not be appropriate at this stage. The evaluation of the CTSAS can be repeated with different samples to assess and possibly reduce further the length of the scale. Appendix S contains the final items for all scales with loading weights, communalities, and Cronbach α values for each of the scales.

Convergent validity. The Pearson product moment correlation (*r*) values and Spearman (*rho*) values for testing convergent validity of the CTSAS (90 items) with the related construct Need for Cognition Scale (NCS –SF) using Canadian sample revealed fairly good results with significant values for five core cognitive scales. This indicated that most of the items in the subscale converged with the NCS-SF items. However, the CS 5 Explanation had non-significant values for the core scale and the subscales. The use of two methods Pearson *r* and Spearman *rho* values showed negligible difference. However, this is a preliminary evaluation of the CTSAS and more testing is needed for final decisions.

The findings of EDA support the revision of the four items related to socio-cultural relevance of CTS identified earlier. Similarly, the EFA results support re-examination of the items of the two subscales (*detecting arguments* in CS 2 and *conjecturing alternatives* in CS 4) that failed to emerge in the EFA structure. The CFA findings strongly support revision of three items (46, 89 and 109) which showed low regression weights and high error variances. The results of the reliability (α) testing, the EFA, and the CFA are encouraging, but not complete and the scale will require considerable testing using varied samples and settings. Further, using alternate theoretical perspectives for item reduction and including testing for convergent and discriminant validity, criterion validity, consequential validity, and social reliability will add to the strength and usefulness of the scale.

Reviewed conceptual framework. The development and evaluation of the CTSAS was based on the conceptual framework that assumed the six core cognitive skills and 16 sub skills that are interrelated and interdependent. The data reduction strategies applied to test the structural validity of the CTSAS resulted in retaining 14 of 16 sub skills (Figure 5.1). The two sub skills "*detecting arguments*" from CS 2 Analysis and "*conjecturing alternatives*" from CS 4 Inference failed to emerge with item loadings. However, the CS 4 Inference demonstrated a close to fit model with CFA with the remaining two sub skills (q*uerying evidence and drawing conclusions*). The results suggest further review and modification of the items and subsequent retesting of items for this scale. The proposed conceptual framework premises that the core cognitive skills and sub skills are interdependent and interrelated gained support as seen from the scatter plots. These premises are also supported by the acceptable levels of inter factor correlations and inter- item correlations. The homogeneous nature of the variables loading with large loading weights support strong relationships between variables (items) and factors (subscales) indicating these items are a more accurate measure of the constructs. The modified conceptual framework after EFA and CFA is presented in Figure 5.1. Although the end goal is to have an instrument that can be normalized to reflect the six stages of critical thinking (Paul & Elder, 2006), this normalization cannot be achieved until the scale is finalized.

Figure 5.1. Modified Conceptual Framework after EFA and CFA results



Note: CS 2-Analysis and CS 4-Inference retained two of the three subscales each
CS1-Interpretation, CS 3-Evaluation and CS 4- Inference were found to be close to fit models.
14 of the 16 subscales were retained.

Limitations

The primary limitation of the study was the inadequate sample size for the conduct of CFA. The consequences of this limitation have already been discussed. As noted earlier, the three poor fit models, in particular, CS 2 and CS 5 were probably the result of unmet sampling assumptions for performing CFA. The alternate model test was not tried because of unavailability of a sample as it is desirable to use a different sample to test CFA when items and scales derived from the EFA are performed with the same sample. The sample characteristics (age, gender, cohort variability) were not compared as the baseline data were not available for the Canadian sample to ensure whether the sample was representative of the target population. The age differences have already been discussed. The Indian sample (entire students from the five colleges studied) included all the four years of the study cohort whereas in Canada only Years III and IV students were in a university environment. The differences observed in the factor loadings of EFA point to possible influence of socio-cultural and demographic variables on critical thinking skills. This was not studied at this time as this is preliminary testing of the CTSAS. Additional testing and refining of items is required. The study is limited to testing some aspects (content validity, structural validity, reliability and convergent validity) of Messick's unitary concept of construct validity framework; discriminant validity, criterion validity and consequential validity were not attempted in this study.

Next Steps

The study aimed to establish preliminary psychometrics of CTSAS. Any scale construction involves a series of research processes for re-examining and refining the scale. Hence, several studies may be required in the future using diverse samples and settings for refining the scale. While different researchers may use somewhat different paths, recommended steps necessary for further refinement of the CTSAS scale are presented.

Both the Indian and Canadian samples demonstrated similar results in the EDA and resulted in exclusion of four items that were related to socio-cultural aspects of critical thinking. These items need to be revised by either changing the structure of the items, wordings or language used in writing those items, and re-testing them in future in similar samples to ensure socio-cultural relevance of the scale.

The newly constructed 90 item scale should be retested using the similar samples and analyzing them using Item Response Theory (IRT) to find a difficulty index and discriminating index of items for refining the scale. This will ensure adequate preliminary testing using two theoretical approaches. The modified conceptual framework may be tested again with the refined scale including CT constructs identified by other experts to develop items and examine if the lost two subscales would emerge with further EFA and CFA or those items could be re-tested using the IRT. Both EFA and CFA should be replicated with the new scale using similar samples for further data reduction and model fit indices in order to shorten the scale.

Although the researcher believed that CTS is universal, the development and testing of the scale was initiated with nursing students because it is the researcher's passion to help nursing students and she desired to make a difference in the nursing profession. In future, the researcher has plans to further test the scale using students from other disciplines to examine similarities and differences among them.

The remaining elements of Messick's unitary concept (test-retest (stability) reliability, convergent validity, and discriminant validity), and consequential validity should be tested with the new scale using a similar or comparable sample. The use of scales measuring self concept, anxiety scale, and study habits scale could be choices for discriminant validity and the Need for Cognition or California Critical Thinking Disposition Inventory could serve for convergent validity.

Once the scale is finalized, it should be tested using different ethnic and cultural samples of student populations to examine the influence of ethnicity, culture, and language on CTS. The use of nursing students from schools using differing curricular frameworks could also be tested at this time. This is where one could test students from various nursing education formats – e.g., the traditional four year program, the 2 year liberal arts followed by two years nursing, the accelerated second degree nursing formats or the emerging inter-professional nursing education program. The scale may then be tested on larger student samples in diverse settings to ensure socio-cultural sensitivity of the scale to avoid cultural bias.

Once the scale has been subjected to various tests as described, above, further testing for its utility in a variety of settings can be completed to ensure consequential validity. For example:

- A comparative study may be conducted with teachers and students to examine the similarities and differences in item loadings and pattern of structural validity of the scale.
- A comparison across age groups, gender and different levels of cohort groups to determine how the scale discriminates among the various populations using statistical techniques such as Factorial ANOVA.
- A study may be undertaken for testing the scale for establishing social reliability.

Implications and Conclusions

For students, teachers, and nursing education programs, the current study results imply that, once refined, the CTSAS will be a valid measuring tool for assessing CTS of students, teachers and nurses. When the analyses are complete, there are many implications for use.

Students can effectively utilize this scale for continuous monitoring of their thinking skills for self improvement which can enhance their learning process. For the general student population, this scale is based on the belief that CTS is universal. The scope of expanding the usability of this scale by the students from multi-disciplines in effectively performing the meta-cognitive process for improving their thinking is wide open.

The scale may be valuable for teachers to monitor their own CTS as part of reflective teaching practice. As well, the scale will assist in reflecting on the effect of innovative teaching methodologies on developing and improving CTS in students. The teachers could use the items of the scale to develop rubrics for measuring CTS demonstrated by students in nursing process recordings, journaling, group dynamics, discussions, e-learning, and student portfolio analysis to evaluate the accomplishments of CTS through discipline-specific content. The scale can be useful for monitoring how nursing education programs demonstrate evidence for developing and measuring CTS in students for accreditation purposes.

For practitioners, the CTSAS can assist in self regulating their meta-cognitive process to reflect back on their patient care decisions and improve on their thinking skills to enhance informed decision making in their daily practice. For employers and administrators, once the scale is standardized by establishing norms, the scale could be an invaluable tool for employers using CTS as one of the criteria for assessing the suitability of candidates combined with other professional competencies for job placement. For researchers, mentors and trainers of staff, this CTSAS can be used for future research in strengthening CTS through in-service education programs and mentorship programs and measuring the impact of strong CTS on job satisfaction and retention of practicing nurses.

The conclusion of this study is that the preliminary evaluation of psychometric properties of the CTSAS revealed structural fidelity of three constructs of CTS and the other three constructs demonstrated nearer to close fit indices. Confirmatory Factor Analysis with a larger sample size would have probably yielded clearer results. Contrary to the inconsistent findings reported in the studies using the CCTST (which is also based on the APA definition of CTS) for measuring CTS, the current research, using the same construct definitions for generating items for the CTSAS, revealed inspiring results and showed strong support for the APA construct definition especially for three (Interpretation, Evaluation and Inference) cognitive skills and partial support for the remaining three constructs (Analysis, Explanation and Self regulation) that define CTS.

The results also support Messick's (1995) claim of the unitary concept of construct validity and affirms that content validation is the key to construct validity. Carefully crafting items and rigor in using stringent statistical methods for evaluating the experts` ratings of the items during the content validation process of this study resulted in satisfying outcome on the construct validity of CTSAS. The use of well planned and executed scale construction steps, researcher's personal involvement in administering the tests and providing directions, and the use of powerful statistical strategies for data reduction ensured the scientific soundness of the research methodology employed to control extraneous variables thus yielding intervention fidelity which contributes to the internal validity of the findings.

Further, the researcher ensured statistical conclusion validity and structural fidelity of the scale by the use of more than adequate experts sample size (17) and several criteria for evaluating the experts' ratings (content validation), large sample size (868) for conducting EFA, the alternate methods of factor extractions and rotations, conservative criteria (absolute value 0.45 was set to suppress variables to ensure minimum 20% overlapping variance) used for retaining items in EFA, the conservative values used for region of rejection of the hypothesis in CFA, and several model fit indices (χ^2/df, CFI and RMSEA) to test the goodness of fit of the model specified. All these contributed to the strengths of inferences and the conclusions drawn from this study.

The research conducted was based on the perspectives of Cronbach's vision of Generalizability Theory and followed with caution throughout the research process except the study was unable to satisfy the sampling assumptions for CFA. The researcher recognizes that these results, while promising, are only preliminary. Additional work is required to ensure a valid and reliable scale for use as outlined above. Nevertheless, the results are most promising and encouraging.

References

Adams, M. H., Whitlow, J. F., Stover, L. M., & Johnson, K. W. (1996). Critical thinking as an educational outcome: An evaluation of current tools of measurement. *Nurse Educator, 21*(3), 23-32.

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological measurement, 45,* 131-142.

Austin, Z., Gregory, P. A. M., & Galli, M. (2008). "I just don't know what I'm supposed to know": Evaluating self-assessment skills of international pharmacy graduates in Canada. *Research in Social and Administrative Pharmacy, 4,* 115-124.

Austin, Z., Gregory, P. A. M., & Chiu, S. (2008). Use of reflection-in-action and self- assessment to promote critical thinking among pharmacy students. *American Journal of Pharmaceutical Education, 72(3),* 1-8.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood: Prentice Hall.

Bentler, P. M. (1988). Comparative fit indices in structural models. *Psychological Bulletin, 107,* 238-246.

Bondy,K. N., Koenigseder, L. A., Ishee, J. H., & Williams, B. G. (2001). Psychometric properties of California Critical Thinking Tests. *Journal of Nursing measurement, 9*(3), 309-328.

Brennan, R. L. (2006). *Educational Measurement* (4th ed.). American Council of Education, Praeger publishers.

Brunt, B. A. (2005). Models, measurement, and strategies in developing critical thinking skills. *The Journal of Continuing Education in Nursing, 36*(6), 255-262.

Cacioppo, J.T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, 42(1), 116-131.

Cacioppo, J.T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, *48*, 306-307.

Cacioppo, J. T., Petty, R.E., Feinstein, J. A., & Jarvis, W. B. J (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119* (2), 197-253.

Canadian Association of School of Nursing (2005). CASN Accreditation-recognition of excellence. Retrieved from www.casn.ca

Carter, L. M., & Rukholm, E. (2008). A study of critical thinking, teacher-student interaction, and discipline-specific writing in an online educational setting for registered nurses. *The Journal of Continuing Education in Nursing, 39*(3), 133-138.

Cody, W. K. (2002). Critical thinking and nursing science: Judgment or Vision? *Nursing Science Quarterly, 15*, 184-189.

Colluciello, M. L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students-a conceptual model for evaluation. *Journal of Professional Nursing, 13* (4), 236-245.

Commission on Collegiate Nursing Education. (2003). *Standards for accreditation of baccalaureate and graduate nursing programs.* Retrieved from www.aacn.nche.edu/Accreditation/NEW_STANDARDS.htm

Culhane, S. E., Morera, O. F. & Watson, P. J. (2006). The assessment of factorial invariance in need for cognition using Hispanic and Anglo samples. *The Journal of Psychology, 140*, 53-67.

Culhane, S. E., Moreara, O. F., & Hosch, H. M. (2004). The factor structure of the Need for Cognition –Short Form in a Hispanic sample. *The Journal of Psychology. 138*, 77-88.

Daroszewski, E. B., Kinser, A. G., & Loyd, S. L. (2004) Online, directed journaling in community health advanced practice nursing clinical education. *Journal Nursing Education, 43*, 175-180.

DeVellis, R. F. (2003). *Scale development: Theory and application.* Thousand Oaks, CA: Sage.

Dickieson, P., Carter, L. M., & Walsh, M. (2008). Integrative thinking and learning in undergraduate nursing education: Three strategies. *International Journal of Nursing Education Scholarship. 5*(1), 1-13.

Dunn, J. G. H., Bouffard, M., & Rogers, W. T. (2000). Assessing item content-relevance in sport psychology scale-construction research: Issues and recommendations. *Measurement in Physical Education and Exercise Science, 3*(1), 15-36.

Evans, C. J., Kirby, J. R., & Fabrigar, L. R. (2003). Approaches to learning, need for cognition, and strategic flexibility among university students. *British Journal of Educational Psychology, 73*, 507-528.

Facione, P. A. (1990). Critical thinking: *A statement of expert consensus for purposes of educational assessment and instruction: Executive summary, The Delphi report*. American Philosophical Association, California Academic Press. Retrieved from http://www.isigntassessment.com/article.html

Facione P. A., & Facione, N. (1994, August). *Class Room Assessment of Higher Order Thinking.* Paper presented at the 14th Annual International Conference on Critical Thinking and Educational Reform, Sonoma, CA.

Facione, N. C., & Facione, P. A. (1996). Assessment design issues for evaluating critical thinking in nursing. *Holistic Nursing Practice, 10*(3), 41-53.

Facione , P. A. (2007). Critical Thinking what it is and why it counts? Update 2007, Retrieved from http://www.insightassessment.com/articles.html

Facione, P. A., Facione, N. C., Blohm, S & Gittens, C. A (2008). *California critical thinking skills test manual*. CA: Insight Assessment, The California Academic Press.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). Thousand Oaks: Sage.

Frey, B. B., Petersen, S., Edwards, L. M., Pedrotti, J. T., & Peyton, V. (2005). Item-writing rules: Collective wisdom. *Teaching & Teacher Education, 21*, 357-364.

Galbraith, R. M., Hawkins, R. E., & Holmboe, E. S. (2008). Making self-assessment more effective. *Journal of Continuing Education in the Health Professions, 28(1),* 20-24.

Hayens, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment. 7*(3), 238-247.

Hoelter, J. W. (1983). The analysis of covariance structure: Goodness-of-fit indices. *Sociological Methods & Reserach, 11,* 325-344.

Horan, P. M., DiStefano, C., & Motl, R. W. (2003). Wording effects in Self-Esteem Scales: Methodological artifact or response style? *Structural Equation Modeling, 10*(3), 435-455.

Hu, l., & Bentler, P. M. (1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Ibarreta, G. I., & McLeod, L. (2004). Thinking aloud on paper: An experience in journal writing. *Journal of Nursing Education, 43*(3), 134-137.

Indian Nursing Council. (2004). Syllabus: Basic B.Sc. Nursing. New Delhi: INC.

Ironside, P. M. (2003). New pedagogies for teaching thinking: The lived experiences of students and teachers enacting narrative pedagogy. *Journal of Nursing Education, 42*(11), 509-516.

Ironside, P. M. (2004). Covering content' and teaching thinking: Deconstructing the additive curriculum. *Journal of Nursing Education, 43*(1), 5-12.

Janicek, M. (2006). How to read, understand, and write discussions section in medical articles. An exercise in critical thinking. *Medical Science Monitor, 12*(6), 28-36.

Johns, C. (1995). Framing learning through reflection with Carper's fundamental ways of knowing in nursing. *Advanced Nursing Practice*, *22*(2), 226-234.

Kessler, P., & Lund, C. (2004). Reflective journaling: Developing online journal for distance education. *Nurse Educator, 29*(1), 20-24.

Khosravani, S., Manoochehri, H., & Memarian, R. (2005). Developing critical thinking skills in nursing students by group dynamics. *The Internet Journal of Advanced Nursing Practice, 7* (2), Retrieved from http://www.ispub.com/ostia/index.php?xmlFilePath=Journals/ijanp/vol7n2/skills/xml

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling*. (2nd ed.). New York, NY: The Guilford Press.

Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with California Critical Thinking Tests. *Nurse Educator, 22* (5), 29-33.

Marsh, H. W., Balla, J. R. & McDonald, R. P. (1988). Goodness-of-fit indexes in Confirmatory Factor Analysis: The effect of sample size. *Psychological Bulletin, 103 (3),* 391-410.

May, B. A., Edell, V., Butell, S., Doughty, J., & Longford, C. (1999). Critical thinking and clinical competence: A study of their relationship in BSN seniors. *Journal of Nursing Education, 38,* 100-110.

McCarthey, P., Schuster, P., Zehr, P., & McDougal, D. (1999). Evaluation of critical thinking in a baccalaureate nursing program. *Journal of Nursing Education, 38* (3), 142-144.

MacCallum, R.C., Browne, M.W., & Sugawara, H.M. (1996). Power Analysis and determination of sample size for covariance Structure Modeling. *Psychological Methods, 1*(2), 130-149.

McGrath, P. J. (2003). Issues and innovations in nursing education. *Journal of Advanced Nursing, 43* (6), 569-577.

McGrath, J. P., Hesketh, K. L.,Lang, S., & Estabrooks, C. A. (2003). A study of critical thinking and research utilization among nurses. *Western Journal of Nursing Research, 25* (3), 322-337.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S. (1990). Validity of test interpretation and use. (Report No. ETS-RR-90-11) Princeton, N.J: Educational Testing Service.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50* (9), 741-749.

Morrison, S., & Free, K. W. (2001). Writing multiple-choice test items that promote and measure critical thinking. *Journal of Nursing Education, 40*, 17-24.

National League for Nursing Accreditation Commission, Inc. (2004, revised 2008). *Accreditation manual with interpretive guidelines by program type for post secondary and higher degree programs in nursing.* Retrieved from http://www.nlnac.org/

Oreman, M. H. (1997). Evaluating critical thinking in clinical practice. *Nurse Educator, 22*(5), 25-28.

O'Sulliven, P. S., Belvins-Stephens, W. L., Smith, F. M., & VaghanWrobel, B. (1997). Addressing the National League for Nursing critical thinking outcome. *Nurse Educator*, *22*(1), 23-29.

Paul, R. W., & Elder, L. (2006). *Critical thinking: Learn the tools the best thinkers use* (Con. ed.). Pearson, NJ: Prentice Hall.

Paul, R. W. (2007). *Critical thinking in every domain of knowledge and belief.* Key note address at the 27th international conference on critical thinking, Berkeley, CA. Retrieved from http://www.criticalthinking.org

Paul, R. W. (2009, July). *Exploring the relationship between content of thinking in fostering intellectual discipline.* Paper presented at the 29th international conference on critical thinking, Berkeley, CA. Retrieved from http://www.criticalthinking.org

Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (8th ed.). Philadelphia: Lippincott Williams & Wilkins.

Rubenfeld, M. G., & Scheffer, B. K. (1999). *Critical thinking in nursing: An integrative approach.* (2nd ed.). Philadelphia: Lippincott.

Sanders, J., Gass, R., & Wieserman, R., & Bruschke, J. (1992). Ethnic comparison and measurement of argumentativeness, verbal aggressiveness and need for cognition. *Communication Reports*, *5*(1), 50-56.

Saskatchewan Registered Nurses Association. (2007). Standards and Foundation Competencies for the Practice of Registered Nurses. Retrieved from http://www.srna.org/images/stories/pdfs/about_us/ends_standards_codes.pdf

Saucier, B. L., Stevens K. R., & Williams, G. B. (2000). Critical thinking outcome of computer assisted instruction versus nursing process. *Nursing and Health Care Perspectives, 21*, 240-246.

Shin, K., Jung, D., Shin, S., & Kim, M. (2006). Critical thinking disposition and skills of senior nursing students in associate, baccalaureate, and RN to BSN programs. *Journal of Nursing Education, 45*(6), 233-237.

Simpson, E., & Courtney, M. (2002). Critical thinking in nursing education: Literature review. *International Journal of Nursing Practice, 8*(2), 89-98.

Spelic, S. S., Parsons, M., Hercinger, M., Andrews, A., Parks, J., & Norris, J. (2001). Evaluation of critical thinking outcomes of a BSN program. *Holistic Nursing Practice, 15*(3), 27-34.

Staib, S. (2003). Teaching and measuring critical thinking. *Journal of Nursing Education, 42*(11), 498-508.

Stein, B., Haynes, A., & Understein, J. (December 6-9, 2003). “Assessing Critical Thinking.” Paper presented at SACS Annual meeting in Nashville, Tennessee. Retrieved from http://iweb.tntech.edu.sti/SACS%20presentations%20paper.pdf

Stone, C. A., Davidson, L. J., Evans J. L., & Hansen, M. A. (2001). Validity evidence for using a general critical thinking test to measure nursing students’ critical thinking. *Holistic Nursing Practice, 15*(4), 65-74.

Sumner, W. G. (1940). *Folkways. A study of the sociological importance of usages, manners, customs, mores and morals.* New York, NY: Ginn and Company.

Tabachinick, B.G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education, Inc.

Tiwari, A., Lai, P., So, M., & Yuen K. (2005). A comparison of effect of problem based learning and lecturing on the development of students' critical thinking. *Medical Education, 40* (6), 547-554.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.

Twibell, R., Ryan, M., & Hermiz, M. (2005). Faculty perception of critical thinking in student clinical experience. *Journal of Nursing Education, 44*(2), 71-79.

Wagner, T. A. & Harvey, R. J. (2006). Development of a new critical thinking test using item response theory. *Psychological Assessment, 18*(1), 100-105.

APPENDICES

Appendix A

Schedule for Data Collection

Contact addresses of Colleges of Nursing, Kerala, India.

Name and address of the Colleges	Scheduled dates
1. St. Joseph's College of Nursing, Dharmagiri, kothamangalam, Ernakulam Dt, Kerala, India. Postal Code: 686691.	Test July 10, 2010 (Saturday)
2. College of Nursing KVM Trust P B no. 13 Cherthala, Alappuzha Dt. Kerala. India 688524	July 22 & 23, 2010 (Thursday & Friday) Re-test August 11, 2010 (Wednesday)
3. Theophilus College of Nursing & Director of Nursing Education, P. Geevarghese School of Nursing, Devagiri P.O, Kangazha, Kottayam Dt. Kerala, India, 686 555	Test July 08, & 09, 2010 (Thursday& Friday) Retest July 26 2010 (Monday)
4. Aswini College of Nursing Thrissur, Kerala, India, 680751	Test July 28 & 29, 2010 (Wednesday & Thursday)
5. Pushpagiri College of Nursing, Tiruvalla, Pathanamtitta Dt. Kerala, India. 689101,	Test July13 & 14, 2010 (Tuesday & Wednesday) Retest July 30, 2010 (Friday)

Data collection Schedule for Term I students at Saskatoon, Regina & Prince Albert

University of Saskatchewan Sites	Dates	Time
Saskatoon		
Year IV	Nov 01/10	12:30-2:30
Year III	October 14/10	12:30-1:30
Prince Albert		
Year IV	October 18/10	12:00 – 1:00
Year III	October 19/10	12:00-1:00
Regina		
Year IV	Nov 08/10	12:30-1:30
Year III	Nov 08/10	3:30-4:30

Appendix B

UNIVERSITY OF SASKATCHEWAN

Behavioural Research Ethics Board (Beh-REB)

Certificate of Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	BEH#
Lynnette L. Stamler	Nursing	10-84

INSTITUTION(S) WHERE RESEARCH WILL BE CONDUCTED
University of Saskatchewan

STUDENT RESEARCHERS
Girija Nair

SPONSOR
UNFUNDED

TITLE
Developing a Critical Thinking Self-Assessment Scale and Evaluating its Preliminary Psychometrics for Undergraduate Nursing Students

ORIGINAL REVIEW DATE	APPROVAL ON	APPROVAL OF	EXPIRY DATE
10-Apr-2010	29-Apr-2010	Ethics Application Consent Protocol	29-Apr-2011

Full Board Meeting ☐ Date of Full Board Meeting:

Delegated Review ☒

CERTIFICATION

The University of Saskatchewan Behavioural Research Ethics Board has reviewed the above-named research project. The proposal was found to be acceptable on ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research project, and for ensuring that the authorized research is carried out according to the conditions outlined in the original protocol submitted for ethics review. This Certificate of Approval is valid for the above time period provided there is no change in experimental protocol or consent process or documents.

Any significant changes to your proposed method, or your consent and recruitment procedures should be reported to the Chair for Research Ethics Board consideration in advance of its implementation.

ONGOING REVIEW REQUIREMENTS

In order to receive annual renewal, a status report must be submitted to the REB Chair for Board consideration within one month of the current expiry date each year the study remains open, and upon study completion. Please refer to the following website for further instructions: [illegible]

John Rigby, Chair
University of Saskatchewan
Behavioural Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University, 1602 – 110 Gymnasium Place
Saskatoon SK S7N 4J8

Appendix B 1

UNIVERSITY OF SASKATCHEWAN

College of Nursing

107 Wiggins Road
Saskatoon SK S7N 5E5 Canada
Telephone: (306) 966-6244
Facsimile: (306) 966-6609

May 21, 2010

Dear Girija Nair

RE: Developing a Critical Thinking Self-Assessment Scale and Evaluating its Preliminary Psychometrics for Undergraduate Nursing Students.

Thank you for sending the above research proposal and ethics approval for consideration by our College Research Committee. I am pleased to inform you that the Committee has approved your study. This approval allows you to contact students for inclusion in your study but does allow for data collection to take place during class time. The Committee also recognized the high burden placed on the students through the time commitment for data collection. It is also suggested that email may be a good route for contacting students when successful in the draw. Please note that this approval does not extend to health science students from any other College, Unit, or Department.

We wish you well in your project and would kindly ask that you share a copy of your report with us as we are very interested in hearing the results.

Please feel free to contact me via telephone or email phil.woods@usask.ca if you require further information.

Kind regards

Dr Phil Woods
Associate Dean Research, Innovation and Global Initiatives

Appendix C

Box.1. Definitions of Core Cognitive Skills of Critical Thinking

Interpretation. To comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria.

Analysis. To identify the intended and actual inferential relationships among questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions.

Evaluation. To assess the credibility of the statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intend inferential relationships among statements, descriptions, or other forms of representation.

Inference. To identify and secure elements needed to draw reasonable conclusions; to form conjunctures and hypotheses; to consider relevant information; and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts descriptions and questions or other forms of representation

Explanation. To state the results of one's reasoning; to justify reasoning in terms of evidential, conceptual, methodological, criteriological and contextual considerations upon which one's results were based; and to present reasoning in the form of cogent arguments.

Self-Regulation. Self-consciously to monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgment with a view toward questioning, confirming, validating or correcting either one's reasoning or one's results.

Source: Facione, P. (1990, p. 6-11) American Philosophical Association Delphi Report.

Appendix D

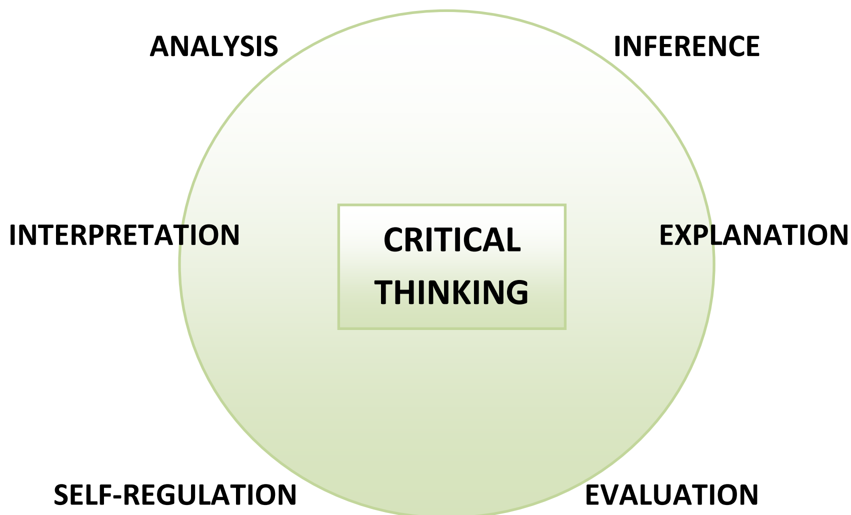
Box 2. Core Cognitive Skills and Sub-Skills

Core skills	Sub skills
Interpretation:	categorization, decoding significance and clarifying meaning,.
Analysis:	examining ideas, identifying arguments, and analyzing arguments.
Evaluation:	assessing claims and assessing arguments.
Inference:	querying evidence, conjecturing alternatives, and drawing conclusions.
Explanation:	stating results, justifying procedures, and presenting the arguments.
Self-Regulation:	self-examination and self-correction.

Source: Facione, APA, (1990).

Appendix E

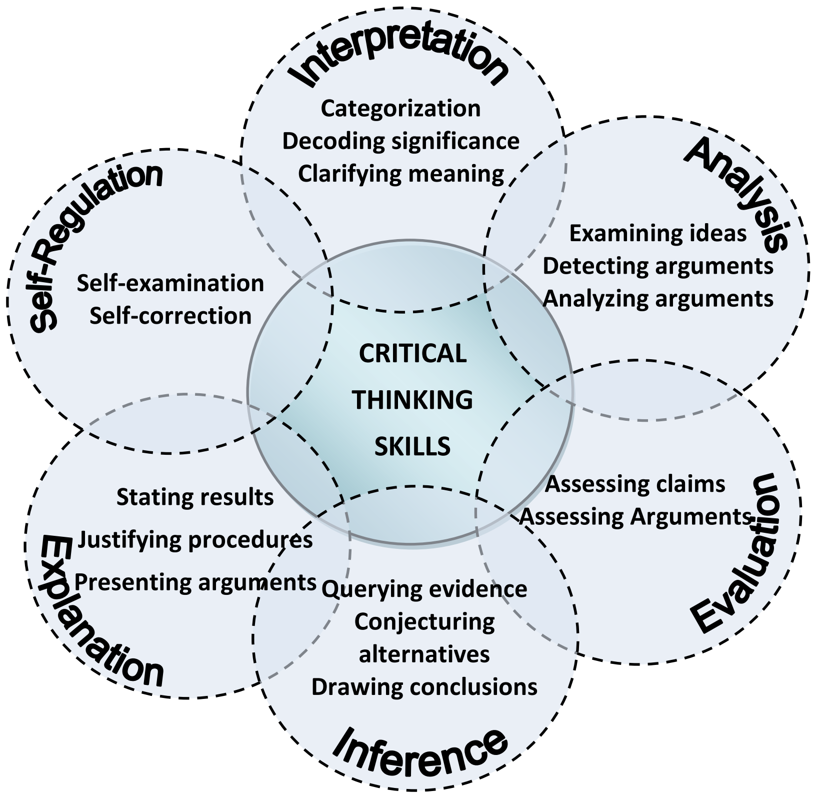
Figure 2.1. Core Critical Thinking Skills



(Source: Facione, APA, 1990)

Appendix F

Figure 2.2. Researcher Modified Conceptual Framework of Core Cognitive and Sub-skills of CTS



(Source: Concepts adapted from Facione, APA, 1990)

Appendix G

Definitions of Core Cognitive Skills and Sub-skills of Critical Thinking Skills

1. Interpretation. "To comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria" (Facione, 1990, p. 6).

Sub skills of Interpretation

1.1. Categorization. "To apprehend or appropriately formulate categories, distinctions or frameworks for understanding, describing or characterizing information. To describe experiences, situations, beliefs, and events" (Facione, 1990, p. 6).

1.2. Decoding significance. "To detect, attend to, and describe the informational content, affective purport, directive functions, intentions, motives, purposes, social significance, values, views, rules, procedures, criteria or inferential relationships expressed in convention-based communication systems such as in language, social behaviors, drawings, numbers, graphs, tables, charts, signs and symbols" (Facione, 1990, p. 7).

1.3. Clarifying meaning. "To paraphrase or make explicit, through stipulation, description, analogy or figurative expression, the contextual, conventional or intended meanings of words, ideas, concepts, statements, graphs, numbers, symbols, rules, events etc" (Facione, 1990, p. 7).

2. Analysis. "To identify the intended and actual inferential relationships among questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information or opinions" (Facione, 1990, p. 7)

Sub-skills of Analysis

2.1. Examining ideas. “To determine the role various expressions play or are intended to play in the arguments, reasoning or persuasion; to define terms; to compare or contrast ideas, concepts, or statements; to identify issues or problems and determine their component parts, and also to identify the conceptual relationships of the parts” (Facione, 1990, p. 7) .

2.2. Detecting arguments. “Given a set of statements, descriptions, questions or graphic presentations to determine whether or not the set expresses or is intended to express, a reason or reasons in support of or contesting some claim, opinion or point of view” (Facione, 1990, p. 7).

2.3. Analyzing arguments. “Given the expression of a reason or reasons intended to support or contest some claim, opinion or point of view, to identify and differentiate (a) the intended main conclusion, (b) the premises and reasons advanced in support of the main conclusion, (c) further premises and reasons advanced as backup or support for those premises and reasons intended as supporting the main conclusion (d) additional unexpressed elements of that reasoning such as intermediary conclusions, unstated assumptions, or presuppositions, (e) the overall structure of the argument or intended chain of reasoning, and (f) any items contained in the body of expressions being examined which are not intended to be taken as part of the reasoning being expressed or its intended background ” (Facione, 1990, p. 8).

3. Evaluation. “To assess the credibility of the statements or other representations which are accounts or descriptions of a person’s perception, experience, situation, judgment, belief or opinion; and to assess the logical strength of the actual or intend inferential relationships among statements, descriptions, or other forms of representation” (Facione, 1990, p. 8).

Sub skills of Evaluation

3.1. Assessing claims. “To recognize the factors relevant to assessing the degree of credibility to ascribe to a source of information or opinion; to assess the contextual relevance of questions, information, principles, rules or procedural directions; and to assess the acceptability, the level of confidence to place in the probability of evidence of any given representation of an experience, situation, judgment, belief or opinions”(Facione, 1990, p. 8).

3.2. Assessing arguments**.** “To judge whether the assumed acceptability of the premises of a given argument justify one’s accepting as true (deductively certain) or very probably true (inductively justified) the expressed conclusion of that argument; to anticipate or raise questions or objections, and to assess whether these point to significant weakness in the argument being evaluated; to determine whether an argument relies on false or doubtful assumptions or presuppositions and determine how crucially these affect its strength; to judge between reasonable and fallacious inferences; to judge the probative strength of an argument’s premises and assumptions with a view toward determining the acceptability of the argument; to determine and judge the probative strength of an argument’s intended or unintended consequences with a view toward judging the acceptability of the argument; and to determine to the extent to which possible additional information that might strengthen or weaken the argument” (Facione, 1990, p. 8).

4. Inference. “To identify and secure elements needed to draw reasonable conclusions; to form conjunctures and hypotheses; to consider relevant information; and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs,

opinions, concepts descriptions and questions or other forms of representation" (Facione, 1990, p. 9).

Sub skills of Inference

4.1. Querying evidence. "To recognize premises which require support and to formulate a strategy for seeking and gathering information that might supply that support; to judge the information relevant to deciding acceptability, plausibility or relative merits of a given alternative, question, issue, theory, hypotheses or statement is required, and to determine plausible investigatory strategies for acquiring that information" (Facione, 1990, p. 9).

4.2. Conjecturing alternatives. "To formulate multiple alternatives to resolving problem; to postulate a series of suppositions regarding a question; to project alternative hypotheses regarding an event; to develop a variety of plans to achieve some goal; to draw out presuppositions and project the range of possible consequences of decisions, positions, policies, theories or beliefs" (Facione, 1990, p. 9).

4.3. Drawing conclusions. "To apply appropriate modes of inference in determining what position, opinion or point of view one should take in a given matter or issue; given a set of statements, descriptions, questions or other forms of representation, to educe with proper level of logical strength, their inferential relationships and the consequences or the presuppositions which they support, warrant, imply or entail; to employ successfully various sub-species of reasoning, as for example, analogically, arithmetically, dialectically, scientifically etc.; to determine which of several possible conclusions is most strongly warranted or supported by the evidence at hand, or which should be rejected or regarded as less plausible by the information given" (Facione, 1990, p. 9).

5. Explanation. "To state the results of one's reasoning; to justify reasoning in terms of evidential, conceptual, methodological, criteriological and contextual considerations upon which one's results were based; and to present reasoning in the form of cogent arguments" (Facione, 1990, p. 10).

   Sub skills of Explanation

   5.1. Stating results. "To produce accurate statements, descriptions, or representations of the results of one's reasoning activities so as to analyze, evaluate, infer from, or monitor those results" (Facione, 1990, p. 10).

   5.2. Justifying procedures. "To present the evidential, conceptual, methodological, criteriological, and contextual considerations which one used in forming one's interpretations, analysis, evaluation or inferences, so that one might accurately record, evaluate, describe, or justify those processes to oneself and to others" (Facione, 1990, p. 10).

   5.3. Presenting arguments. "To give reason for accepting the claim; to meet objections to the method, conceptualizations, evidence, criteria or contextual appropriateness for inferential, analytical or evaluative judgments" (Facione, 1990, p. 10).

6. Self-Regulation. "Self-consciously to monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis and evaluation to one's own inferential judgment with a view toward questioning, confirming, validating or correcting either one's reasoning or one's results" (Facione, 1990, p. 10).

Sub skills of Self-regulation

6.1. Self- examination. “To reflect on one’s own reasoning and verify both the results produced and the correct application and execution of the cognitive skills involved; to make objective and thoughtful meta-cognitive self-assessment of one’s opinions and reasons for holding them; to judge to the extent which one’s thinking is influenced by deficiencies in one’s knowledge, or by stereotypes, prejudices, emotions or any other factors which constrain the objectivity or rationality; to reflect on one’s motivations, values, attitudes and interests with a view toward determining that one has endeavored to be unbiased, fair-minded, thorough, objective, respectful of the truth, reasonable and rationale in coming to one’s analysis, interpretations, evaluations, inferences or expressions” (Facione, 1990, p. 10-11).

6.2. Self- correction. “Where self-examination reveals errors or deficiencies, to design reasonable procedures to remedy or correct those mistakes and their causes” (Facione, 1990, p. 11).

(Permission to duplicate for non-commercial uses may be purchased for $0.80 per copy from California Academic Press, (c) 1990. The material is purchased from the press).

Appendix H

List of Expert Judges Panel

Name of the Discipline	Address
1. Theology	Lutheran Theological Seminary Saskatoon, SK
2. Theology	St Andrews College Saskatoon SK
3. Nursing	California State University, Long Beach Department of Nursing 1250 Bellflower Blvd. Long Beach, California 90840
4. Nursing	College of Nursing University of Saskatchewan
5. Nursing	College of Nursing University of Saskatchewan
6. Nursing	College of Nursing University of Saskatchewan
7. Nursing	Government College of Nursing Kolkata, INDIA, 700089
8. Nursing	College of Nursing Mangalore, India.
9. Geology & Teaching Effectiveness	University Learning Centre The Gwenna Moss Centre for Teaching and Department of Geography and Planning University of Saskatchewan
10. Computer Science	Computer Science University of Saskatchewan
11. Health Sciences (Inter-professional Education)	Educational Support & Development, College of Medicine University of Saskatchewan

12. Health Sciences Inter-professional Education	College of Medicine University of Saskatchewan
13. Educational Psychology	Department of Educational Psychology & Special Education College of Education University of Saskatchewan
14. Educational Philosophy	College of Education University of Saskatchewan
15. Science (Physics)	Dept of Physics and Engineering Physics University of Saskatchewan
16. Social Sciences	St . Thomas Moore College University of Saskatchewan
17. Kinesiology	College of Kinesiology University of Saskatchewan

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Interpretation - Subscale : Categorization

Category: Interpretation							
Item #	Subscale: 1.1. Categorization	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks Modifications /rejections*
1.1.1	When presented with a problem, first I try to figure out the various aspects of the problem.	3	5	0.75	0.766	.009	modified to - content
1.1.2	I try to define the character of the problem to recognize it.	3	5	0.688	0.672		modified to -nature (rejected)
1.1.3	I sort the information into sub sets.	4	3	0.938	0.875	<.01	retained
1.1.4	I clearly describe my experience in a given situation.	3	5	0.625	0.688		rejected
1.1.5	I classify data or findings using a diagram.	4	3	0.81	0.844	<.01	Changed to - framework
1.1.6	I use concept mapping to figure out the problem.	4	4	0.81	0.84	<.01	I breakdown problem into discrete parts (language)
1.1.7	I break the complex assignments into manageable sub-tasks to define concepts.	4	4	0.81	0.828	<.01	I break the complex ideas into manageable sub-ideas (retain)
1.1.8	I figure out the various components of a concept or opinion.	3	4	0.75	0.781	<.01	retained
1.1.9	I categorize the information for understanding the problem.	4	5	0.81	0.813	<.01	I categorize similar and related information into groups (clarity)
1.1.10	I detect person's views and intentions on the issue raised. (M)	2	5	0.31	0.437		Rejected
1.1.11	I ensure all the steps of a problem are considered.	2	5	0.56	0.578		Rejected

Criteria for retaining or rejecting, and modifying an item. (Accumulative evidence for Validity decision)

Right-tail probability value of validity coefficient VI_k (Aiken, 1985, p.134) is found from the table for 17 experts, a value of 0.66 is significant at <0.036. I-CVI Validity index according to Polit & Beck for these items are almost closer to VI_k. Which is a better scientific value than *I-CVI*, as it takes into account the variations in the score as the value is calculated from the summated scores of raters.

Wherever ambiguity was a concern the items are modified even if the items have the significant ***VI_k* &*I-CVI*** values**.**

SEVEN items retained and language modified according to expert's suggestion. The VI_k for these items are highly significant the range is very high indicating highly ambiguous items due to mainly technical language. SEVEN items are retained and some with modifications.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Interpretation – subscale: Decoding significance.

Category: Interpretation

Item #	Subscale: Decoding significance.	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
1.2.1	I figure out a person's purpose in asking a question.	4	3	0.875	0.88	<.01	Retained
1.2.2	I observe the facial expression or nonverbal cues people use in a given situation.	4	2	1	0.92	<.01	I observe the facial expression people use in a given situation
1.2.3	I appreciate the meaning of an individual's gesture in a given situation.	4	5	0.875	0.813	<.01	I clarify the meaning of an individual's gesture in a given situation
1.2.4	I detect the use of sarcastic questions in a debate.	4	5	0.687	0.66		Rejected
1.2.5	I sort the information into sub sets.	0	5	0.375	0.38		Rejected (M)
1.2.6	I classify the whole information into specifics to interpret the data.	3	5	0.562	0.594		Rejected
1.2.7	I cluster the data and reduce it to derive its meaning.	3	5	0.562	0.625		Rejected
1.2.8	I pay attention to the social relevance of the information in the text.	4	4	0.937	0.844	<.01	I try to identify the social relevance of the information presented in the text.
1.2.9	I identify the author's views and intentions in the issue presented.	4	2	1	0.859	<.01	Retained
1.2.10	I figure out the relationship among the concepts to understand their meaning.	4	4	0.875	0.84	<.01	I look for various relationships among concepts to understand the meaning.
1.2.11	I figure out the meaning of a given issue / situation based on my assumptions.	4	5	0.687	0.7		I examine the values rooted in the information presented.
1.2.12	I use a picture or graph to explain the relationship of a complex question.	3	5	0.625	0.641		Rejected

SEVEN items are retained and some of them modified to address ambiguity.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Interpretation – subscale: Clarifying Meaning

Category: Interpretation

Item #	Subscale: Clarifying Meaning	Mdn	Range	**I-CVI**	***VI_k***	*P-value*	*Remarks*
1.3.1	I restate or paraphrase another person's statements to clarify the meaning.	4	2	1	0.95	<.01	Retained
1.3.2	I figure out an example which helps explain the concept /opinion.	4	2	1	0.91	<.01	Retained
1.3.3	I clarify my thoughts by explaining to someone else.	4	2	0.937	0.84	<.01	Retained
1.3.4	I figure out the distinction between opinions /concepts/ideas to remove ambiguity.	4	4	0.81	0.81	<.01	I try to differentiate between opinions and ideas to remove ambiguity
1.3.5	I engage in brainstorming to figure out the meaning of the given problem.	4	5	0/5	0.64		Rejetced
1.3.6	I figure out the context to derive the meaning of another's point of view.	3	5	0.81	0.7	<05	Rejetced
1.3.7	I look for analogies of the words and concepts to clarify meaning.	4	3	0.937	0.88	<.01	Retained
1.3.8	I am comfortable when I figure out the meaning of expressions in a graph.	3	5	0.625	0.58		Rejetced
1.3.9	I figure out the meaning of another's point of view.	4	4	0.937	0.92	<.01	Retained
1.3.10	I seek clarification of the meanings of opinion /belief/ the points of view raised by others.	4	5	0.81	0.78	<.01	Retained
1.3.11	Thinking drives my all actions. (M)	1		0.312	0.359		Rejected
1.3.12	Given a statement or question, I find out its purpose.	3	5	0.625	0.672		Rejected
1.3.13	I recognize my own lack of understanding.	3	5	0.625	0.641		Rejected

SEVEN items are retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Analysis – subscale: Examining Ideas

Category: Analysis							
Item #	Subscale: Examining Ideas	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
2.1.1	I identify the phrases used by people to persuade others to agree with their opinion.	3	5	0.687	0.69		Rejected
2.1.2	I examine the similarities and differences among the opinions posed for a given problem.	4	5	0.75	0.77	<.01	Retained
2.1.3	I explain the concepts supported by reasons or examples. (M)	4	5	0.625	0.67		Rejected
2.1.4	I examine the arguments for a valid reason.	4	5	0.625	0.67		Rejected
2.1.5	I examine the interrelationships among concepts or opinions posed.	4	5	0.81	0.8	<.01	Retained
2.1.6	I provide my points of view/opinions during discussions on issues.	3	5	0.5	0.55		Rejected
2.1.7	I analyze the breadth and depth of problem based on the information.	4	5	0.687	0.72	<.05	Rejected
2.1.8	I can't stay focused on the issues/opinions/ ideas while analysing them. (-ve)	3	5	0.312	0.36		Rejetced
2.1.9	I seek evidence when another's view point contradicts my belief.	4	5	0.625	0.719	<.05	I look for a supporting reason when examining opinions.
2.1.10	I look for relevant information to answer the question at issue.	4	5	0.625	0.703	<.05	Retained
2.1.11	I examine the proposals for solving a given problem.	4	5	0.687	0.75	<.05	Retained
2.1.12	Given a situation, I look for specifics to analyze its various aspects.	4	5	0.875	0.813	<.01	Retained

SIX items are retained and one modified

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Analysis – subscale: Detecting Arguments

Category: Analysis

Item #	Subscale: Detecting Arguments	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
2.2.1	Given a reading paragraph, I determine the main claim.	4	5	0.81	0.8	<.01	Given a paragraph to read:
2.2.2	I figure out the reasons to support the author's claim.	4	5	0.81	0.83	<.01	I figure out what my reasons are for supporting or not supporting the author's claim.
2.2.3	Given a reading paragraph, I state questions to find evidence for reasons.	3	5	0.75	0.77	<.01	I ask questions in order to seek evidence to support or refute the author's claim.
2.2.4	Given a reading passage, I figure out if author's arguments include both for and against the claim.	4	4	0.875	0.86	<.01	Retained
2.2.5	I seek supporting reasons when a person is advancing a claim.	4	3	0.875	0.91	<.01	Retained
2.2.6	I agree with the opinion when presented with valid evidence.	3	5	0.625	0.64		Rejected
2.2.8	I find it hard to understand another's arguments on issues.	1	5	0.312	0.39		Rejected
2.2.9	My argument against a belief is based on the context in which it is presented.	4	5	0.625	0.63		Rejected
2.2.10	Given a question/statement / point of view I find out its purpose. (M)	3	5	0.625	0.59		Rejected
2.2.11	I neither agree nor disagree with others' views until I am clear.	3	5	0.563	0.578		Rejected
2.2.12	I am fair-minded to others' arguments even if I disagree with them.	4	5	0.687	0.719	<.05	Rejected

IX items retained and some with modification

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Analysis – subscale: Analyzing Arguments

Category: Analysis

Item #	Subscale: Analyzing Arguments	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
2.3.1	Given a brief argument, I identify the author's chief claim.	4	5	0.81	0.78	<.01	Retained
2.3.2	It is laborious to analyze the premises or assertions others make in support of the main conclusion. (-ve)	0	5	0.312	0.36		Rejected
2.3.3	I look for background reasons to support the premises of the claim.	3	5	0.625	0.64		Rejected
2.3.4	I figure out unstated assumptions in one's reasoning for a claim.	4	5	0.81	0.76	<.01	Retained
2.3.5	I look for the overall structure of the argument.	4	5	0.875	0.84	<.01	Retained
2.3.6	I avoid getting into arguments and would rather agree with another's views or opinions. (negative -ve)	1	5	0.25	0.33		Rejected
2.3.7	I avoid assessing another's arguments because it results in conflicts. (-ve)	1	5	0.25	0.31		Rejected
2.3.8	I look for the intended chain of reasoning in an opinion or point of view.	3	5	0.81	0.77	<.01	I figure out the process of reasoning for an argument.
2.3.9	I seek evidence for the reasons supporting the arguments.	4	5	0.875	0.86	<.01	Retained
2.3.10	I analyse the breadth and depth of problem based on the information gathered. (M)	3	5	0.625	0.656		Rejected
2.3.11	I figure out the assumptions implicit in the author's reasoning.	3	5	0.81	0.77	<.01	Retained
2.3.12	I map out the inferential flow of reasoning for an argument.	3	5	0.687	0.641		Rejected

SIX items retained one modified

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Evaluation - subscale : Assessing Claims

			Category: Evaluation				
Item #	Subscale: Assessing Claims	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
3.1.1	I figure out the relevant factors to assess credibility of the opinion / belief raised.	4	4	0.875	0.86	<.01	Retained
3.1.2	I assess the credible authority of information sources about a topic based on the presenter.	3.5	5	0.687	0.7	<.05	I assess the credible authority of the source of information supporting the claim. (retained)
3.1.3	I figure out if a given claim is true or false based on the supported knowledge.	4	3	0.937	0.89	<.01	Retained
3.1.4	I am more comfortable receiving an instruction than evaluating others' statements to make a decision. (-ve)	2	5	0.437	0.48		Rejected
3.1.5	I find it difficult to evaluate another's point of view. (-ve)	2.5	5	0.5	0.53		Rejected
3.1.6	I seek help from others when the situation demands a decision. (-ve)	3	5	0.5	0.53		Rejected
3.1.7	I propose my beliefs and assumptions after assessing their usefulness to the issues posed.	3	5	0.687	0.66		Rejected
3.1.8	I demand the credible authority of the source of information supporting the claim.	2.5	5	0.625	0.72	<.05	Retained
3.1.9	I recognize that learning problems result from a natural desire to avoid frustration. (M)	1	5	0.25	0.38		Rejected
3.1.10	I am aware of my egocentric view when assessing a claim.	3	5	0.625	0.59		Rejected
3.1.11	I assess the contextual relevance of an opinion or claim posed.	3	3	0.875	0.84	<.01	
3.1.12	I seek the accuracy of the evidence supporting a given judgment or opinion.	3.5	3	0.937	0.88	<.01	Retained

FIVE items are retained with one modified.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Evaluation - subscale : Assessing Arguments

Category: Evaluation

Item #	Subscale: Assessing arguments	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
3.2.1	I assess the probability of a premise used in concluding an argument.	3	4	0.81	0.797	<.01	I assess the chances of success or failure in using a premise to conclude an argument.
3.2.2	I rarely examine the flaws in an argument. (-ve)	3	5	0.75	0.703	<.05	Retained
3.2.3	I examine the logical strength of the causal reasoning in an argument.	4	5	0.937	0.86	<.01	Retained (underlying reasoning..
3.2.4	I support an argument without examining the relevance of a given situation. (-ve)	3	5	0.625	0.58		Rejected
3.2.5	I figure out the implications of the arguments supporting the conclusion.	3	5	0.687	0.64		Rejected
3.2.6	I search for new data to confirm or refute a given opinion/ claim.	4	5	0.75	0.75	<.01	Retained (Remove opinion)
3.2.7	I examine the applicability of the principle proposed for deciding what to do in a given situation.	4	5	0.687	0.69		Rejected
3.2.8	I search for additional information that might support or weaken the argument.	4	4	0.875	0.84	<.01	Retained
3.2.9	I raise questions when the assumptions supporting the arguments are doubtful or false.	4	5	0.81	0.77	<.01	I ask questions when the assumptions supporting the arguments are false.
3.2.10	I map out the inferential flow of reasoning for an argument. (M)	3	5	0.5	0.53		Rejected
3.2.11	I examine the logical reasoning of an objection to an opinion.	3	5	0.875	0.77	<.01	Retained
3.2.12	I do not accept a conclusion which is not based on logic.	3	5	0.75	0.7	<.05	I look for conclusions which are logical.

EIGHT items are retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Inference - subscale : Querying Evidence

Category: Inference

Item #	Subscale: Querying evidence	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
4.1.1	When developing a persuasive argument, I search for useful information to support my point of view.	4	3	0.875	0.891	<.01	Retained
4.1.2	I seek relevant information to support another's point of view.	4	3	0.875	0.859	<.01	Retained
4.1.3	I seek useful information to refute an argument when supported by doubtful reasons.	4	3	0.875	0.88	<.01	Retained (change unsure)
4.1.4	I collect evidence supporting the availability of information to back up opinions.	4	5	0.81	0.77	<.01	Retained
4.1.5	I think accepting some missing information is useful for a more or less reasonable opinion than a competing opinion. (-ve)	3	5	0.625	0.61		Rejected
4.1.6	I plan an information search to reveal if any information is available.	3	5	0.562	0.547		Rejected
4.1.7	I seek for evidence / information before accepting a solution.	3	5	0.75	0.73	<.01	Retained
4.1.8	It is too much work to do a literature search to support or refute a view point. (-ve)	2	5	0.375	0.44		Rejected
4.1.9	I consider another's view points and seek evidence during discussions.	4	5	0.937	0.86	<.01	I consider opposing views in support of information when controversial issues are examined.
4.1.10	Given a reading paragraph, I determine the main claim. (M)	2	5	0.31	0.39		Rejected
4.1.11	I cannot accept a conclusion without understanding the supporting evidence.	3	4	0.81	0.77	<.01	Retained

SEVEN items are retained with one modified.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Inference – subscale: Conjecturing Alternatives

Category: Inference

Item #	Subscale: Conjecturing Alternatives	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
4.2.1	I cluster the related composite data to infer the problem in a given situation.	2	5	0.5	0.61		Rejected
4.2.2	I develop a couple of assumptions to address the problem or issue.	3	5	0.562	0.63		Rejected
4.2.3	I figure out alternate hypotheses / questions, when I need to solve a problem.	4	5	0.937	0.89	<.01	Retained
4.2.4	Given a problem to solve, I develop a set of options for solving the problem.	4	3	0.937	0.91	<.01	Retained
4.2.5	Whether or not one agrees, I state the difficulties and the benefits of adopting a given a set of priorities for decision making.	4	4	0.875	0.84	<.01	Retained
4.2.6	I systematically analyse the problem using multiple sources of information to draw inferences.	3	5	0.625	0.67		Retained with modification (multiple sources)
4.2.7	I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.	4	5	0.75	0.73	<.01	Retained
4.2.8	I make links between concepts to derive hypotheses.	4	5	0.562	0.64		Rejected
4.2.9	I seek help from friends to get an easy answer to the problem.	3	5	0.5	0.5		Rejected
4.2.10	I sort the information into subsets. (M)	2	5	0.312	0.41		Rejected
4.2.11	I consider others' points of view when formulating all possible solutions.	4	4	0.875	0.84	<.01	I talk to others to get feedback on various ways of constructing alternate hypotheses.
4.2.12	I identify the consequences of various options to solving a problem.	4	3	0.937	0.875	<.01	Retained

SEVEN items retained.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Inference - subscale : Drawing Conclusions

Category: Inference

Item #	Subscale: Drawing Conclusions	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
4.3.1	I depend on statistical techniques for drawing inferences.	3	5	0.625	0.69	<.05	Retained
4.3.2	I consider opposing views in support of information when controversial issues/opinions are examined.	4	5	0.687	0.72	<.05	Rejected (overlap)
4.3.3	I gather more information when conclusions are made from opposing views.	3	5	0.75	0.73	<.01	Retained (multiple source of)
4.3.4	I use both deductive and inductive rules to interpret information.	4	5	0.75	0.73	<.01	Retained (deductive and inductive reasoning)
4.3.5	I arrive at conclusions that are supported with strong evidence.	4	3	0.937	0.92	<.01	Retained
4.3.6	I actively analyse my thinking before jumping to conclusions.	4	5	0.875	0.84	<.01	Retained (remove actively)
4.3.7	I figure out the logical relationship of the reasons supporting the conclusions.	4	5	0.75	0.72	<.05	Retained
4.3.8	It is difficult for me to draw conclusions from the data I gather. (-ve)	2	5	0.5	0.6		Rejected
4.3.9	I foresee the consequences of my inferences and actions.	3	5	0.625	0.66		Rejected
4.3.10	I confidently reject an alternative solution when it lacks evidential reasoning.	3	5	0.81	0.75	<.01	Retained
4.3.11	I reject an alternative solution when its consequences are unclear.	3	5	0.687	0.67		Rejected
4.3.12	I make links between concepts to derive hypotheses. (M)	3	5	0.5	0.563		Rejected
4.3.13	I categorize the pros and cones of a solution before accepting it.	3	5	0.687	0.703	<.05	(Retain) modify -I figure out the

EIGHT items retained.

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Explanation - subscale : Stating Results

Category: Explanation

Item #	Subscale: Stating Results	Mdn	Range	***I-CVI***	VI_k	*P-value*	*Remarks*
5.1.1.	Given a situation, I can inductively analyze and state the results.	4	5	0.75	0.73	<.01	I can describe the results of a problem using inferential evidence.
5.1.2	I describe the antecedents in a given problem when stating the results.	3	5	0.687	0.609		Retain -I can describe the history of a given problem when stating the results
5.1.3	I state accurately each part of the problem to draw inferences.	2	5	0.562	0.578		Rejected
5.1.4	I state the reason for holding a particular view.	3	5	0.75	0.734	<.01	Retained
5.1.5	I ensure that the inference presented is accurate.	3	5	0.562	0.58		Rejected
5.1.6	I objectively judge and state the beliefs/opinions that refute what I already think.	4	5	0.687	0.656		Rejected
5.1.7	I precisely articulate the inferences to address a given problem.	4	5	0.687	0.656		Rejected
5.1.8	I logically present inferences to address a given problem.	4	5	0.81	0.75	<.01	Retained
5..1.9	I clearly present the inferences to address a given problem.	4	5	0.75	0.703	<.05	Retained
5.1.10	I clarify another's propositions and accept with reasonable evidence. (M)	3	5	0.562	0.531		Rejected
5.1.11	I monitor the process of reasoning to obtain results.	1	5	0.437	0.53		Rejected
5.1.12	It is uninteresting to go over the results to examine the validity of the finding. (-ve)	3	5	0.437	0.422		Rejected

FIVE items retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Explanation- subscale: Justifying Procedures

Category: Explanation

Item #	Subscale: Justifying Procedures	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
5.2.1.	I state my choice of using a particular method to solve the problem.	4	4	0.937	0.88	<.01	Retained
5.2.2	My values enable me to analyse pros and cones before taking an action.	3	5	0.562	0.594		Rejected
5.2.3	I keep a log of the steps followed in working through a problem.	3	5	0.81	0.73	<.01	Retained
5.2.4	I explain how I understand a key concept for clarity in my thinking.	3	5	0.875	0.75	<.01	I can explain a key concept to clarify my thinking
5.2.5	I state the criteria when evaluating a piece of literature.	3	5	0.81	0.734	<.01	Retained
5.2.6	I explain the assumptions of using a particular technical method.	3	2	1	0.84	<.01	Retained
5.2.7	I explain the prerequisites for satisfaction of using a method for problem solving.	3	5	0.75	0.7	<.05	Rejected (confusing)
5.2.8	I report the strategy used in deriving a decision with reasons.	3	3	0.937	0.84	<.01	Retained
5.2.9	I design a graphic display to show the evidence to reach an inference.	3	5	0.75	0.7	<.05	Retained (I make a flow chart to show the process of deriving the conclusion.)
5.2.10	I accurately record the processes involved in drawing inferences.	3	5	0.81	0.719	<.05	I provide written record of the process involved in drawing inferences.
5.2.11	I reflect on my opinions and reasons to ensure my premises are correct. (M)	3	5	0.687	0.641		Rejected
5.2.12	I present the evidence to support my conclusion.	3	4	0.875	0.813	<.01	Retained

NINE items are retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category : Explanation – subscale: Presenting Arguments

Category: Explanation

Item #	Subscale: Presenting Arguments	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
5.3.1.	I write essays with adequate arguments supported with reasons for a given policy or situation.	4	5	0.875	0.83	<.01	Retained
5.3.2	I anticipate reasonable criticisms one might raise against one's view points.	3	4	0.81	0.766	<.01	Retained
5.3.3	I respond to reasonable criticisms one might raise against one's view points.	4	3	0.937	0.86	<.01	Retained
5.3.4	I clearly articulate evidence for my own view points.	4	3	0.937	0.88	<.01	Retained
5.3.5	I present more evidence or counter evidence for another's points of view.	4	3	0.875	0.891	<.01	Retained
5.3.6	I accept another person's thinking on matters of personal concern.	3	5	0.5	0.578		Rejected
5.3.7	I provide reasons for rejecting another's claim.	4	3	0.937	0.891	<.01	Retained
5.3.8	I explain the concepts supported with reasons or examples.	4	3	0.937	0.875	<.01	Retained
5.3.9	I ask questions to validate my own inferential judgment.	3	5	0.687	0.641		Rejected
5.3.10	I ask questions to validate others' inferential judgment.	3	5	0.75	0.656		Rejected
5.3.11	I make links between concepts to derive hypotheses. (M)	3	5	0.562	0.547		Rejected
5.3.12	I clearly articulate the reasons for accepting a claim.	4	5	0.875	0.844	<.01	Retained

EIGHT items retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category – Self-regulation- Subscale : Self-examination

Category: Self regulation

Item #	Subscale: Self-Examination	Mdn	Range	***I-CVI***	***VI_k***	*P-value*	*Remarks*
6.1.1.	I reflect on my opinions and reasons to ensure my premises are correct.	4	5	0.937	0.90	<.01	Retained
6.1.2	I recognize my deficiencies and seek to enrich my store of knowledge.	4	5	0.937	0.891	<.01	Retained
6.1.3	I make no changes or revisions on my points of view as I am sure they are correct. (-ve)	3	5	0.562	0.5		Rejected
6.1.4	I suspend another's point of view or opinions that are not supported by a justifiable rationale.	3	5	0.562	0.63		Rejected
6.1.5	Some of my colleagues are inherently poor in stating their assumptions. (-ve)	2	5	0.31	0.34		Rejected
6.1.6	I review sources of information to ensure important information is not overlooked.	3	5	0.75	0.7	<.05	Retained
6.1.7	I explain the concepts supported with reasons and examples.	3	5	0.5	0.53		Rejected
6.1.8	I examine and consider ideas, beliefs and viewpoints even when others do not agree.	3	5	0.75	0.69	<.05	Retained (remove beliefs)
6.1.9	Thinking drives my all actions. (-ve)	3	5	0.625	0.58		Rejected
6.1.10	I always examine my values, thoughts / beliefs based on reasons and evidence.	4	5	0.875	0.766	<.01	Retained (remove always)
6.1.11	I continuously assess my targets and work towards achieving them.	3	5	0.875	0.75	<.01	Retained
6.1.12	I review my reasons and reasoning process in coming to a given conclusion.	4	5	1	0.89	<.01	Retained
6.1.13	I reflect on my thinking skills to identify their consistent and inconsistent applications in understanding or solving a problem.	4	5	0.875	0.828	<.01	how consistently I use my thinking skills in solving a Problem.
6.1.14	I analyze areas of consistencies and inconsistencies in my thinking.	4	5	0.875	0.844	<.01	Retained

NINE items are retained

Appendix I

Results of Content Validation Analysis

Content Validity Coefficient for items for Category: Self-regulation – subscale: Self-correction

Category: Self-regulation

Item #	Subscale: Self-Correction	Mdn	Range	***I- CVI***	VI_k	*P-value*	*Remarks*
6.2.1.	Given a factual deficiency, I will revise the work to correct my opinions and beliefs.	4	5	0.875	0.8	<.01	Retained (willingly)
6.2.2	I willingly modify my position or opinions if revisions warrant change.	4	5	0.937	0.89	<.01	Retained (if need be)
6.2.3	I rarely validate the opinions of others and go by the majority. (-ve)	3	5	0.5	0.47		Rejected
6.2.4	Self-examination is not needed when I am confident of the propositions I make in a group. (-ve)	3	5	0.562	0.55		Rejected
6.2.5	I feel I am correct on my decisions most of the time, so I rarely re-examine resolutions. (-ve)	3	5	0.625	0.63		Rejected
6.2.6	I take charge of my thinking and continually revise and rethink strategies to improve it.	4	2	1	0.92	<.01	Retained (I continually revise and rethink .....
6.2.7	I can participate effectively in discussions with an interdisciplinary team.	4	5	0.81	0.77	<.01	Retained
6.2.8	I respect others' points of view even if they contradict mine.	4	5	0.81	0.78	<.01	Retained
6.2.9	I regularly reflect and critique on my own thoughts.	4	4	0.937	0.89	<.01	Retained
6.2.10	I reflect on my thinking to improve the quality of my judgment.	4	2	1	0.89	<.01	Retained
6.2.11	I think self-awareness is the best way to understand others.	3	5	0.75	0.734	<.01	Retained (I am aware of my strengths and weakness.)
6.2.12	I am able to control the undue influence of my values and beliefs on my thinking.	3	5	0.687	0.7	<.05	I am aware of my values and belief and control its undue influence on my thinking
6.2.13	Being aware of the social norms that restrict my thinking, I control their influence on my thinking.	4	5	0.875	0.813	<.01	Retained
6.2.14	I seek relevant information to support other's points of view.	3	5	0.562	0.563		Rejected(M)

TEN Items retained

Appendix – J
Detailed Description of Interpretation of Exploratory Data Analyses

Core Scale 1 Interpretation. Summated scores for each of the three subscale (SS) *categorization* (SS 1), *decoding significance* (SS 2), and *clarifying meaning (*SS 3) of the core scale interpretation[1] were explored for normality of data distribution. The results showed one outlier (case 251) in SS 1 with a score of 0 for the scale with maximum possible score of 42. Two outliers (cases 115 and 155) were found in SS 3 with a score of five (5) and six (6) respectively for the scale with a maximum possible score of 42. When the analysis was repeated with exclusion of these cases the results showed reasonable values of Mean, SD, skew, kurt. The Mean and Mdn were very close to each other and the value of Mdn was within the upper limit of mean at a 95% Confidence Interval (CI). (Table 4.4)

Core Scale 2 Analysis. Data screening for the three subscales *examining ideas* (SS 4), *detecting arguments* (SS 5), and *analysing arguments* (SS 6) of the core cognitive skill revealed two (2) outliers in SS 4 (cases 99 & 109) with a score of 4 and 2 respectively for the scale with maximum possible score of 36. Analysis with exclusion of these cases yielded normal distribution with improved values of Mean, Mdn, SD, skew and kurt. (Table 4.4)

Core Scale 3 Evaluation. Exploratory Data Analysis for the summated scores of the two subscales (SS 7 & SS 8) in evaluation revealed *assessing claim* (SS 7) with a normal distribution and *assessing arguments* (SS 8) with three outliers (Cases 95, 96, 110) with scores of 5, 10 and 10 respectively for the scale (maximum possible score of 48). Analysis with exclusion of these cases provided better distribution of data and improved values of skew and kurt. (Table 4.4)

---

[1]Core scales are indicated in the text with regular font and subscales are with italic font

Table 4.4

Exploratory Data Analysis for Composite Scores for Each Core Scale (improved results)

	1. Interpretation (868)			2. Analysis (882)			3. Evaluation (882)	
Sub Scales →	1.1 *Cat	1.2 DS	1.3 CM	2.1 EI	2.2 DA	2.3 ANA	3.1 AC	3.2 AA
Mean	25.90	28.17	27.15	23.82	22.79	22.30	19.66	30.62
Lower limit (CI)	25.39	27.75	26.70	23.42	22.38	21.89	19.31	30.12
Upper limit (CI)	26.42	28.59	27.59	24.22	23.08	22.71	20.01	31.13
Mdn	26.00	29.00	27.00	25.00	23.00	22.50	19.77	31.00
Variance	60.29	41.29	45.95	37.08	39.58	38.68	20.00	57.692
SD	7.764	6.426	6.78	6.089	6.291	6.219	7.37	7.596
Skewness	-.206	-.348	-.167	-.302	-.067	-.160	-.269	-.202
SE of skewness	.082	.082	.082	.082	.082	.082	.082	.082
Kurtosis	-.484	-.340	-.684	-.473	-.654	-.469	-.526	-.476
SE of kurtosis	.165	.165	.165	.165	.165	.165	.164	.164
Maximum	3.00	11.00	9.00	6.00	7.00	6.00	4.00	11.00
Minimum	42.00	42.00	42.00	36.00	36.00	36.00	30.00	48.00

*Cat: Categorization; DS: Decoding Significance; CM: Clarifying Meaning; EI: Examining Ideas; DA: Detecting Argument; ANA: Analysing Argument; AC: Assessing Claim; AA: Assessing Arguments.

Core Scale 4 Inference. Composites scores for the three subscales *querying evidence* (SS 9), *conjecturing alternatives* (SS 10), and *drawing conclusions* (SS 11) demonstrated two (2) outliers (cases 441 & 630) in SS 9, two (2) cases (403, 476) in SS 10, and five (5) cases (216, 338, 476, 486, 770) in SS 11. The maximum score in SS 9 and 10 was 42 each and for SS11 were 48. The scores obtained by these outlying cases were: SS 9 had a score ≤ 8, SS 10 with a score ≤ 7, and SS 11 with a score ≤ 10. Exclusion of these cases from the analysis improved the values of Mean, Mdn, SD, skew and kurt. (Table 4.5)

Core Scale 5 Explanation. Data screening for the three subscales *stating results* (SS 12), *justifying procedures* (SS 13), and *presenting arguments* (SS 14) in the core scale Explanation showed normal distribution in SS 12 and SS 13 and three outliers (cases 74, 476, & 700) in SS 14 scored ≤ 5, for the scale with a maximum possible score of 48. These cases were excluded and the analysis was repeated yielding improved values and normal distribution. (Table 4.5)

Core Scale 6 Self-regulation. The scale Self regulation with two subscales *self-examination* (SS 15) and *self-correction* (SS 16) were explored for normality and found that SS15 had a normal distribution and there were two (2) outliers (cases 216 and 141) in SS 16 (Table 4.5). Analysis of items for normality in this scale revealed three items with skewed distribution and outliers. These items were excluded from the analysis.

Table 4.5

Exploratory Data Analysis for Composite Scores for Each Core Scale (improved results)

	4. Inference (874)			5. Explanation (877)			6. Self Regulation (882)	
Sub Scales →	4.1 *QE	4.2 CA	4.3 DC	5.1 SR	5.2 JP	5.3 PA	6.1 SE	6.2 SC
Mean	28.23	28.78	31.12	18.37	32.71	29.73	37.41	42.59
Lower limit (CI)	27.77	28.33	30.62	18.01	32.08	29.20	36.83	41.95
Upper limit (CI)	28.69	29.23	31.61	18.73	33.33	30.26	37.98	43.22
Mdn	29.00	29.0	32.00	18.00	32.00	30.00	38.00	44.00
Variance	49.46	46.10	55.75	28.92	88.20	64.1	75.60	93.31
SD	7.032	6.789	7.466	5.378	9.392	8.007	8.695	9.659
Skewness	-.333	-.327	-.205	-.081	-.076	-.018	-.330	-.384
SE of skewness	.083	.083	.083	.083	.083	.083	.082	.082
Kurtosis	-.431	-.461	-.378	-.553	-.532	-.495	-.508	-.519
SE of kurtosis	.165	.165	.165	.165	.165	.165	.164	.164
Maximum	8.00	9.00	10.00	4.00	7.00	9.00	12.00	15.00
Minimum	42.00	42.00	48.00	30.00	54.00	48.00	54.00	60.00

*QE: Querying Evidence; CA: Conjecturing Alternatives; DC: Drawing Conclusions;
SR: Stating Results; JP Justifying Procedures; PA: Presenting Arguments;
SE: Self- Examination; SC: Self-Correction.

Exploratory descriptive analysis of the summated scores for the six core scale (Interpretation, Analysis, Evaluation, Inference, Explanation and Self-Regulation) results prompted six (6) cases more to be excluded in addition to the 13 identified with data screening for normality with individual items. Case 476 was an extreme case and was common to many scales. A total of 19 cases were excluded from further analysis of reliability and Exploratory Factor Analysis (EFA). Excluding 19 cases from the total 887 will not limit the application of other statistical techniques for further data analysis. The remaining sample size of 868 is “very good” according to Tabachnick and Fidell (2007, p. 613) for data reduction strategies such as EFA.

Appendix K

Exploratory Data Analysis: Evidence for Normality of Distribution of CTSAS

Item #	Indian sample (887)				Canada Sample (144)				India re-test Yr 3 & 4 (n=251)				India first test Yr 3 & 4,(n=410)			
(Missing Cases)	Skew	Kurt	Box Plot	Outliers	Skew	Kurt	Box Plot	Outliers	Skew	Kurt	Box Plot	Outliers	Skew	Kurt	Box Plot	Outliers
1. Interpretation																
1 cat (1)	-.484	-.987	N		-.871	-.342	N		-.689	-.772	N		-.719	-.677	N	
2 cat	.015	-.930	N		-.066	-.975	N		-.204	-.861	N		-.200	-.814	N	
3 cat	.105	-1.054	N		-.047	-.700	N		-.224	-.619	N		-.004	-1.004	N	
4 cat	-.646	-.643	N		-241	-.356	N		-.204	-.374	N		-.351	-.877	N	
5 cat	-.646	-.643	N		-.423	-.270	N		-.570	-.437	N		-.682	-.528	N	
6 cat	-.652	-.651	N		-.435	-.309	N		-.600	-.429	N		-.708	-.442	N	
7 cat	-.202	-.942	N		-.455	-.291	N		-.453	-.571	N		-.282	-.835	N	
8 deco(2)	-.447	-1.018	N		-.432	-.426	N		-.559	-.644	N		-.495	-.831	N	
9 deco	-.623	-.534	N		-.147	-.741	N		-.631	-.280	N		-.675	-.433	N	
10 deco*	-1.646	1.916	Flag?	13	-1.056	.491	N?		-1.176	-.469	N	2	-1.648	-2.101	Flag	52
11 deco	.286	-.896	N		-.386	-.509	N		-.332	-.696	N		-.275	-.769	N	
12 deco	-.414	-.811	N		-.384	-.463	N		-.493	-.359	N		-.435	-.748	N	
13 deco	-.417	-.712	N		-.494	-.253	N		-.505	-.478	N		-.437	-.613	N	
14 deco	-.108	-1.053	N		-.336	-.147	N		-.413	-.749	N		-.177	-.953	N	
15 clmg	-.272	-1.002	N		-.335	-.550	N		-.283	-.754	N		-.369	-.878	N	
16 clmg	-.582	-.792	N		-.722	-.463	N		-.632	-.331	N		-.619	-.661	N	
17 clmg	-.826	-.283	N		-.744	-.044	N		-.812	.025	N		-.867	-.185	N	2
18 clmg	-.215	-.804	N		-.222	-.446	N		-.541	-.350	N		-.281	-.775	N	
19 clmg	-.196	-.856	N		-.325	-.760	N		-.364	-.435	N		-.199	-.820	N	
20 clmg	-.484	-.666	N		-.318	-.497	N		-.430	-.588	N		-.591	-.414	N	
21 clmg	-.427	-.797	N		-.261	-.411	N		-.528	-.522	N		-.544	-.589	N	
2. Analysis																
22 exm (2)	-.471	-.661	N		-.545	-.207	N		-.421	-.449	N		-.450	-.707	N	
23 exm	-.220	-.773	N		-.024	-.394	N		-.243	-.473	N		-.154	-.792	N	
24 exm	-.640	-.387	N		-.531	-.490	N		-.583	-.473	N		-.609	-.426	N	
25 exm	-.817	-.259	N		-.796	-.724	N		-.833	-.174	N		-.827	-.183	N	1-(591)
26 exm	-.384	-.738	N		-.373	-.033	N		-.532	-.230	N		-.437	-.704	N	
27 exm	-.244	-.718	N		-.434	-.057	N		-.442	-.325	N		-.272	-.642	N	

Item #	Indian sample (887)				Canada Sample (143)				India retest(yr3&4, 251)				India first test yr 3 & 4,			
(Missing Cases)	Skew	Kurt	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers
28 darg (2)	-.612	-.528	N		-.859	-.756	N?	1	-.598	-.447	N		-.674	-.376	N	
29 darg	-.469	-.725	N		-.451	-.614	N		-.540	-.572	N		-.523	-.683	N	
30 darg	-.186	-1.045	N	2	-.620	-.047	N		-.488	-.518	N		-.281	-.959	N	
31 darg	-.194	-.926	N		-.458	-.309	N		-.276	-.605	N		-.290	-.878	N	
32 darg	-.457	-.532	N		-.557	.092	N		-.512	-.416	N		-.493	-.467	N	
33 darg	-.572	-.638	N		-.665	.068	N		-.524	-.453	N		-.762	-.357	N	
34anarg (2)	-.202	-.799	N		-.473	-.124	N		-.356	-.682	N		-.252	-.716	N	
35anarg	-.013	-.841	N		-.270	-.291	N		-.250	-.588	N		-.020	-.798	N	
36 anarg	-.054	-.469	N		-.220	-.389	N	1	-.628	-.214	N		-.711	-.304	N	
37anarg	-.412	-.572	N		-.221	-.482	N		-.473	-.230	N		-.443	-.510	N	
38anarg	-.186	-.616	N		-.105	-.424	N		-.332	-.579	N		-.348	-.408	N	
39anarg	-.723	-.288	N		-.689	-.422	N	1	-.753	-.056	N	2	-.767	-.240	N	
3. Evaluation																
40asscl (1)	-.432	-.591	N		-204	-.487	N		-.696	-.079	N		-.399	-.769	N	
41asscl	-.234	-.711	N		-.653	-.044	N		-.363	-.581	N		-.301	-.716	N	
42asscl	-.806	-.190	N		-.552	-.020	N		-.571	-.415	N		-.326	-.054	N	
43asscl	-.279	-.510	N		-.530	-.366	N		-.559	-.263	N		-.326	-.451	N	
44asscl (2)	-.632	-.428	N		-.374	-.129	N		-.538	-.540	N		-.712	-.258	N	
45asarg	-.593	-.478	N		-.683	-.344	N		-.676	-.090	N		-.574	-.412	N	
46asarg	-.105	-.428	N		-.360	-.714	N		-.182	-.721	N		.121	-.744	N	
47asarg	-.352	-.782	N		-.501	-.047	N		-.613	-.188	N		-.473	-.585	N	
48asarg	-.351	-.723	N		-.601	-.108	N		-.505	-.390	N		-.388	-.651	N	
49asarg	-.425	-.652	N		-.836	-.547	N		-.577	-.371	N		-.407	-.748	N	
50asarg	-.531	-.643	N		-.678	-.004	N		-.498	-.538	N		-.546	-.634	N	
51asarg	-.332	-.585	N		-.376	-.595	N		-.463	-.501	N		-.423	-.448	N	
52asarg	-.660	-.343	N		-.912	-.390	N		-.741	-.137	N		-.740	-.180	N	
4. Inference																
53qevid (3)	-.831	-.029	N		-.830	-.176	N		-.638	-.397	N		-.767	-.105	N	
54qevid	-.462	-.671	N		-.818	-.813	N		-.623	-.108	N		-.564	-.549	N	
55qevid	-.366	-.637	N		-.839	1.920	N		-.402	-.558	N		-.378	-.549	N	
56qevid	-.309	-.657	N		-.542	-.235	N	1	-.588	-.045	N		-.411	-.495	N	
57qevid	-.773	-.304	N		-.450	-.492	N		-.738	-.217	N		-.811	-.114	N	

Item #	Indian sample (887)				Canada Sample (143)				India retest(yr3&4, 251)				India first test yr 3 & 4,			
(Missing cases)	Skew	Kurt	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers
58qevid	-.431	-.534	N		-.463	.036	N	5	-.536	-.057	N		-.526	-.364	N	
59qevid	-.937	-.073	N		-.611	-.311	N	3	-.697	-.299	N		-1.061	-.308	N	4
60cjalt (4)	-.629	-.388	N		-.450	-.252	N		-.683	-.170	N		-.656	-.488	N	
61cjalt	-.643	-.446	N		-.448	-.445	N		-.493	-.646	N		-.725	-.270	N	
62cjalt	-.479	-.317	N		-.470	-.130	N		-.582	-.137	N		-.582	.061	N	
63cjalt	-.362	-.696	N		-.451	-.086	N		-.436	-.274	N		-.381	-.717	N	
64cjalt	-.649	-.243	N		-.535	-.090	N		-.854	-.586	N	6	-.691	-.235	N	2
65cjalt	-.622	-.337	N		-.678	-.187	N		-.464	-.620	N		-.675	-.097	N	
66cjalt	-.595	-.431	N		-.270	-.439	N		-.713	-.113	N		-.691	-.154	N	
67dcon (1)	-.122	-.930	N		-.084	-.748	N		-.147	-.918	N		-.123	-.952	N	
68dcon	-.308	-.626	N		-.613	-.387	N		-.423	-.446	N		-.360	-.565	N	
69dcon	-.956	.207	N	4	-.728	.105	N		-.895	.231	N		-1.107	.615	N	4
70dcon	-.300	-.523	N		-.560	-.062	N		-.409	-.471	N		-.332	.565	N	
71dcon	-.926	.009	N	4	-.744	.403	N		-.769	-.093	N		-1.102	.548	N	4
72dcon	-.513	-.700	N		-.071	-.640	N		-.510	-.468	N		-.611	-.472	N	
73dcon	-.450	-.346	N		-.613	-.245	N		-.658	-.032	N		-.606	-.028	N	
74dcon	-.308	-.669	N		-1.102	1.145	N	2	-.508	-.268	N		-.510	-.475	N	
5. Explanation																
75sresu	-.213	-.604	N		-.470	-.436	N		-.355	-.605	N		-.239	-.611	N	
76sresu	-.218	-.774	N		-.515	-.095	N	3	-.295	-.766	N		-.208	-.826	N	
77sresu	-.371	0.707	N		-.793	-.154	N		-.640	-.331	N		-.501	-.612	N	
78sresu	-.239	-.648	N		-.388	-.345	N	9	-.400	-.526	N		-.386	-.533	N	
79sresu	-.196	-.601	N		-.363	-.526	N		-.418	-.585	N		-.296	-.526	N	
80Jproc	-.552	-.517	N		-.853	-.951	N		-.720	.018	N		-.636	-.361	N	
81Jproc	-.182	-.647	N		-.029	-1.139	N		-.382	-.423	N		-.176	-.669	N	
82Jproc	-.314	-.654	N		-.569	-.064	N		-.366	-.711	N		-.335	-.817	N	
83Jproc	-.179	-.767	N		-.269	-.271	N		-.279	-.718	N		-.206	-.748	N	
84Jproc	-.254	-.647	N		-.220	-.028	N		-.288	-.553	N		-.316	-.697	N	
85Jproc	-.226	-.638	N		-.516	-.118	N	5	-.273	-.447	N		-.323	-.405	N	
86Jproc	-.061	-1.126	N	4	-.388	-.982	N		-.387	-.817	N		-.120	-1.106	N	
87Jproc	-.042	-.1.007	N	3	-.216	-.903	N		-.215	-.902	N		-.079	-.964	N	
88Jproc	-.738	-.301	N		-.678	-.017	N		-.635	-.288	N		-.908	-141	N	
89parg (5)	-.176	-.958	N		-.887	-.320	N	17	-.263	-.743	N		-.194	-.827	N	

Item #	Indian sample (887)				Canada Sample (143)				India retest(yr3&4, 251)				India first test yr 3 & 4,			
Item # (Missing cases)	Skew	Kurt	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers	Sk	K	Box Plot	Outliers
90parg	-.188	-.806	N		-.654	-.146	N		-.367	-.461	N		-.253	-.806	N	
91parg	-.347	-.738	N		-.550	-.154	N		-.317	-.461	N		-.253	-.806	N	
92parg	-.389	-.600	N		-.765	.958	N		-.413	-.459	N		-.529	-.452	N	
93parg	-.243	-.645	N		-.177	-.317	N		-.520	-.404	N		-.229	-.632	N	
94parg	-.486	-.590	N		-.738	-.559	N		-.633	-.269	N		-.616	-.325	N	
95parg	-.362	-.734	N		-.994	.962	N		-.634	-.101	N		-.424	-.557	N	
96parg	-.377	-.621	N		-.576	.111	N		-.510	-.054	N		-.434	-.509	N	
6. Self-Regulation																
97sexm (2)	-.322	-.694	N		-.824	.422	N		-.567	-.325	N		-.416	-.485	N	
98sexm	1.075	.191	N		-.665	.013	N		-.907	-.081	N	1	-1.186	-.667	N	2
99sexm	-.511	-.399	N		-.704	-.516	N		-.519	-.467	N		-.533	-.307	N	
100sexm	-.568	-.452	N		-.810	.971	N		-.753	-.063	N		-.590	-.377	N	
101sexm	-.765	-.129	N		-.369	-.015	N		-.753	-.063	N		-.852	-.228	N	4
102sexm	-.605	-.357	N		-.479	-.475	N		-.602	-.077	N		-.691	-.184	N	
103sexm	-.513	-.482	N		-.468	-223	N		-.443	-.442	N		-.584	-.364	N	
104sexm	-.455	-.629	N		-.527	-.261	N		-.632	-.036	N		-.443	-.687	N	
105sexm	-.404	-.469	N		.429	-.317	N		-.641	-.211	N		-.490	-.466	N	
106scorr (2)	-.505	-.631	N		-.411	-.547	N		-.560	-.248	N		-.564	-.470	N	
107scorr	-.580	-.404	N		-.376	-.488	N		-.659	-.282	N		-.701	-.056	N	
108scorr	-.531	-.548	N		-.160	-.910	N		-.550	-.311	N		-.459	-.716	N	
109scorr	-.498	-.559	N		-.594	-.342	N		-.546	-.358	N		-.657	-.299	N	
110scorr*	-.801	-.185	N		-1.046	-.689	Flag?	9	-.636	-.372	N	1	-.978	-.192	N	4
111scorr	-.286	-.743	N		-.464	-.562	N		-.364	-.542	N		-.432	-.450	N	
112scorr	-.543	-.516	N		-.332	-.555	N		-.647	-.319	N		-.611	-.366	N	
113scorr*	-1.168	-.468	Flag?	3	-.635	-.364	N	Flag?	-1.002	-.150	N	1	-.1.463	1,536	Flag	4
114scorr*	-.980	-.357	Flag?		-.766	-.446	N	6	-.912	.437	N	4	-.904	.046	N	1
115scorr	-.689	-.142	N		-.675	-.810	N		-.725	-.117	N		-.752	-.021	N	

* These items (10, 110, 113, & 114) were excluded from further analysis.
The remaining items meet the normality of distribution.

Appendix L

Detailed Description of the Reliability Analysis

A detailed description of the reliability analysis for CS 1 is in the text. This appendix gives the details for the other core scales which demonstrated low item total correlations.

In the Indian sample, similar findings were noted in Analysis (CS 2). With overall Cronbach α 0.857, item 33 showed item total correlation < 0.3 (0.288, item α 0.858) which is negligible. In this CS, the *analyzing arguments* (SS5 - 6 item scale), item 28 and *detecting arguments* item 33 showed item total correlation < 0.3 (0.286 and 0.232) and an overall α value for the SS = 0.666. The analysis was repeated with the exclusion of these two items, which resulted in an improved reliability for this subscale with α = 0.690. The item total statistics for improved reliability are presented in the SPSS output Table 4.6.2. A decision was made to perform EFA with and without this item. The remaining four items have an average inert-item correlation of 0.56 which is good to yield an alpha level of 0.80 (DeVellis, 2003).

Table 4.6.2

SPSS output: Item Total Statistics – CS 2 Analysis

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
29. I figure out what my reasons are for supporting or not supporting the author's claim.	10.83	15.672	.439	.203	.646
30. I ask questions in order to seek evidence to support or refute the author's claim.	11.17	13.918	.550	.310	.572
31. I figure out if author's arguments include both for and against the claim.	11.23	14.921	.502	.264	.606
32. I seek supporting reasons when a person is advancing a claim.	10.80	16.761	.403	.165	.666

The CS 3 Evaluation analysis showed an overall Cronbach α 0.817. Item 46 *"I rarely examine the flaws in an argument"* showed item total correlation < 0.3 (0.260, and item α = 0.820), which is a little greater than the overall α for the scale. However, repeating the analysis and excluding this item slightly improved the CS α to 0.822. The SS 8 in this scale with an original α = 0.733 improved to 0.746 and the reliability of all other items in the scale improved to ≥ 0.7. The decision was made to run EFA with and without this item.

The overall reliability of CS 4 was 0.875 although in the core scale the item total correlation for the item 67 *"I depend on statistical techniques for drawing conclusions"* was 0.264 (α = 0.876) and was very similar to the overall α for the CS. Deletion of this item showed negligible results. However, this item belongs to the SS 11 and separate analysis of SS 11 showed item total correlation of 0.298 and α = 0.748. However, this is a negligible change. It was decided to run EFA with and without this item. The remaining two core scales 5 and 6 showed very good reliability.

Appendix M

Test-Retest - Stability Reliability Values

Table 4.7

Test Retest Reliability: Pearson-r and Spearman-rho- N= 251 (India)

Core & Sub scales	India Test - Retest Yr 3 & 4 (N=251)		Pearson
	Pearson r	Spearman rho	R^2 %
CS 1. INTERPRETATION	.500**	.509**	25
SS 1. Categorization	.483**	.487**	23.33
SS 2. Decoding Significance	.361**	.369**	13.03
SS 3. Clarifying Meaning	.346**	362**	11.97
CS 2. ANALYSIS	.431**	.431**	17.67
SS 4. Examining Ideas	.411**	.425**	16.89
SS 5. Detecting Arguments	.368**	.377**	13.54
SS 6. Analysing Arguments	.394**	.401**	15.52
CS 3. EVALUATION	.527**	.529**	27.77
SS 7. Assessing Claims	.457**	.462**	20.88
SS 8. Assessing Arguments	.460**	.478**	21.16
CS 4. INFERENCE	.500**	.482**	25
SS 9. Querying Evidence	.463**	.429**	21.44
SS 10. Conjecturing Alternatives	.423**	.408**	17.89
SS 11. Drawing Conclusions	.416**	.403**	17.31
CS 5. EXPLANATION	.464**	.479**	21.53
SS 5. Stating Results	.375**	.382**	14.06
SS 11. Justifying procedure	.286**	.281**	8.18
Ss 12. Presenting Arguments	.392**	.397**	15.36
CS 6. SELF-REGULATION	.518**	.526**	26.83
SS 15. Self Examination	.493**	.494**	24.30
SS 16. Self Correction	.451**	.441**	20.34

**. Correlation is significant at the 0.01 level (2-tailed).

Appendix N

Assumptions and Decisions Criteria for Using Exploratory Factor Analysis

Figure 4.3. Assumptions and Decisions Criteria for Using Exploratory Factor Analysis.

Assumptions	Criteria /Recommended acceptable values
1. Multivariate Normality 1.1. Exploratory Data Analysis (EDA) 1.2. Sensitivity analysis.	Examination for skewness and kurtosis, and missing data. Examination of histograms and probability curve and stem and leaf. Exclusion of items and cases with missing values, outliers which might influence factor solution.
2. Factorability of R (Pearson correlation) 2.1. Linearity 2.2. Bartlett's test of Sphericity	Inspection of R-Matrix for Absence of Multicollinearity and singularity. Squared Multiplication Correlation (SMC) greater than 0.3 ensuring variables are fairly correlated and Greater than 0.9 indicating multicollinearity and a value of one (1) indicating singularity (Field, 2009). Assess scatter plots for linearity among pairs of variables. Significant Chi-Square reveals the R-Matrix is not an identity matrix and there exists some relationship between the variables.
4. Sample size	Large enough that correlations are reliably estimated. General *rule of thumb* is: 1. A sample size 300 is good, 2. 500 is very good, and 3. 1000 excellent (Comery & Lee, 1992). OR 4. Ten to 15 cases per variable (Field, 2009).
5. Kaiser-Mayor-Olkin Measure of Sampling Adequacy(KMO)	1. Values between 0.5 and 0.7 are mediocre, 2. Values 0.7 and 0.8 good, 3. Values 0.8 and 0.9 are great, and 4. Values above 0.9 are superb correlation between pairs of variables (Filed, 2009).
6. Kaiser Criterion for Retaining Factors	1. Eigenvalues greater than one (1) (eigenvalue one represents a substantial amount of variance).

	2. A value of communalities is 0.3 and above. (communalities: explains the amount of variance in each variable by the retained factors). 3. Examination of Cattle's scree plot for retaining factors that lie above the point of inflection. 4. Factor extraction using first orthogonal rotation and Oblimin rotations, especially after the varimax rotation if the transformation matrix is unsymmetrical.
6.1 Factor Loadings:	Comrey and Lee (1992) suggested that: a. loading in excess of .71 is excellent and yields 50% of overlapping variance, b. 0.63 (40% of overlapping variance) very good and c. 0 .55 (30% of overlapping variance) good and d. 0.45 (20% of overlapping variance) fair (as cited in Tabachinic and Fidell, 2007). Hence, a value of 0.45 was set as cut off point for obtaining a substantive importance of factor loadings, and for the purpose of interpretation. (Variables' loading with greater weights is an indication that the variables are more a pure measure of the factor or variables are homogeneous and together share the conceptual meaning of the construct).
6.2. Interpretability Criteria.	a. There must be at least three items for a factor with significant loadings. b. Variables that load on a factor share conceptual meaning. c. Variables load on different factors measure different constructs. d. The rotated factor pattern demonstrates simple structure. e. The variables relatively load high on one factor and low on others. f. If an item loads on to more than one factor the item will be retained with the factor that shows the highest value.

Appendix O

Results of EFA Analysis

Table 4.9.1

SPSS output: Factor Loadings for Exploratory Factor Analysis of the CS 1- Interpretation Canada Sample (N=144)

Item #	Items (Variables)	Component			Commu h^{2*}
		1 DS*	2 Cat*	3 CM*	
9.	I clarify the meaning of an individual's gesture in a given situation.	.804			.604
11.	I try to identify the social importance of the information presented in the texts.	.732			.494
13.	I examine the values rooted in the information presented.	.710			.610
8.	**I figure out a person's purpose in asking a question.**	**.615**			.421
14.	I identify the author's views and intentions in the issue presented.	.559			.388
12.	I look for various relationships among concepts to understand the meaning.	.450			
20.	I figure out the meaning of another's point of view.	---	---	---	.509
1.	When presented with a problem, first I try to figure out the content of the problem.	---	---	---	.327
5.	I break the complex ideas into manageable sub-ideas.		-.815		.635
4.	I break down problem into discrete parts.		-.794		.619
6.	I categorize similar and related information into groups		-.724		.531
2.	I sort the information into sub sets.		-.708		.550
3.	I classify data using a framework.		-.702		.517
7.	I classify whole information into specifics.		-.551		.367
17.	I clarify my thoughts by explaining to someone else.			.859	.641
16.	I figure out an example which explains the concept /opinion.			.714	.484
15.	I restate another person's statements to clarify the meaning.			.649	.469
21.	I seek clarification of the meanings of another's opinion or points of view.			.608	.547
19.	I look for analogies of the words and concepts to clarify meaning.			.589	.581
18.	I try to differentiate between opinions and ideas to remove ambiguity.			.546	.543
	Percent of Variance 51.01%				

*DS: Decoding significance; Cat: Catgorization; CM: Clarifying Meaning; Commu h^2: Communalities.

Appendix O 1

Comparison of the Results of Assumption Tests for Performing Exploratory Factor Analysis

No	Scales	India Total (N= 887)			India Gr. 1 (n= 477)			India Gr. 2 (n= 440)			Canada (N=144)			Ind	Can
	Factors Extracted (F)	KMO	Variance %	Bartlett's Test	KMO	Variance	Bartlett's Test	KMO	Variance	Bartlett's Test	KMO	Variance	Bartlett's Test	Rel* α	Rel* α
CS 1 N=882	Interpretation F 5 Varimax F 3 Oblimin	.889	48.591 38.124	3244.299* *df.* (190)	.836 (442)	F 6 V* 51.841 F 4 O* 41.554 F 3 O 35.619	1354.617* *df.* (190)	.885 (440)	F 4 V 46.638 F 3 O 41,552	20.33.27 * *df.* (190)	.863	F 3 V 51.01	1084.792 * *df.*(190)	.834	.890
CS 2 N=885	Analysis F 3 Varimax F 3 Oblimin	.918	42.555	3415.699* *df.* (153)	.888 (442)	F 4 V 44.808 F3 O 39.121	1420.977* *df.* (153)	.912 (438)	F 3 V 45.916 ***	2149.21* *df.* (153)	.892	F 4 V 63.204 F 3 O 57.367	1252.204 * *df.* (153)	.857	.917
CS 3 N=882	Evaluation F 2 Varimax F 2 Oblimin	.888	40.222	2294.932* *df.* (78)	.856 (442)	F 3 V 46.111 F 2 O 38.160	996.215* *df* .(78)	.890 (440)	F 3 V 50.681 F 2 O 42.774	1364.44* *df* .(78)	.838	F 2 V 58.351 F 2 O 49.519	722.674* *df.* (78)	.822	.836
CS 4 N=875	Inference F 4 Varimax F 3 Oblimin	.933	44.333 40.231	4528.371* *df.* (231)	.899 (439)	F 4 V 43.132 F 3 O 38.209	2006. 474* *df.* (231)	.912 (436)	F 5 V 50.762 F 3 O 41.294	2487.31* *df.* (231)	.900	F 4 V 57.587 F 3 O 52.708	1475.230 * *df.*(231)	.875	.927
CS 5 N=877	Explanation F 4 Varimax F 3 Oblimin	.938	49.592 45.444	5828.932* *df.* (231)	.908 (439)	F 5 V 51.775 F 3 O 41.990	2592.428* *df.* (231)	.924 (438)	F 4 V 53.470 F 3 O 47.824	3459.99* *df.* (231)	.894	F 4 V 63.650 F 3 O 59.056	1848.561 * df. (231)	.748	.845
CS 6 N=882	Self- Regulation F 2 Varimax F 2 Oblimin	.939	44.301	4300.756* *df.* (120)	.917 (442)	F 3 V 48.738 F 2 O 41.849	1958.407* *df.* (120)	.940 (440)	F 1 V 40.205 F 2 O 46.387**	2418.18* *df.*(120)	.911	F 3 V 63.598 F 2 O 56.932	1303.223 * *df.* (120)	.887	.932

*Significant, $p < .001$. ** (Eigenvalues for two factors 6.433 and .989); *V – Varimax ; *O – Oblimin. Rel*=Reliability-India, Reliability Canada.

*** Varimax and Oblimin rotations 3 factor solution did not provide an interpretable solution. Item 33 which had low reliability and loaded as a single item in factor three (3) when excluded from analysis produced better interpretable solution and more accurate distribution of items across the factors with better loading weights.

Appendix O 2

Table 4.10

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 2 Analysis -Indian Sample (N=882)

Pattern Matrix[a]

Item #	Items (Variables)	Component 1 AA*	 2 EI*	 3 DA*	 Comu h^{2*}
38	I figure out the assumptions implicit in the author's reasoning.	.628			.469
37	I figure out the process of reasoning for an argument.	.528			.475
31	I figure out if author's arguments include both for and against the claim.	.689			.445
35	I figure out unstated assumptions in one's reasoning for a claim.	.664			.411
30	I ask questions in order to seek evidence to support or refute the author's claim.	.562			.409
36	I look for the overall structure of the argument.	--	--	--	.432
29	I figure out what my reasons are for supporting or not supporting the author's claim.	.571			.321
34	Given a brief argument, I identify the author's chief claim.	.548			.389
27	Given a situation, I look for specifics to analyze its various aspects.	--	--	--	.335
39	I look for evidence for the reasons supporting the arguments.	--	--	--	.361
25	I look for relevant information to answer the question at issue.		.760		.512
22	I examine the similarities and differences among the opinions posed for a given problem.		.652		.429
23	I examine the interrelationships among concepts or opinions posed.		.608		.456
26	I examine the proposals for solving a given problem.		.686		.420
24	I look for supporting reasons when examining opinions.		.622		.390
28	Given a paragraph to read: I determine the main claim.	--	--	--	.227
33	My arguments will depend on the sources of supporting information.		.	.711	.665
32	I seek supporting reasons when a person is advancing a claim.				.447
		.519			
	Percent of Variance 42.555%				
	Reliability α (Core scale) 0.837	0.790	0.702	---	

Factor 1: *AA-Analysing Arguments (8 items); Factor 2: *EI-Examining Ideas (5 items);
Factor 3: *DA-Detecting Arguments (1 item). Zero loading (3 items). 14 of the 18 items retained.
h^2 : Communalities

Appendix O 3

Table 4.10.1

SPSS output:– Factor Loadings for Exploratory Factor Analysis of CS 2 Analysis with Inclusion and Exclusion of Item 33 -Indian Sample (N=882)*

Item #	Items (Variables)	Component (Varimax)			Component (Oblimin)			Component (excluding 33)		
		1 AA*	2 EI*	3 DA*	1 AA*	2 EI*	3 DA*	1 AA*	2 EI*	3 DA*
38	I figure out the assumptions implicit in the author's reasoning.	.658			.660			.565		
37	I figure out the process of reasoning for an argument.	.638			.570			.732		
31	I figure out if author's arguments include both for and against the claim.	.629			.682					.594
35	I figure out unstated assumptions in one's reasoning for a claim.	.544			.701					.710
30	I ask questions in order to seek evidence to support or refute the author's claim.	.543			.524			--	--	--
36	I look for the overall structure of the argument.	.528			--	--	--	.739		
29	I figure out what my reasons are for supporting or not supporting the author's claim.	.497			.521			--	--	--
34	Given a brief argument, I identify the author's chief claim.	.450			.520			--	--	--
27	Given a situation, I look for specifics to analyze its various aspects.	--	--	--	--	--	--	--	--	--
-39	I look for evidence for the reasons supporting the arguments.	--	--	--	--	--	--	.547		
25	I look for relevant information to answer the question at issue.		.678			.763			.676	
22	I examine the similarities and differences among the opinions posed for a given problem.		.609			.618			.551	
23	I examine the interrelationships among concepts or opinions posed.		.597			.581			-.494	
26	I examine the proposals for solving a given problem.		.597			.663			.739	
24	I look for supporting reasons when examining opinions.		.573			.601			.577	
28	Given a paragraph to read: I determine the main claim.	--	--	--				.536		
33	My arguments will depend on the sources of supporting information.			.799			.741	Item Excluded		
32	I seek supporting reasons when a person is advancing a claim.			.485	.468			--	--	--

*Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. Rotation converged in 15 iterations.
*AA : Analysing Arguments; EI: Examining Ideas; DA: Detecting Arguments.

Appendix O 4

Table 4.10.2

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 2 Analysis – Canada Sample (N=144)

Item #	Items (Variables)	Component 1 AA*	2 EI*	3 DA*	Comu h^{2*}
38	I figure out the assumptions implicit in the author's reasoning.	.772			.711
37	I figure out the process of reasoning for an argument.	.823			.791
31	I figure out if author's arguments include both for and against the claim.			.605	.543
35	I figure out unstated assumptions in one's reasoning for a claim.	.652			.646
30	I ask questions in order to seek evidence to support or refute the author's claim.			.687	.635
36	I look for the overall structure of the argument.			.687	.568
29	I figure out what my reasons are for supporting or not supporting the author's claim.			.642	.675
34	Given a brief argument, I identify the author's chief claim.	--	--	--	.528
27	Given a situation, I look for specifics to analyze its various aspects.		.511		.430
39	I look for evidence for the reasons supporting the arguments.		.580		.466
25	I look for relevant information to answer the question at issue.		.643		.698
22	I examine the similarities and differences among the opinions posed for a given problem.		.711		.513
23	I examine the interrelationships among concepts or opinions posed.		.733		.540
26	I examine the proposals for solving a given problem.		.580		.459
24	I look for supporting reasons when examining opinions.		.635		.527
28	Given a paragraph to read: I determine the main claim.			.813	.611
33	My arguments will depend on the sources of supporting information.	.741	.640		.415
32	I seek supporting reasons when a person is advancing a claim.		.475		.574
	Percent of Variance	57.367%			

Factor 1: *AA: Analyzing arguments (4 items) Factor 2: *EI: Examining ideas (9 items); item 33 loaded in Fact 1 and 2; Factor 3: *DA: Detecting arguments (5 items); 0 loading three (4) items. 14 of the 18 items retained

Appendix O 5

Table 4.11

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 3 Evaluation with Inclusion and Exclusion of item # 46 –*

A comparison of Indian Sample (N=882)and Canadian Sample (N=144).

Item No		Varimax (India)		Oblimin (India)		Oblimin (Canada)		Oblimin (India)		Comu h^{2*}
		1	2	1	2	1	2	1	2	
		AC*	AA*	AC	AA	AC	AA	AC	AA	
43.	I assess the contextual relevance of an opinion or claim posed.	.602		.615		**.661**	**-.603**	.596		.410
47.	I examine the logical strength of the underlying reason in an argument.	.599		.599		.684		.640		.446
45.	I assess the chances of success or failure in using a premise to conclude an argument.	.595		.635			.661	.639		.356
51.	I examine the logical reasoning of an objection to a claim.	.551		.533		.748		.626		.451
46.	I rarely examine the flaws in an argument.	.539		.594			.523	Excluded		.295
40.	I figure out the relevant factors to assess credibility of the opinion / belief raised.	.515		.506		.612		.591		.361
44.	I seek the accuracy of the evidence supporting a given judgment.	.482		.454			-.680	.541		.400
42.	I figure out if a given claim is true or false based on the supported knowledge.	--	--	--	--	**.498**	**-.723**	--	--	.277
49.	I search for additional information that might support or weaken an argument.		.784		-.810	.652			.782	.619
48.	I search for new data to confirm or refute a given claim		.742		-.752	.689			753	.561
50.	I ask questions when the assumptions supporting the arguments are false.		.613		-.605	.665			.624	.398
52.	I look for conclusions which are logical.		.480	--	--	.694		.506		.359
41.	I assess the credible authority of the source of information supporting the claim.	--	--	--	--	**-.540**	**.617**	--	--	.295
	Percent of Variance : 40.222%; Reliability α (Core Scale) 0.796	α ➡		0.746	0.647	0.845	0.744			

*Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations. Factor 1: *AC: Assessing Claim (7 items); *AA: Assessing Arguments (3 items); 0 loading (3) (Oblimin); *h^2 : Communalities

Appendix O 6

Table 4.12

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 4 Inference – A Comparison of Factor Loadings - Indian Sample (N=875) and Canadian Sample (N=144).

Item No		Component (Varimax)				Oblimin(Ind.)			Oblimin (CA)			Com
		1 DC	2 QE	3 DC	4 C A	1 DC*	2 QE*	3 CA*	1 DC	2 QE	3 CA	h^{2*}
66	I identify the consequences of various options to solving a problem.	.636				.542					.633	.469
64	I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.	.634				.679			.664			.479
61	Given a problem to solve, I develop a set of options for solving the problem.	.596				.578			.495			.400
62	Whether or not one agrees, I state the difficulties and the benefits of adopting a given a set of priorities for decision making.	.575				.523			.489			416
65	I talk to others to get feedback on various ways of constructing alternate hypotheses.	.556				--	--	--			.716	.387
60	I figure out alternate hypotheses / questions, when I need to solve a problem.	.497				--	--	--				.336
63	I systematically analyse the problem using multiple sources of information to draw inferences.	--	--	--		.604			.545			.434
55	I seek useful information to refute an argument when supported by unsure reasons.		.707				.789			.813		.562
54	I seek relevant information to support another's point of view.		.598				.663			.799		.424
56	I collect evidence supporting the availability of information to back up opinions.		.577				.620			.779		.457
57	I seek for evidence / information before accepting a solution.		.519				.504			.532		.405
58	I consider opposing views in support of information when controversial issues are examined.		.518				.538					.358

59	I cannot accept a conclusion without understanding the supporting evidence.		.491				.452		.568			.471
53	When developing a persuasive argument, I search for useful information to support my point of view.		.484				.471			.620		.350
71	I analyse my thinking before jumping to conclusions.			.687		.753			.606			.548
69	I arrive at conclusions that are supported with strong evidence.			.686		.620						.542
72	I confidently reject an alternative solution when it lacks evidence.			.544		--	--	--	.810			.373
70	I use both deductive and inductive reasoning to interpret information.			.478		.559			.640			.475
73	I figure out the logical relationship of the reasons supporting the conclusions.	--	--	- -		.493			.861			.431
74	I figure out the pros and cons of a solution before accepting it.	--	--	--		.528			.513			.393
67	I depend on statistical techniques for drawing inferences.				.807			.770				.655
68	I gather multiple source of information when conclusions are made from opposing views.				.466	--	--	--				.390
	Percent of Variance 40.231%; Reliability (Core scale) α = 0.858					.819	.736	--	.912	.825	--	

Factor 1 loaded with 10 items *DC: Drawing conclusions; Factor 2 loaded with 7 items *QE: Querying evidence and Factor 3 * CA: Conjecturing Alternatives with 1 item.

Appendix O 7

Table 4.13

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 5 Explanation – A Comparison of Indian Sample (N=882) with Canadian Sample (144).

Item No	Items (Variables)	Component (Varimax)				Oblimin* (Ind.)			Oblimin (CA)			Comu
		1 SR*	2 JP*	3 PA*	4 JP	1 SR	2 JP	3 PA	1 PA	2 JP	3 SR	h^{2*}
95	I explain the concepts supported with reasons or examples.	.623						**.474**	.574			.479
77	I state the reason for holding a particular view.	.613				---	---	---			.641	.492
76	I can describe the history of a given problem when stating the results	.591				---	---	---				.432
96	I clearly articulate the reasons for accepting a claim.	.553				---	---	---	.573		.527	.503
78	I can logically present results to address a given problem.	.525				**.539**						.464
88	I present the evidence to support my conclusion.	.503				---	---	---			.873	.415
75	I can describe the results of a problem using inferential evidence.	.467				**.533**			.643			.413
80	I state my choice of using a particular method to solve the problem.		.734			**.758**						.603
81	I keep a log of the steps followed in working through a problem.		.649			**.689**				.888		.559
82	I can explain a key concept to clarify my thinking.		.637			**.664**					.564	.500
83	I state the criteria when evaluating a piece of literature.		.501			**.515**				.591		.452
85	I report the strategy used in deriving a decision with reasons.		.445			**.502**				.530		.444
84	I explain the assumptions of using a particular method.	--	--	--	--	**.500**				.684		.422
79	I clearly present the inferences to address a given problem.	--	--	--	--	**.519**			.498			.426
90	I anticipate reasonable criticisms one might raise against one's view points.			.704				**.658**	.860			.567
91	I respond to reasonable criticisms one might raise against one's view points.			.663				**.641**	.784			.528
92	I clearly articulate evidence for my own view points.			.562				**.634**				.455
93	I present more evidence or counter evidence for another's points of view.			.527				**.581**	.693			.399
94	I provide reasons for rejecting another's claim.	.463		.509				**.700**	.511		.472	.480
87	I provide written record of the process involved in drawing inferences.				.778		**.754**			.832		.675
86	I make a flow chart to show the process of deriving the conclusion.				.745		**.717**			.758		.637
89	I write essays with adequate arguments supported with reasons for a given policy or situation.				.646		**.649**					.565
	Percent of Variance: 45.444%											
	Reliability (Core scale) α =0.876					0.834	0.713	715	.892	.670	.886	

*Oblimin: Factor 1 *SR : Stating results 9 items; Factor 2 *JP : Justifying Procedure 3 items; Factor 3 *PA : Presenting Arguments 6 items 0 Loading – 4 items.

Appendix O 8

Figure 4.5: Scatter plot for CS 6 Self-regulation (Varimax Rotation)

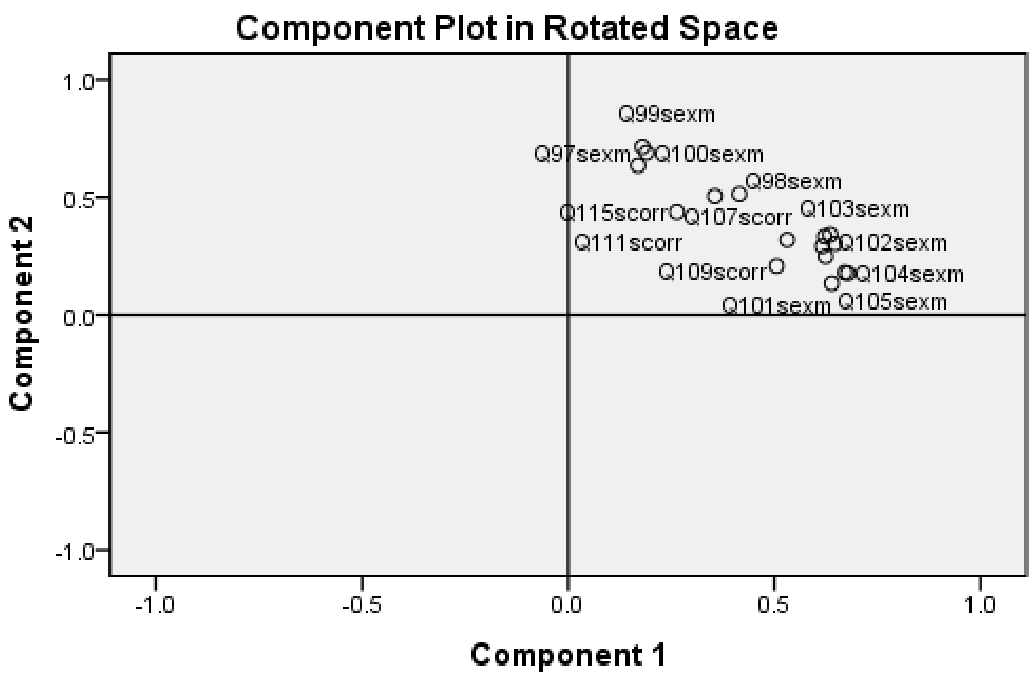
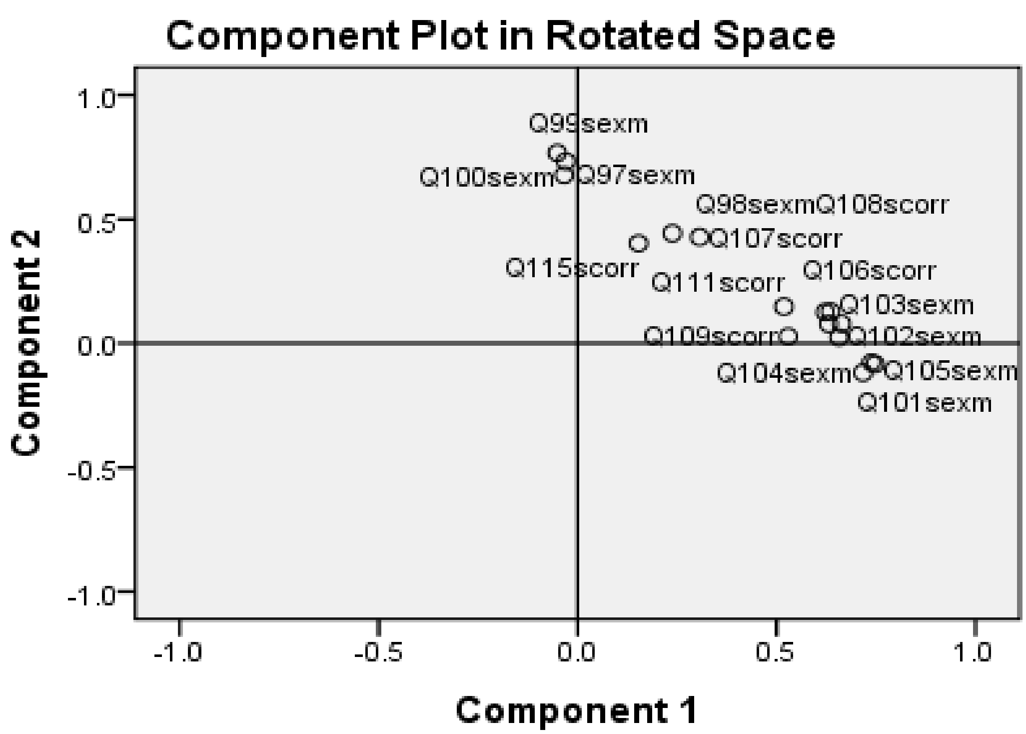


Figure 4.6. Scatter Plot for CS 6 Self-regulation (Oblimin Rotation)



Appendix O 9

Table 4.14

SPSS output: Factor Loadings for Exploratory Factor Analysis of CS 6: Self- regulation–A Comparison of Indian Sample (N=877) with Canadian sample (144)*

Item No	Items (Variables)	Component (Varimax)		Oblimin (India)		Oblimin (Canada)		Commu
		1 SE*	2 SC*	1 SE	2 SC	1 SE?	2 SC?	h^{2*}
104	I reflect on my thinking skills to identify how consistently I use my thinking skills in solving a problem.	.677		.745			.726	.491
105	I analyze areas of consistencies and inconsistencies in my thinking.	.673		.742			.968	.484
112	I reflect on my thinking to improve the quality of my judgment.	.646		.663				.507
101	I examine my values, thoughts / beliefs based on reasons and evidence.	.634		.711		.816		.420
103	I review my reasons and reasoning process in coming to a given conclusion.	.630		.625			.454	.519
108	I continually revise and rethink strategies to improve my thinking.	.625		.630			.676	.495
102	I continuously assess my targets and work towards achieving them.	.619		.648		.505		.447
106	I willingly revise my work to correct my opinions and beliefs.	.617		.634			.765	.461
111	I regularly reflect and critique on my own thoughts.	.530		.519				.380
109	I can participate effectively in discussions with an interdisciplinary team.	.505		.531			.515	.294
99	I review sources of information to ensure important information is not overlooked.		.712		.759	.711		.542
100	I examine and consider ideas and viewpoints even when others do not agree.		.693		.739	.607		.514
97	I reflect on my opinions and reasons to ensure my premises are correct.		.638		.682	.880		.435
107	I willingly modify my position or opinions if need be.				.508		.581	.430
98	I recognize my deficiencies and try to improve my knowledge.		.		503	.897		.375
115	Being aware of the social norms that restrict my thinking, I control their influence on my thinking.	---	----	--	--	.532		.252
	Percent of Variance: 44.301%; Reliability (Core scale) α= 0.872			0.860	0.606	0.833	0.865	

*Factor 1 *SE: Self-Examination 10 items; Factor 2 *SC Self-Correction 5 items; and 0 loading 1 item. Loading weights were better with oblimin rotation although two more items loaded into subscale self correction. The average loading weight for varimax F1 is 0.615 as against oblimin 0.645, (10 items) and average loading weights in F2 for five items is 0.638 (Oblimin) as against 0.681(varimax) for three items. Hence the oblimin solution provided a more accurate measure of variables to construct. The decision was to use the oblimin solution for interpretation and reliability analysis.

Appendix O 10

Table 4.16

Internal Consistency - Cronbach's Reliability Values: A Comparison of Pre and Post EFA

Scale #	Scale & Sub scales	Total # of items / items retained	India (N = 884)		Canada (N =144)		
			Before EFA	After EFA	Before EFA	After EFA	
CS 1.	**INTERPRETATION**	**(17/ 20)**	**0. 838**	**0.826**	**0.895**	**0.877**	
SS 1	Categorization	(7/7)	0.766	0.766	0.826	0.824	
SS 2	Decoding Significance	(4 /6)	0.628	0.612	0.797	0.810	
SS 3	Clarifying Meaning	(6/7)	0.674	0.678	0.829	0.782	
CS 2	**ANALYSIS**	**(13/18)**	**0.860**	**0. 837**	**0.917**	**0. 896**	
SS 4	Examining Ideas	(5/6)	0.720	0. 702	0.811	0.833	
SS 5	Detecting Arguments	(2/6)	0.672	---	0.825	---	
SS 6	Analyzing Arguments	(6+2=8)	0.725	0. 790	0.855	0.852	(2 items loaded from SS 5)
CS 3	**EVALUATION**	**(11/13)**	**0.836**	**0.796**	**0.821**	**0.858**	
SS 7	Assessing Claims	(5+2=7)	0.685	0.746	0.806	0.845	(2 items loaded from SS 8)
SS 8	Assessing Arguments	(3/8)	0.740	0.647	0.707	0.744	
CS 4	**INFERENCE**	**(17/22)**	**0.880**	**0.858**	**0.927**	**0.912**	
SS 9	Querying Evidence	(7/7)	0.741	0. 736	0.822	0.825	
SS 10	. Conjecturing Alternatives	(1/7)	0.755	---	0.844	---	
SS 11	Drawing Conclusions	(8+2=10)	0.754	0. 819	0.845	0.912	(2 items loaded from SS10)
CS 5	**EXPLANATION**	**(20/22)**	**0.904**	**0.876**	**0.931**	**0.932**	
SS 12	Stating Results	(5+4=9)	0.740	0.834	0.790	0.886	(4items loaded from SS11)
SS13	Justifying procedure	(4/9)	0.822	0.713	0.875	0.670	
SS14	Presenting Arguments	(7/8)	0.784	0.715	0.897	0.892	
CS 6	**SELF REGULATION**	**(14/16)**	**0.938**	**0.872**	**0.902**	**0.922**	
SS 15	Self Examination	(9+1=10)	0.812	0.860	0.900	0.883	(I item loaded from SS16)
SS 16	Self Correction	(4/7)	0.836	0.606	0.882	0.865	
	Cronbach α for total scale			0.960		0.97	

Note: Retained all six core scales with 14 subscales.

Appendix P

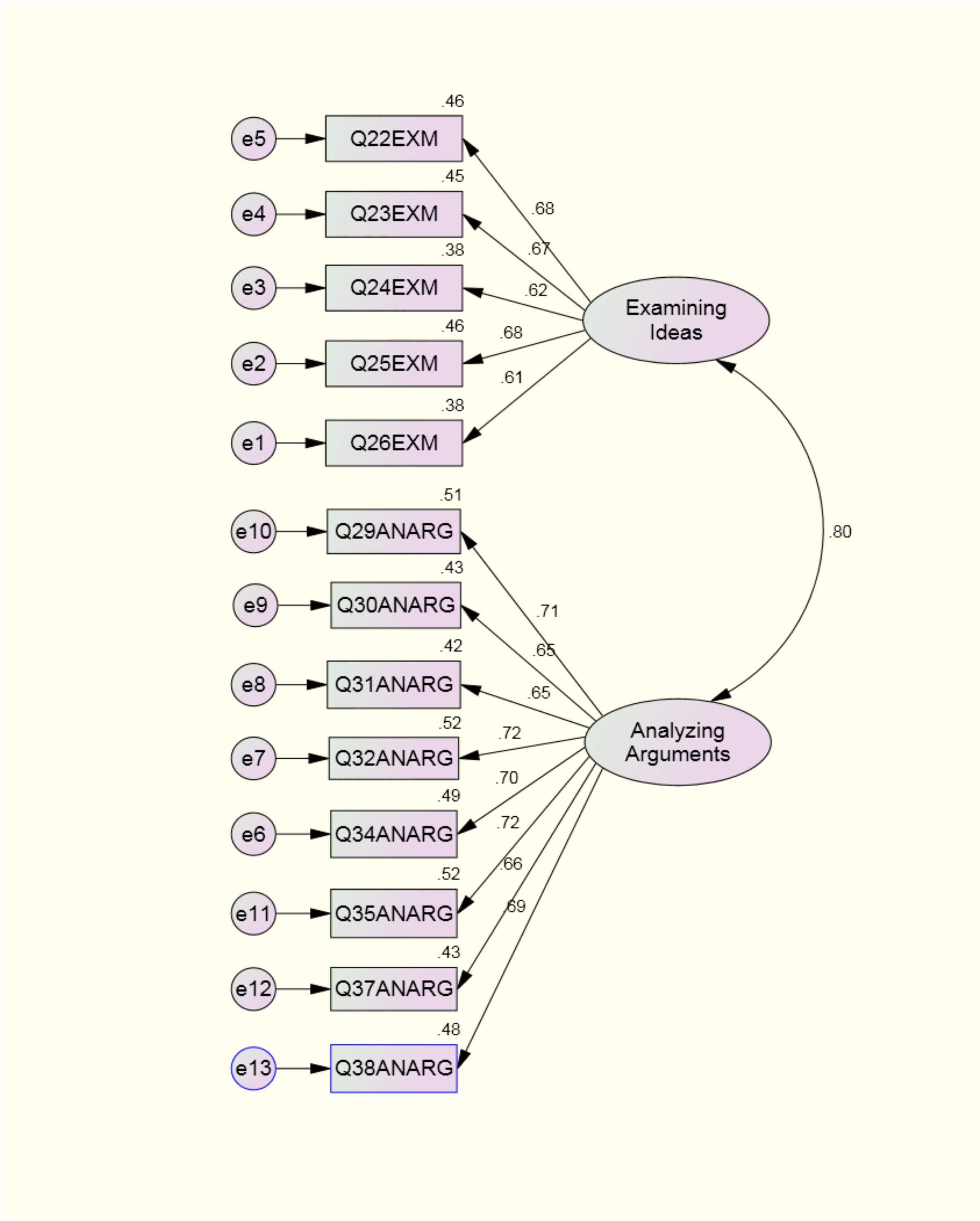
Confirmatory Factor Analysis: Criteria for Assessing Model Fit.

Figure 4.7 Criteria Used for Assessing Model Fit in This Study

Criteria / tests	Acceptable values for decisions on model fit
1. Chi-Square (χ^2)	A non-significant Chi-Square (χ^2) value. A value close to zero indicates little difference between the expected and observed covariance matrices. Usually with large sample size chi-square tends to be significant. Hence, one very rough "rule of thumb" is that a good fitting model is indicated when the ratio of the chi-square to the degrees of freedom is less than 2 (Tabachnck & Fidell, 2007).
2. Comparative Fit Index (CFI).	CFI ranges from 0 to 1 with a larger value indicating model fit. acceptable model fit is indicated by a CFI value of 0.95 or greater (Bentler, 1988; Hu & Bentler, 1999; Tabachnck & Fidell, 2007).
3. Root Mean Square Error of Approximation (RMSEA)	The RMSEA is related to residual in the model. The RMSEA index is used because it allows for "computation of confidence intervals, and takes in to account model parsimony" (Evans, Kirby & Fabrigar, 2003, p.518). The RMSEA values range from 0 to 1 with a smaller RMSEA indicating better model fit. Values of 0.05 or lower constitute a good fit, values of 0.051 to 0.08 constitute an acceptable fit, values of 0.081 to 0.1 constitute a marginal fit and values greater than 0.1 p constitute a poor fit (Hu & Bentler, 1999; MacCallum, Browne, & Sugawara, 1996).
4. CFA requires a model a priori.	The number of factors and the items that load on each factor, a model supported by theory and an error explicitly (Kline, 2005).

Appendix P 1

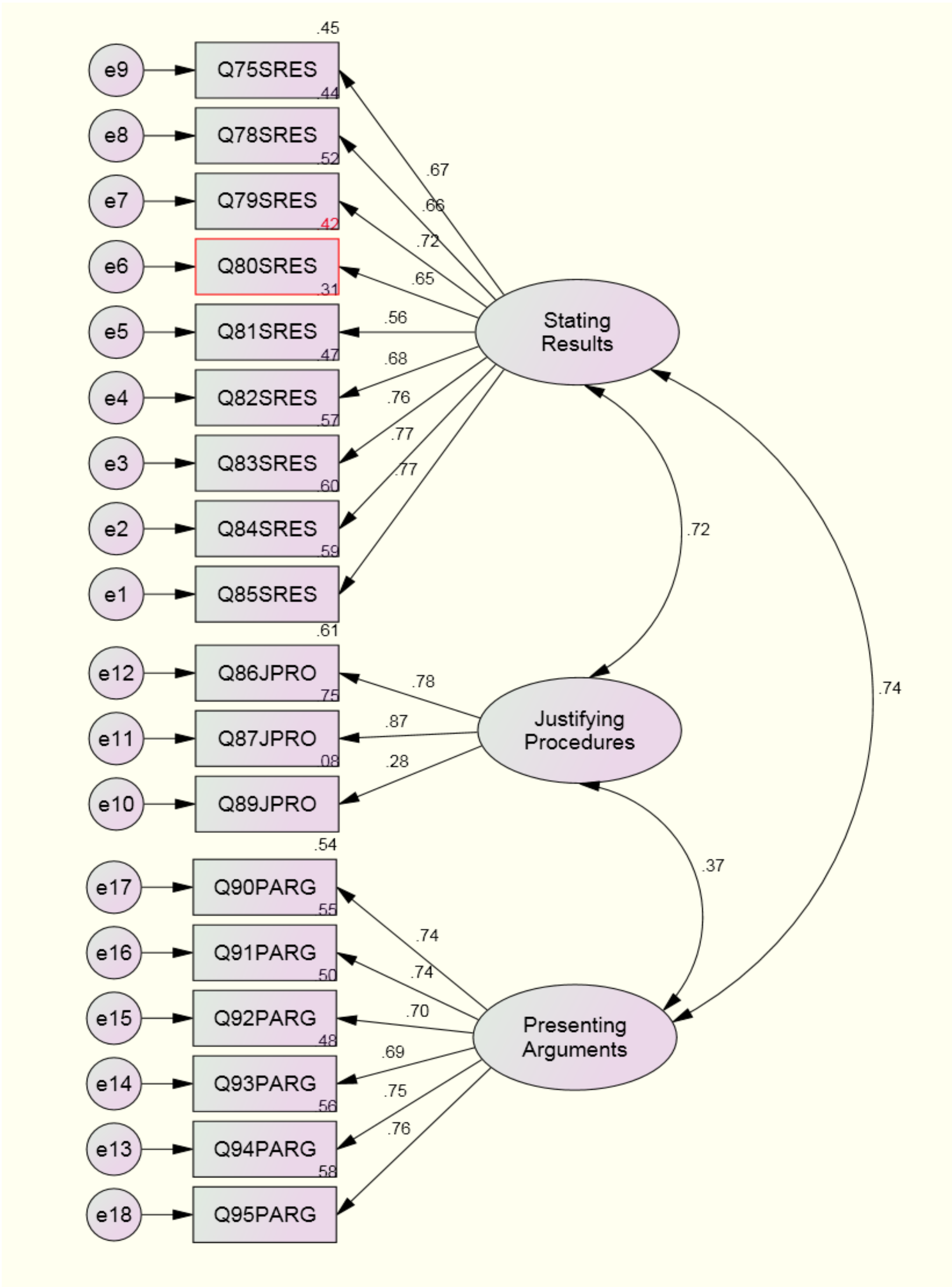
Figure 4.8.4 AMOS Results of Confirmatory Factor Analysis for CS 2 Analysis



Note: SS 5 *Detecting augments* did not emerge as a factor in the EFA

Appendix P 2

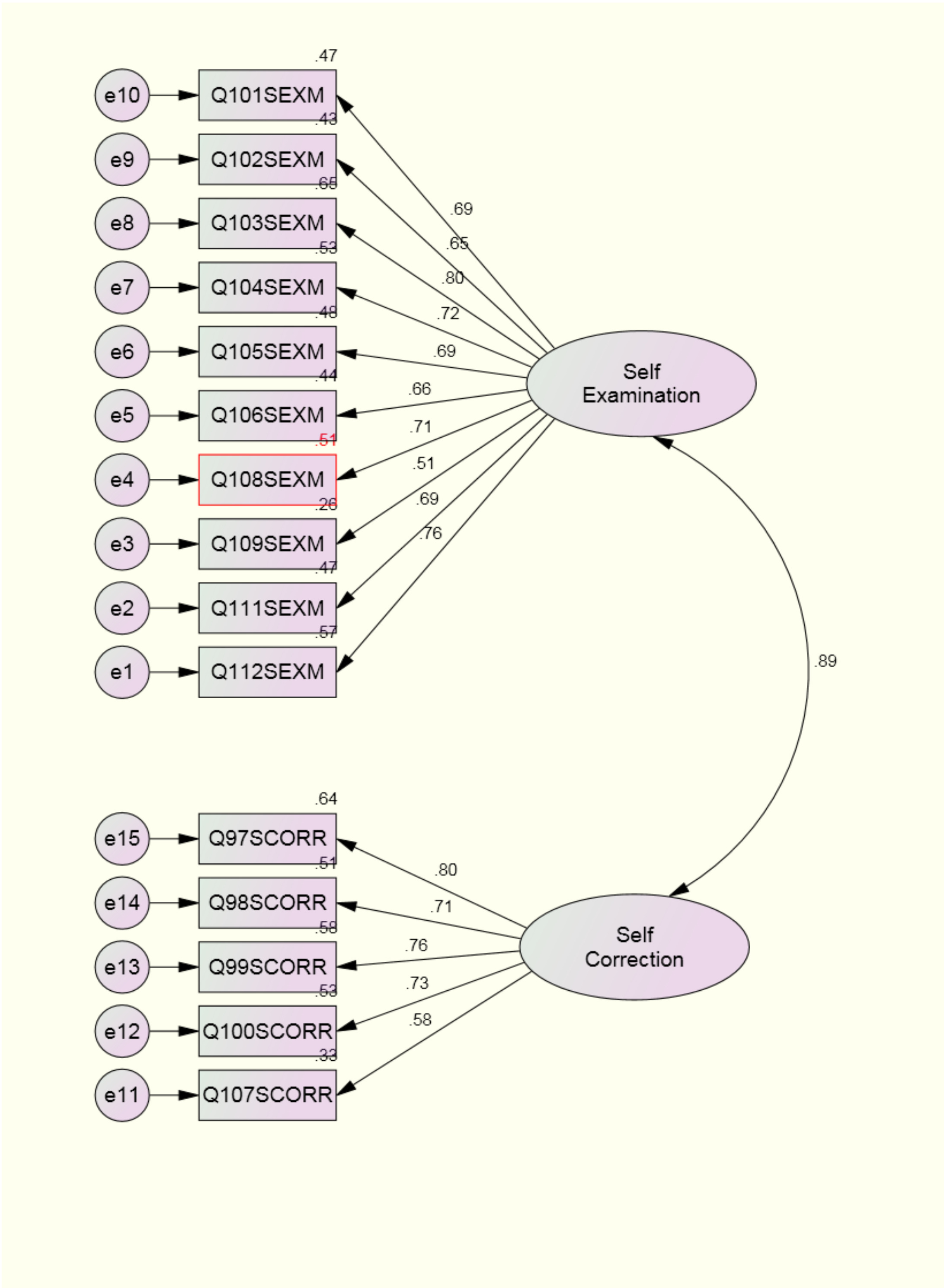
Figure 4.8.5 AMOS Results of Confirmatory Factor Analysis for CS 5: Explanation



Note: Item 89- $R^2 = 0.08$ $(1-R^2)$ $1- 0.08 = 0.98$ high error variance.

Appendix P 3

Figure 4.8.6 AMOS Results of Confirmatory Factor Analysis for CS 6 Self-regulation



Appendix Q

Permission Letter for Using NCS-SF Scale

Date: **Mon, 05 Apr 2010 11:38:26 -0500**
From: "John Cacioppo" <jcaciopp@uchicago.edu> **Block Address**
To: "'ggn417@mail.usask.ca'" <ggn417@mail.usask.ca>

Subject: **NCS**

Reply | Reply All | Forward | Print | Delete

Dear Giriga

You have my permission to use the Need for Cognition Scale in your research. Per your request, I have attached a review which includes the scale in the Appendix.

I wish you the best in your studies.

Sincerely,

John Cacioppo

John T. Cacioppo, Ph.D.

Tiffany & Margaret Blake Distinguished Service Professor,
Director, Center for Cognitive and Social Neuroscience, and

Director, Arete Initiative of the Office of Research and National Laboratories

The University of Chicago

5848 S. University Avenue

Chicago, Illinois 60637

Cacioppo@uchicago.edu

(773) 702-1962 (Phone)

(773) 702-4580 (Fax)

Appendix R

Table 4.18

Comparison of Correlation Coefficient Values (r and rho) for Convergent Validity of the CTSAS with NCS Canada (N=144)

No	Name of scales	Pearson r	P values	Spearman rho	P values
CS 1	Interpretation	.177*	p <.05	.105*	p < .036
1.1	Categorization	.202**	p <.01	.201**	p < .01
1.2	Decoding Significance	.125	NS***	.107	NS
1,3	Clarifying Meaning	.068	NS	.040	NS
CS2	Analysis	.274**	p <.01	.231**	p < .01
2.1	Examining Ideas	.181*	p <.05	.150*	p <.05
2.2	Analysing Arguments	.293**	p <.01	.247**	p <.01
CS3	Evaluation	.224**	p <.01	.199**	p <.01
3.1	Assessing Claim	.155*	p <.05	.137	NS
3.2	Assessing Arguments	.265**	p <.01	.244**	p <.01
CS4	Inference	.285**	p <.01	.273**	p <.01
4.1	Querying Evidence	.305**	p <.01	.286**	p <.01
4.2	Drawing Conclusions	.230**	p <.01	.217**	p <.01
CS5	Explanation	.124	NS	.090	NS
5.1	Stating Results	.090	NS	.070	NS
5.2	Justifying Procedures	.012	NS	.013	NS
5.3	Drawing Conclusions	.185**	p <.01	.137	p < .057
CS6	Self Regulation	.189**	p <.01	.170*	p <.05
6.1	Self Examination	.167*	p <.05	.162*	p <.05
6.2	Self Correction	.187**	p <.01	.157*	p <.05

* significant at $p < .05$;

**significant at $p < .01$

***NS: Non-Significant.

Appendix S

Retained Items in CTSAS: Values of Loading Weights, Cronbach Alpha and Communalities

S. No	Items	SS1	SS3	SS2	h^2	Remarks
CS 1	INTERPRETATION	CAT*	CM*	DS*		Good model fit
1	5. I break the complex ideas into manageable sub-ideas.	.704			.526	
2	2. I sort the information into sub sets.	.641			.536	
3	3. I classify data using a framework.	.711			.512	
4	1. When presented with a problem, first I try to figure out the content of the problem.	.562			.488	
5	4. I break down problem into discrete parts.	.530			.401	
6	6. I categorize similar and related information into groups	.510			.451	
7	7. I classify whole information into specifics.	.532			.424	
8	20. I figure out the meaning of another's point of view.		.750		.533	
9	21. I seek clarification of the meanings of another's opinion or points of view.		.642		.567	
10	17. I clarify my thoughts by explaining to someone else.		.551		.624	
11	15. I restate another person's statements to clarify the meaning.		.529		.578	
12	10. I try to identify the social importance of the information presented in the texts			.633	.496	
13	14. I identify the author's views and intentions in the issue presented.			.729	.501	
14	13. I examine the values rooted in the information presented.			.543	.435	
15	11. I look for various relationships among concepts to understand the meaning.			.518	.383	
16	16. I figure out an example which explains the concept /opinion.			.481	.410	
17	1. I look for analogies of the words and concepts to clarify meaning.			.486	.403	
	Reliability Cronbach α	0.766	0.612	0.678		

*Cat: Categorization; *CM: Clarifying Meaning; *DS: Decoding Significance

CS 2	ANALYSIS	SS 6 AA*	SS 4 EI*	SS 5 DA*	h^2	Lost one SS 5 Poor fit model
18	38. I figure out the assumptions implicit in the author's reasoning.	.658			.469	
19	37. I figure out the process of reasoning for an argument.	.638			.475	
20	31. I figure out if author's arguments include both for and against the claim.	.629			.445	
21	35. I figure out unstated assumptions in one's reasoning for a claim.	.544			.411	
22	30. I ask questions in order to seek evidence to support or refute the author's claim.	.543			.409	
23	36. I look for the overall structure of the argument.	.528			432	
24	29. I figure out what my reasons are for supporting or not supporting the author's claim.	.497			.321	
25	34. Given a brief argument, I identify the author's chief claim.	.450			.389	
26	25. I look for relevant information to answer the question at issue.		.678		.512	
27	22. I examine the similarities and differences among the opinions posed for a given problem.		.609		.429	
28	23. I examine the interrelationships among concepts or opinions posed.		.597		.456	
29	26. I examine the proposals for solving a given problem.		.597		.420	
30	24. I look for supporting reasons when examining opinions.		.573		.390	
	Reliability Cronbach α	0.790	0.702			

*AA: Analyzing Arguments
*EI: Examining Ideas
*DA: Detecting Arguments (SS Failed to emerge in EFA)

CS 3	EVALUATION	SS 7 AC*	SS 8 AA*	h^{2*}	Good model fit
31	43. I assess the contextual relevance of an opinion or claim posed.	.615		.410	
32	47. I examine the logical strength of the underlying reason in an argument.	.599		.446	
33	45. I assess the chances of success or failure in using a premise to conclude an argument.	.635		.356	
34	51. I examine the logical reasoning of an objection to a claim.	.533		.451	
35	46. I rarely examine the flaws in an argument.	.594		.295	-16 in CFA
36	40. I figure out the relevant factors to assess credibility of the opinion/belief raised.	.506		.361	
37	44. I seek the accuracy of the evidence supporting a given judgment.	.454		.400	
38	49. I search for additional information that might support or weaken an argument.		.782	.619	
39	48. I search for new data to confirm or refute a given claim		753	.561	
40	50. I ask questions when the assumptions supporting the arguments are false.		.624	.398	
	Reliability Cronbach α	0. 746	0.647		

*AC: Assessing claim
*AA: Assessing arguments
*h^2: Communalities

CS 4	INFERENCE	SS 11 DC*	SS 10 CA*	SS 9 QE*	h^{2*}	Good model fit Lost SS 10
41	66. I identify the consequences of various options to solving a problem.	.542			.469	
42	64. I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.	.679			.479	Not reached fit indices
43	61. Given a problem to solve, I develop a set of options for solving the problem.	.578			.400	
44	62. Whether or not one agrees, I state the difficulties and the benefits of adopting a given a set of priorities for decision making.	.523			416	
45	63. I systematically analyse the problem using multiple sources of information to draw inferences.	.604			.434	
46	55. I seek useful information to refute an argument when supported by unsure reasons.			.789	.562	
47	54. I seek relevant information to support another's point of view.			.663	.424	
48	56. I collect evidence supporting the availability of information to back up opinions.			.620	.457	
49	57. I seek for evidence / information before accepting a solution.			.504	.405	
50	58. I consider opposing views in support of information when controversial issues are examined.			.538	.358	
51	59. I cannot accept a conclusion without understanding the supporting evidence.			.452	.471	
52	53. When developing a persuasive argument, I search for useful information to support my point of view.			.471	.350	
53	71. I analyze my thinking before jumping to conclusions.	.753			.548	
54	69. I arrive at conclusions that are supported with strong evidence.	.620			.542	
55	70. I use both deductive and inductive reasoning to interpret information.	.559			.475	
56	73. I figure out the logical relationship of the reasons supporting the conclusions.	.493			.431	
57	74. I figure out the pros and cons of a solution before accepting it.	.528			.393	
	Reliability Cronbach α	0.819		0.736		

*DC: Drawing Conclusions
*CA: Conjecturing Alternatives: (SS Failed to emerge in EFA)
*QE: Querying Evidence.
h^2 : Communalities

CS 5	EXPLANATION	SS 12 SR*	SS 13 JP*	SS 14 PA*	h^2	Comments
58	95. I explain the concepts supported with reasons or examples.			.474	.479	Poor fit model
59	78. I can logically present results to address a given problem.	.539			.464	
60	75. I can describe the results of a problem using inferential evidence.	.533			.413	
61	80. I state my choice of using a particular method to solve the problem.	.758			.603	
62	81. I keep a log of the steps followed in working through a problem.	.689			.559	
63	82. I can explain a key concept to clarify my thinking.	.664			.500	
64	83. I state the criteria when evaluating a piece of literature.	.515			.452	
65	85. I report the strategy used in deriving a decision with reasons.	.502			.444	
66	84. I explain the assumptions of using a particular method.	.500			.422	
67	79. I clearly present the inferences to address a given problem.	.519			.426	
68	90. I anticipate reasonable criticisms one might raise against one's view points.			.658	.567	
69	91. I respond to reasonable criticisms one might raise against one's view points.			.641	.528	
70	92. I clearly articulate evidence for my own view points.			.634	.455	
71	93. I present more evidence or counter evidence for another's points of view.			.581	.399	
72	94. I provide reasons for rejecting another's claim.			.700	.480	
73	87. I provide written record of the process involved in drawing inferences.		.754		.675	
74	86. I make a flow chart to show the process of deriving the conclusion.		.717		.637	
75	89. I write essays with adequate arguments with reasons for a given policy or situation.		.649		.565	
	Reliability Cronbach α	0.834	0.713	.715		

*SR: Stating Results;
*JP: Justifying Procedures;
*PA: Presenting Arguments

CS 6	SELF-REGULATION	SS 15 SE*	SS16 SC*		h^2	Comments
76	104. I reflect on my thinking skills to identify how consistently I use my thinking skills in solving a problem.	.745			.491	Not reached fit indices
77	105.I analyze areas of consistencies and inconsistencies in my thinking.	.742			.484	
78	112. I reflect on my thinking to improve the quality of my judgment.	.663			.507	
79	101. I examine my values, thoughts / beliefs based on reasons and evidence.	.711			.420	
80	103. I review my reasons and reasoning process in coming to a given conclusion.	.625			.519	
81	108. I continually revise and rethink strategies to improve my thinking.	.630			.495	
82	102. I continuously assess my targets and work towards achieving them.	.648			.447	
83	106. I willingly revise my work to correct my opinions and beliefs.	.634			.461	
84	111. I regularly reflect and critique on my own thoughts.	.519			.380	
85	109. I can participate effectively in discussions with an interdisciplinary team.	.531			.294	
86	99. I review sources of information to ensure important information is not overlooked.		.759		.542	
87	100. I examine and consider ideas and viewpoints even when others do not agree.		.739		.514	
88	97. I reflect on my opinions and reasons to ensure my premises are correct.		.682		.435	
89	107. I willingly modify my position or opinions if need be.		.508		.430	
90	98. I recognize my deficiencies and try to improve my knowledge.		503		.375	
	Reliability Cronbach α	0.860	0.606			

*SE: Self-examination
*SC: Self-correction

Appendix T

CRITICAL THINKING SELF-ASSESSMENT SCALE

(CTSAS)

GIRIJA NAIR
PhD (Candidate)

Supervisor
Dr. Lynnette Leeseberg Stamler

Please wait for the instruction to begin

College of Nursing

College of Graduate Studies & Research

University of Saskatchewan

Canada

CTSAS is a self-assessment scale with items indicating one's ability to perform critical thinking skills in daily life including professional learning.

Respond to each statement in terms of how frequently you perform these skills.

CTSAS

Directions:

Use a No.2 soft lead pencil only. Do not use a pen or marker.

Complete your back ground information on one side of the CTSAS scan sheet.

Bubble in and write your identification number.

Indicate how frequently you perform with each of the 115 numbered cognitive skill statements by filling the appropriate place on the CTSAS scan sheet.

Here are two examples:

Category response description:

'0' Never; '1'Rarely, '2' Occasionally; '3' Usually; '4' Often; '5' Frequently; '6' Always

	0	1	2	3	4	5	6
Example A: I analyse the situation before making a decision.	○	○	○	○	○	○	●
E.g. A	Never						Always
Example B: I often seek solutions from others.	○	●	○	○	○	○	○
E. g. B							

The location of the ● response to E.g. A shows someone who performs the CT skill **always** with example A.

The location of the ● response to E.g. B shows someone who performs the CT skill **rarely** with example B.

If you erase a response, be sure the erasure is clean

Section A

CRITICAL THINKING SELF-ASSESSMENT SCALE

1. When presented with a problem, first I try to figure out the content of the problem.
2. I sort the information into sub sets.
3. I classify data using a framework.
4. I break down problem into discrete parts.
5. I break the complex ideas into manageable sub-ideas.
6. I categorize similar and related information into groups
7. I classify whole information into specifics.
8. I figure out a person’s purpose in asking a question.
9. I clarify the meaning of an individual’s gesture in a given situation.
10. I observe the facial expression people use in a given situation.
11. I try to identify the social importance of the information presented in the texts.
12. I look for various relationships among concepts to understand the meaning.
13. I examine the values rooted in the information presented.
14. I identify the author’s views and intentions in the issue presented.
15. I restate another person’s statements to clarify the meaning.
16. I figure out an example which explains the concept /opinion.
17. I clarify my thoughts by explaining to someone else.
18. I try to differentiate between opinions and ideas to remove ambiguity.
19. I look for analogies of the words and concepts to clarify meaning.
20. I figure out the meaning of another’s point of view.
21. I seek clarification of the meanings of another’s opinion or points of view.
22. I examine the similarities and differences among the opinions posed for a given problem.
23. I examine the interrelationships among concepts or opinions posed.
24. I look for supporting reasons when examining opinions.
25. I look for relevant information to answer the question at issue.
26. I examine the proposals for solving a given problem.
27. Given a situation, I look for specifics to analyze its various aspects.
28. Given a paragraph to read: I determine the main claim.

29. I figure out what my reasons are for supporting or not supporting the author's claim.
30. I ask questions in order to seek evidence to support or refute the author's claim.
31. I figure out if author's arguments include both for and against the claim.
32. I seek supporting reasons when a person is advancing a claim.
33. My arguments will depend on the sources of supporting information.
34. Given a brief argument, I identify the author's chief claim.
35. I figure out unstated assumptions in one's reasoning for a claim.
36. I look for the overall structure of the argument.
37. I figure out the process of reasoning for an argument.
38. I figure out the assumptions implicit in the author's reasoning.
39. I look for evidence for the reasons supporting the arguments.
40. I figure out the relevant factors to assess credibility of the opinion / belief raised.
41. I assess the credible authority of the source of information supporting the claim.
42. I figure out if a given claim is true or false based on the supported knowledge.
43. I assess the contextual relevance of an opinion or claim posed.
44. I seek the accuracy of the evidence supporting a given judgment.
45. I assess the chances of success or failure in using a premise to conclude an argument.
46. I rarely examine the flaws in an argument.
47. I examine the logical strength of the underlying reason in an argument.
48. I search for new data to confirm or refute a given claim
49. I search for additional information that might support or weaken an argument.
50. I ask questions when the assumptions supporting the arguments are false.
51. I examine the logical reasoning of an objection to a claim.
52. I look for conclusions which are logical.
53. When developing a persuasive argument, I search for useful information to support my point of view.
54. I seek relevant information to support another's point of view.
55. I seek useful information to refute an argument when supported by unsure reasons.
56. I collect evidence supporting the availability of information to back up opinions.
57. I seek for evidence / information before accepting a solution.
58. I consider opposing views in support of information when controversial issues are examined.

59. I cannot accept a conclusion without understanding the supporting evidence.
60. I figure out alternate hypotheses / questions, when I need to solve a problem.
61. Given a problem to solve, I develop a set of options for solving the problem.
62. Whether or not one agrees, I state the difficulties and the benefits of adopting a given a set of priorities for decision making.
63. I systematically analyse the problem using multiple sources of information to draw inferences.
64. I figure out the merits and demerits of a solution while prioritizing from alternatives for making decisions.
65. I talk to others to get feedback on various ways of constructing alternate hypotheses.
66. I identify the consequences of various options to solving a problem.
67. I depend on statistical techniques for drawing inferences.
68. I gather multiple source of information when conclusions are made from opposing views.
69. I arrive at conclusions that are supported with strong evidence.
70. I use both deductive and inductive reasoning to interpret information.
71. I analyse my thinking before jumping to conclusions.
72. I confidently reject an alternative solution when it lacks evidence.
73. I figure out the logical relationship of the reasons supporting the conclusions.
74. I figure out the pros and cons of a solution before accepting it.
75. I can describe the results of a problem using inferential evidence.
76. I can describe the history of a given problem when stating the results
77. I state the reason for holding a particular view.
78. I can logically present results to address a given problem.
79. I clearly present the inferences to address a given problem.
80. I state my choice of using a particular method to solve the problem.
81. I keep a log of the steps followed in working through a problem.
82. I can explain a key concept to clarify my thinking.
83. I state the criteria when evaluating a piece of literature.
84. I explain the assumptions of using a particular method.
85. I report the strategy used in deriving a decision with reasons.
86. I make a flow chart to show the process of deriving the conclusion.
87. I provide written record of the process involved in drawing inferences.

88. I present the evidence to support my conclusion.
89. I write essays with adequate arguments supported with reasons for a given policy or situation.
90. I anticipate reasonable criticisms one might raise against one's view points.
91. I respond to reasonable criticisms one might raise against one's view points.
92. I clearly articulate evidence for my own view points.
93. I present more evidence or counter evidence for another's points of view.
94. I provide reasons for rejecting another's claim.
95. I explain the concepts supported with reasons or examples.
96. I clearly articulate the reasons for accepting a claim.
97. I reflect on my opinions and reasons to ensure my premises are correct.
98. I recognize my deficiencies and try to improve my knowledge.
99. I review sources of information to ensure important information is not overlooked.
100. I examine and consider ideas and viewpoints even when others do not agree.
101. I examine my values, thoughts / beliefs based on reasons and evidence.
102. I continuously assess my targets and work towards achieving them.
103. I review my reasons and reasoning process in coming to a given conclusion.
104. I reflect on my thinking skills to identify how consistently I use my thinking skills in solving a problem.
105. I analyze areas of consistencies and inconsistencies in my thinking.
106. I willingly revise my work to correct my opinions and beliefs.
107. I willingly modify my position or opinions if need be.
108. I continually revise and rethink strategies to improve my thinking.
109. I can participate effectively in discussions with an interdisciplinary team.
110. I respect others' points of view even if they contradict mine.
111. I regularly reflect and critique on my own thoughts.
112. I reflect on my thinking to improve the quality of my judgment.
113. I am aware of my strengths and weakness.
114. I am aware of my values and beliefs and control its undue influence of on my thinking.
115. Being aware of the social norms that restrict my thinking, I control their influence on my thinking.

Section B

NEED FOR COGNITION SCALE

The statements numbered 7.1 to 7.18 show an individual's '*need for cognition*' (enjoy using critical thinking skills). Indicate to what extent the following statements relate to you on a '0 to 4' scale in the CTSAS scan sheet.

0 - Extremely unlike me, 1 - Unlike me, 2 - neutral, 3 - Like me, and 4 - Extremely like me.

116 I would prefer complex to simple problems.

117 I like to have the responsibility of handling a situation that requires a lot of thinking.

118 Thinking is not my idea of fun.

119 I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

120 I try to anticipate and avoid situations where there is likely chance I will have to think in depth about something.

121 I find satisfaction in deliberating hard and for long hours.

122 I only think as hard as I have to.

123 I prefer to think about small, daily projects to long term one.

124 I like tasks that require little thought once I've learned them.

125 The idea of relying on thought to make my way to the top appeals to me.

126 I really enjoy a task that coming up with new solutions to problems.

127 Learning new ways to think doesn't excite me very much.

128 I prefer my life to be filled with puzzles that I must solve.

129 The notion of thinking abstractly is appealing to me.

130 I would prefer a task that is intellectual, difficult and important to one that is somewhat important but does not require much thought.

131 I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

132 It's enough for me that something gets the job done; I don't care how or why it works.

133 I usually end up deliberating about issues even when they do not affect me personally.

Appendix T 1

Information Letter for Student Participants

Dear Participants,

I am inviting you to participate in a research study on "Developing a Self-Assessment Scale and Evaluating its Preliminary Psychometrics for Measuring Critical Thinking Skills of Undergraduate Nursing Students". I am conducting this research as a requirement towards the partial fulfillment of the requirement of PhD Nursing program at the College of Nursing, University of Saskatchewan. The main purpose of the study is to develop the Critical Thinking Self-Assessment Scale (CTSAS) which will enable students to self-monitor and improve their critical thinking skills. Your participation will help me establish a scientific scale which will be beneficial for all students in the future to self-monitor and improve their thinking.

The nature of your participation will involve responding to a self-assessment scale by filling the bubble sheet following the instructions provided. You are also required to furnish some background information which will be used in describing the sample characteristics as a group. It will take approximately 40 to 50 minutes to complete the scale.

The decision to participate in the study is voluntary and this participation is not a regular part of your program of study. Your participation or non participation will not in any way affect your academic status and if you wish, you may feel free to withdraw from the study at any time with no prejudice. Should you at any time decide to withdraw from the study your data will be destroyed. The information that you provide will be kept confidential and will be stored in locked cabinet for five years. The responses will be scored and summated scores will be used for statistical analysis. The results of the study will be reported in aggregate form only. There is no foreseen risk in participating in this study,

If you wish you may place your name on the small piece of paper to be entered into a draw for a $50.00 gift certificate from the university book store, after the draw the names will be destroyed.

If you decide to participate in the study, we ask you to sign this consent form and return with the questionnaire and response sheet.

Thank you for considering participating in this study. Your response will definitely be valuable in establishing the validity of the self-assessment scale. If you have any questions or concerns please do not hesitate to contact either Girija Nair, e-mail: ggn417@mail.usask.ca or the study supervisor Dr. Lynnette Leeseberg Stamler, 966 1477 e-mail: lynnette.stamler@usask.ca

The study has been approved by the Behavioural Research Ethics Board at the University of Saskatchewan on (date) and the College of Nursing Research Committee. If you have any questions about your rights as a participant or concerns about the research project you may contact Research Ethics Office at the University of Saskatchewan at 1 (306) 966-2084.

____

Girija Nair, RN, RM, PhD Candidate
College of Nursing, University of Saskatchewan

____

Dr. Lynnette Leeseberg Stamler
Assistant Dean, Graduate Studies
and Continuing Nursing Education.
Tel: (306)966 1477, lynnette.stamler@usask.ca

Appendix T 1
Consent Form

Permission to Participate

I have read and understand the information provided to me about this research study titled "Developing a Self-Assessment Scale and Evaluating its Preliminary Psychometrics for Measuring Critical Thinking Skills of Undergraduate Nursing Students". I understand that as part of this research study I will be required to provide responses to a set of items in a scale and fill in the bubble sheet with my responses. The self-assessment scale should take about 40 to 50minutes to complete. I also understand that the researcher will be repeating this test after two weeks. I was given enough time to think about it and decide on my participation.

All the information provided in the scale answer sheet will be kept confidential and will not be shared by anyone outside the research team. Although the information from this research study will be used in the researcher's thesis, for publications, and presentations at conferences, all information will be in group form so that it is not possible to reveal the individual identity. Also, the consent form and the bubble sheets where my identity may appear will be kept under safe custody where access is denied to anyone outside the research team. I understand that I can withdraw from the study any time if I wish to do so without affecting my academic status. There is no cost involved to the participant except the time spent on the study in filling the responses. I understand that by signing this consent I do not waive any of my legal rights. There are no foreseen risks in responding to the scale. I also understand that by participating in this study I am not obligated to participate in any future research.

If I wish to be part of the retest, and /or be a part of the draw for a voucher for the bookstore, I have signed the appropriate forms.

The researcher respects my decision to participate or not. Should I choose to withdraw after the study has began (for the repeat testing), I may ask that all the information I have provided be deleted and destroyed, and my request will be respected and the information will be destroyed.

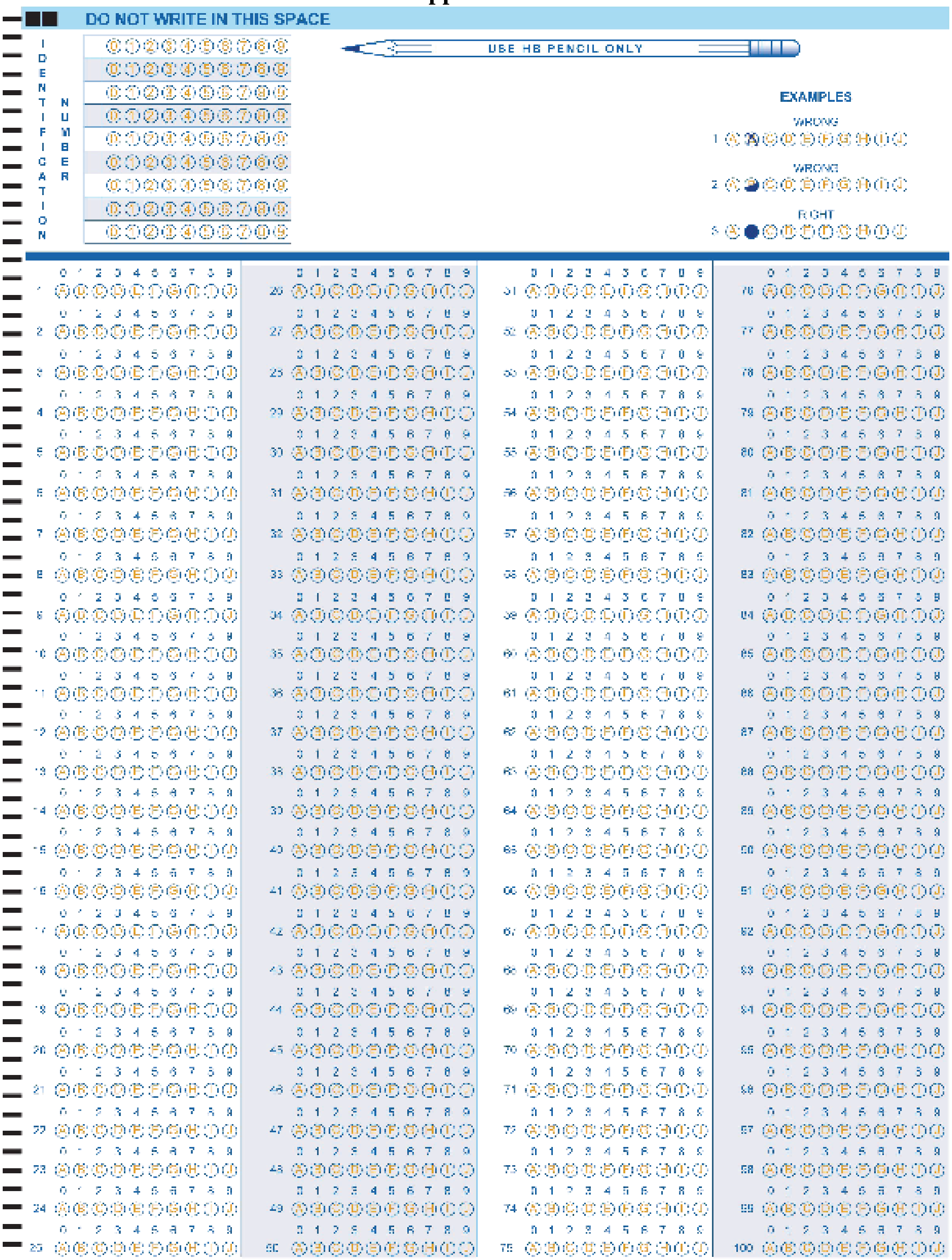
I agree to participate in the study and have all the information that I provide to be used for the research study. I have kept one copy of the consent for my record.

__________________________ _______________________
Signature of the Participant Date

☐ As part of the statistical testing, I agree to complete this package again in two weeks time.
☐ I wish my name to be part of a draw for a gift voucher.
Name______________________________________

Note: Not all students will be required for the second take of the testing.

Appendix T 2


DO NOT WRITE IN THIS SPACE
IDENTIFICATION NUMBER
USE HB PENCIL ONLY
EXAMPLES
WRONG
WRONG
RIGHT

Appendix T 3

Demographic Form for Large Scale Validity Testing

Please answer the following questions:

Name of School__________________________________ Country_______________

Year of Nursing Program _____ First Year

_____ Second Year

_____ Third Year

_____ Fourth Year

Gender _____Male

_____Female

Age _____ Less than 17

_____ 18 – 20

_____ 21 – 24

_____ 25 – 28

_____ 29 or more