USING DIF TO EXAMINE MEASUREMENT INVARIANCE OF SCHOOL CONNECTEDNESS AND SCHOOL SUPPORTIVENESS

by

Gulam A. Khan

A thesis submitted in conformity with the requirements for the degree of Master of Education
Department of Learning, Higher and Adult Education
Ontario Institute for Studies in Education, University of Toronto

© Copyright by Gulam A. Khan 2014
USING DIF TO EXAMINE MEASUREMENT INVARIANCE OF SCHOOL CONNECTEDNESS AND SCHOOL SUPPORTIVENESS

Master of Education, 2014

Gulam A. Khan

Department of Leadership, Higher and Adult Education

Ontario Institute for Studies in Education, University of Toronto

Abstract

One of the most important aspect of an assessment is its ability to reliably assess change without measurement invariance. This study examined the consistency of school connectedness and school supportiveness as measured by a set of sub-items on the COMPASS Student Questionnaire (CSQ) between students in Grade 9 and Grade 12. While exploratory factor analyses (EFA) revealed that the dimensionality of these data remained consistent between Grade 9 and Grade 12 students, differential item functioning (DIF) analyses suggested evidence of measurement invariance. DIF analyses revealed that students in Grade 12 who attended a different school last year more readily endorsed sub-items pertaining to school connectedness with “strongly agree” and “agree” compared to Grade 9 students. Suggestions for future directions with the CSQ are provided.
Acknowledgements

I would like to express my sincerest appreciation to my thesis supervisor, Dr. Ruth Childs for her patience and mentoring, ever since I was considering applying to the M.Ed. program, until the completion of this degree. Dr. Ruth Childs' down-to-earth humility and genuine willingness to help is matched only by her intellectual fortitude. I am thankful for her suggestions and insight that inspired this challenging project, allowing me to increase my knowledge in psychometrics.

To my caring and supportive family I express my deepest gratitude. I would be remiss to not include the innumerable sacrifices my parents made, affording me with a good education. In particular, I would like to express my love and gratitude to my mother whose hard work ethic and support has inspired me to strive for excellence; and to my father for his undying belief in my ability to achieve.

Lastly, I would like to thank God for blessing me with the patience, strength, and support to realize my goals.
# Table of Contents

CHAPTER ONE: INTRODUCTION ........................................................................................................ 1

CHAPTER TWO: A REVIEW OF THE LITERATURE ........................................................................... 5

School Connectedness and School Supportiveness ......................................................................... 5

School Connectedness ..................................................................................................................... 5

Correlates of School Connectedness and School Supportiveness ...................................................... 9

Schools and Adolescent Cognitive and Emotional Development ..................................................... 11

Factor Analysis .................................................................................................................................. 13

Introduction to Factor Analysis ......................................................................................................... 13

Differential Item Functioning and Differential Step Functioning ....................................................... 18

Procedures for Detecting DIF ............................................................................................................ 21

Differential Step Functioning ............................................................................................................ 21

CHAPTER THREE: METHODS ............................................................................................................. 22

Participants ......................................................................................................................................... 22

The COMPASS Student Questionnaire ............................................................................................. 23

Factor Analyses .................................................................................................................................. 24

Exploratory Factor Analyses ............................................................................................................. 24

Differential Item Functioning and Differential Step Functioning ..................................................... 24

CHAPTER FOUR: RESULTS AND DISCUSSION ............................................................................. 26

Descriptive Statistics ......................................................................................................................... 27

Exploratory Factor Analyses ............................................................................................................... 29

DIF and DSF Analyses ....................................................................................................................... 31

CHAPTER FIVE: CONCLUSION ............................................................................................................ 37

Limitations .......................................................................................................................................... 37

Future Directions ............................................................................................................................... 37

REFERENCES ....................................................................................................................................... 39
CHAPTER ONE: INTRODUCTION

Students’ academic success is a key determinant of future success and obtaining employment. According to Statistics Canada (2014) those who prematurely drop out of school are less likely to participate in the labour force than those that have at least a high-school diploma. Consequently, hundreds of millions of dollars are invested annually to ensure students’ academic achievement. The Ontario Ministry of Education budget for grants for student needs, program and other funding was a total of 200 million dollars in 2013-2014 (Statistics Canada, 2013).

While most efforts to promote academic achievement are aimed toward improving the curriculum or facilitating learning through the use of technology, relatively less attention is paid to the social and relational aspects that have been found to be key determinants of academic success (Connell & Wellborn, 1991; Gest, Welsh, & Domitrovich, 2005). Specifically, Gest, Welsh, and Domitrovich (2005) found that students who develop close meaningful relationships with their peers and teachers often report positive feelings about school and generally perform at higher academic levels. As such, being able to collect data on such social and relational aspects related to academic achievement is important. Surveys and questionnaires are commonly used to describe and summarize observations from a group of individuals (Drummond, 1990).

The Oxford Dictionary defines the word “survey” as “the act of viewing, examining, or inspecting in detail [especially] for some specific purpose” (“Survey”, 2014). Phillips (2000) stated that surveys “are among the most important data collection tools available in evaluation for the assessment of change” (p. 2). These descriptions draw attention to two important ideas regarding surveys: (1) the idea of a “specific purpose” behind the survey and (2) the use of surveys for the “assessment of change.” Specifically, the survey should consist of questions focusing on a definite area of interest (or construct) to observe, otherwise the information collected
from the survey will be of limited utility. Furthermore, if surveys are to be used to make meaningful comparisons over time, it is important that the constructs are invariant over time. Marsh, Marco, and Abcy (2002) wrote that “the researcher should investigate whether the underlying factors, the relations between items and factors, the relations among the factors, and measurement error are comparable in the different…groups” (p. 259). It is therefore important to examine the construct validity and measurement invariance over time for researchers to make accurate interpretations of the results obtained from surveys.

When administering a questionnaire for the purpose of tracking changes over time or to make comparisons among different groups of interest, it is important to ensure that the construct remains stable over time (Pitts, West, & Tein, 1996). In other words, the probability of obtaining a particular measure should theoretically be independent of group membership or time effects. Statistically speaking, the definition of measurement invariance is:

The observed random variable $Y$ is said to be measurement invariance with respect to selection on $G$, if $F(y|\eta, g) = F(y|\eta)$ for all $(y, \eta, g)$ in the sample space, where $Y$ denotes an observed random variable with realization $y$; $H$ denotes the latent variable (i.e., factor) with realization $\eta$ that is measured by $Y$, or underlies $Y$; $G$ denotes a random variable with realization $g$ that functions as a selection of a subpopulation with realization $g$ that functions from the parent population by application of a selection function $s(g)$, $0 \leq s(g) \leq 1$ (Wu, Li, & Zumbo, 2007, p. 2).

Demonstrating measurement invariance across groups, across conditions, and over time is central to establishing evidence for construct validity (Yin & Fan, 2003) especially if these fac-
tors are to be used to make inferences about specific groups of individuals regarding a particular construct or set of constructs.

The definition of measurement invariance proposed by Wu, Li, and Zumbo (2007) fits well with the framework of factor analysis. Specifically, a score on a particular latent variable (or factor such as math ability or empathy) can represent an individual’s true score. In this paper, “factor” and “latent variable” are used interchangeably. Factor analysis is a statistical method that is commonly employed not only to develop theoretical constructs, but to establish their validity. Factor analysis can also be employed to assess measurement invariance. Factor analysis will be described in greater detail in the following chapter.

Differential item functioning (DIF) analysis can also be useful for investigating measurement invariance. DIF is a statistical method that identifies items on a test or questionnaire that may be differentially “difficult” for two groups of respondents (e.g., language or gender).

For those items with more than two possible responses (such as those with a Likert-type scale), differential step functioning (DSF) analysis examines whether groups differ at specific score levels.

The purpose of this study to explore the dimensionality and the stability of constructs through examining the measurement invariance of a group of sub-items in a questionnaire assessing high school students’ global health behaviours and attitudes toward and impression of their high school. Exploratory factor analysis (EFA) will first be used to explore the basic factor structure of these data. Second, DIF analyses will be conducted to identify any sub-items that may behave differently for students in Grade 9 compared to students in Grade 12. Lastly, DSF analyses will be conducted on those items exhibiting moderate or strong DIF to investigate ex-
actly at what score level the groups differ. The study will inform the use of these items in the future to measure change in students’ attitudes over time.
CHAPTER TWO: A REVIEW OF THE LITERATURE

This chapter provides a review of the literature pertinent to this study. As the questionnaire in this study was in part developed to investigate school connectedness and school supportiveness, a review of the literature pertaining to these two constructs is provided. An overview of the literature pertaining to correlates of school connectedness and school supportiveness is also included. Next, a brief discussion of the developmental psychology literature relevant to the respondents in this study is presented. The chapter concludes with an overview of exploratory factor analysis (EFA) and differential item functioning, the two main statistical techniques employed in this study.

School Connectedness and School Supportiveness

*School Connectedness*

Students’ connection to school (interchangeably referred to as school connectedness or connectedness in this paper), has been referred to as attachment, membership, bonding, engagement or belonging in a wide range of fields including psychology, health, and education. The construct of school connectedness can be linked to the need to be loved or to belong as part of Maslow’s (1968) hierarchy of needs. According to Libbey (2004), school connectedness can be conceived of as the extent to which a young person feels like they belong at school. Nonetheless, school connectedness is defined and as a result measured in different ways. The Wingspread (2004) declaration on school connections provides the most frequently used construct definition of school connectedness. According to Wingspread (2004), school connectedness is an overarching term encompassed by students’ beliefs regarding the extent to which the adults in the school community are invested in student learning and their interest in students as individuals. Ways in which school connectedness is often represented are through
students’ perceptions of the academic expectations set by their teachers, interactions between students and teachers, and feelings of safety (Wingspread, 2004). It is important to mention that some studies employ a combination of attitudinal and behavioural items to examine the construct of school connectedness where attitudinal items pertain to the perceptions of students’ beliefs of school connectedness and behavioural items pertain to readily observable or quantifiable measures such as school attendance, participation in class, and attention to tasks.

The construct definition of school connectedness is further complicated through the use of multiple terms in its description. Libbey (2004) described 11 distinct terms for school connectedness including school bonding, school climate, school connection, attachment, and orientation to school. With a wide array of definitions for school connectedness, there have been several ways in which researchers have attempted to quantify this construct. While some researchers use self-report, others integrate information from teachers and other school staff to measure school connectedness. This has resulted in questionnaires with as many as 72 items (Karcher & Finn, 2005) and some with as little as 5 items (Resnick et al., 1997). With such a disparity in number of items claiming to measure only slight variations in the commonly used construct definition of school closeness, it complicates the comparison of school connectedness across studies because all of the aforementioned measures are considered to encompass school connectedness. Nonetheless, the most commonly employed measures of school connectedness developed by Resnick and colleagues (1997) are used in studies that typically are a part of a larger questionnaire that assesses global health because of its brevity and good reliability ($\alpha = .75$; Sieving et al., 2001). Moreover, this five-item scale satisfies Wingspread’s (2005) definition of school connectedness.

In examining the literature, a common shortcoming among the studies is that they rarely,
if ever consider the measurement invariance of school connectedness across groups and over time. That is, do these constructs show invariance over time and do they show invariance over specific populations of students? This is a worthwhile question as it allows for meaningful conclusions to be made regarding school connectedness and supportiveness and academic achievement to be made irrespective of grade level. Consequently, psychometricians are searching for evidence that such assessments not only provide valid measures of school connectedness but that they demonstrate measurement invariance over time.

*School Supportiveness*

In lieu of a single universal definition of school supportiveness, the literature presents several construct definitions, terms, and components that encompass school supportiveness. Generally speaking, much of the literature suggests that positive student-teacher relationships and school support networks are the most common and important features of school supportiveness (Finn & Voelkl, 1993; Gest, Welsh, & Domitrovich, 2004; Rosenfeld, Richman, & Bowen, 2000). School warmth has also been found to be associated with school supportiveness (Voelkl, 1995). Specifically, Voelkl (1995) defined school warmth as the degree of teacher warmth, caring, and supportiveness as perceived by the student.

Arthur, Hawkins, Pollard, Catalano, and Baglioni (2002) examined adolescent problem behaviours using the Communities That Care Youth Survey (Hawkins et al., 1992) a self-report survey designed to assess a set of risk and protective factors over several domains including school, family, and peer relationships. In this study, Arthur and colleagues’ (2002) definition of school supportiveness included teacher support, peer support, competence building support in the classroom, and a school environment that provides safety and care.
Rosenfeld, Richman, and Bowen (2000) viewed social support networks within the school as a key component of school supportiveness. Specifically, Rosenfeld, Richman, and Bowen (2000) defined school supportiveness in terms of the extent to which students perceive their parents, friends, and teachers as important sources of social support.

The definition of school supportiveness provided by Klem and Connell (2004) focused particularly on students’ perception of teacher support— that is, the degree to which they perceive their teachers to be invested in their success and education. Other components of the school supportiveness construct included the provision of a personalized, closely-knit, and caring learning environment.

As was the case with school connectedness, quantifying school supportiveness is complicated by the disparity in which component of school supportiveness is emphasized. With some definitions emphasizing the teacher warmth component, and some emphasizing a closely-knit school environment, it is difficult to develop a definitive set of items for a scale that would measure the degree of school supportiveness. Nonetheless, there are several items that are commonly used to assess the level of perceived school supportiveness that usually comprise a global assessment.

To quantify school supportiveness, researchers have used self-report measures in conjunction with teachers’ and parents’ reports of school supportiveness. Some questionnaires such as the Communities that Care Youth Survey are quite lengthy spanning over 12 pages that measure several of the abovementioned domains of school supportiveness. Other questionnaires such as The Hemmingway: Measure of Adolescent Connectedness (Karcher, 2003) which is a 78-item measure that examines 15 distinct subscales of connectedness in adolescents includes a subscale of relatedness with their school and teachers that has been used as a measure for
supportiveness as it arguably measures the same construct. Other measures of school supportiveness have been developed in reference to Battistich’s Middle School Student Questionnaire Measure from the Child Development Project (MSSQM; Battistich, 2000; Battistich, Solomon, Watson, & Schapps, 1997). The MSSQM provides a comprehensive measure of school supportiveness as it takes into account school teacher support, school peer support, competency building in the classroom, and the school environment. Specifically, school teacher support assesses the amount of perceived help, care, and assistance the student receives from the teachers in their school. School peer support measures students’ perceived level of help, care, and dependability from peers in their school. Social competency building assesses the extent to which students perceive they receive assistance with developing life skills in the classroom. Supportive school environment assesses students’ feeling of being safe and cared for by their school.

However, as was the case with school connectedness, the literature on measurement invariance in school supportiveness is limited. It is important to be able to establish measurement invariance especially in situations where one wishes to conduct longitudinal studies using an instrument purported to measure school connectedness and school supportiveness.

Correlates of School Connectedness and School Supportiveness

According to Bronfenbrenner (1979), students are strongly influenced by the social contexts in which they live. Several studies have explored school connectedness and school supportiveness with respect to school outcomes such as academic achievement, school satisfaction, and school engagement (Rosenfeld, Richman, & Bowen, 2000). Findings from other studies have suggested school connectedness and school supportiveness are linked strongly to overall psychosocial well-being in both collectivistic and individualistic cultures (Clark, 1991;
This section will provide an overview of the most commonly reported correlates of school connectedness and school supportiveness.

Studies from the fields of health, psychology, and education have shown that students who show a strong sense of connectedness in general are more engaged in their education and show higher levels of commitment to their school work (Libbey, 2004; Nasir, Lee, Roseberry, & Warren, 2006). Specifically, Hawkins, Guo, Hill, Battin-Pearson, and Abbott (2001) found that higher levels of connectedness between students in Grades 7 and 12 positively correlated with grade point average and negatively in delinquent behaviours in Grade 12. Furthermore, high school connectedness levels were found to be negatively with correlated with remediation, suspension, expulsion, and dropping out. Several other researchers have found similar positive correlations of school connectedness with respect to school-related outcomes (Anderman, 2003; Croninger & Lee, 2001; Klem & Cornell, 2004). These findings have been replicated with students from lower socioeconomic status families (Nasir, Jones, & McLaughlin, 2011).

In the same vein, several studies have indicated that high levels of adolescents' perceived supportiveness from their school teacher, their parents, and their friends is associated with better attendance, more hours of studying, lower rates of problem behaviours, higher school satisfaction, increased engagement in school activities, and higher levels of self-efficacy, and higher grades (Richman & Bowen, 1997, 2000; Rosenfeld, Richman, & Bowen, 2000). Furthermore, Lapan, Wells, Petersen, and McCann (2013) found that schools that provided personalized counselling services increased school connectedness levels and was associated with students’ academic success.

Supportive relationships with adults and peers are a basic psychological need for adolescents (Erickson, 1968). It is unsurprising to find that an important component of both
school connectedness and school supportiveness is student-teacher and peer relationships. Rosenfeld, Richman, and Bowen (2000) found a positive relationship between social support and psychological well-being in adolescents. The lack of supportive relationships has been found to compromise psychosocial well-being and has been implicated in adolescent problem behaviours (Crosnoe, Johnson, & Elder, 2004). Peer support, a major component of school supportiveness has been found to provide social and emotional support and can serve as a buffering agent for problem behaviours (Crosnoe & Elder, 2004). Baker, Dilly, Aupperlee, and Patil (2003) considered peer relationships from an ecological perspective and found that through developing positive peer relationships, students’ positive adjustment with their school environment is facilitated. In regard to school environment, another integral component of school supportiveness, Battistich, Solomon, Watson, and Schapps, (1997) found a decrease in problem behaviours such as bullying and aggression in schools promoting a safe and caring environment. Further, Resnick et al. (1997) found that feelings of being cared about and valued constitute an important protective factor in healthy adolescent development.

With a large body of research pointing to both academic and nonacademic benefits resulting from school connectedness and school supportiveness, it has become common for schools to administer questionnaires that examine these two constructs. It is important to be able to establish their measurement invariance.

Schools and Adolescent Cognitive and Emotional Development

It was previously mentioned that the vast majority of instruments used to assess levels of school connectedness and supportiveness rely on self-report measures and require respondent to provide an accurate representation of their perception of school connectedness and school supportiveness. It is important to consider the abovementioned research in conjunction with the
developmental psychology research on adolescents as it holds notable implications for the
applicability and use of some measures of school connectedness and school supportiveness. This
section of the literature review provides a brief overview of the relevant research pertaining to
the developmental psychology of adolescents.

In regard to cognitive development, it is during the adolescent years that the movement
beyond the limitations of concrete mental operations and the development of the ability to think
in a more abstract manner takes place (Piaget, 1977). Piaget termed this new stage the Formal
Operational Stage which demonstrates how adolescents are able to reflect on abstract and
intangible concepts as opposed to be fixated on what is sensible and concrete. Formal operational
thinking includes proportional thinking and relativistic thinking. Proportional thinking
involves making observations in the absence of visual evidence. That is, relying on unobservable
facts and concepts to make judgments and assertions. Relativistic thinking involves asserting an
opinion based on one’s own bias. Furthermore, at the Formal Operational Stage, adolescents can
consider a wide range of choices rather than demonstrate an all-or-nothing way of thinking
(Piaget, 1977).

Steinberg (2005) examined cognitive and affective development during adolescence and
found that emotion affects basic cognitive processing including decision-making and behavioural
choice. Moreover, patterns of cognitive development are influenced by the emotional and social
context in which reasoning occurs. Research by Klaczynski (1997) found that when faced with a
logical argument, “adolescents are more likely to accept faulty reasoning or shaky evidence
when they agree with the substance of the argument than when they do not” (p.276). Also, the
substance of the argument may be emotionally appealing and invoke memories of certain
experiences which may or may not necessarily be in line with the substance of the argument. To
summarize, both the social and emotional context have been found to influence cognitive performance in adolescents.

The biological, cognitive, social-emotional changes, and developing relationships between parents and peers during the adolescent and teenage years are arguably, next to infancy the most formative experiences of one’s life. Typically, adolescents and teenagers are transitioning into a new school and forging new relationships with teachers and peers. Thus, it is important to consider such influential changes in their life in conjunction with the emotional and cognitive changes experienced by adolescents and teenagers. This becomes increasingly important in accurately interpreting the results obtained from school surveys with adolescents and teenagers.

Factor Analysis

This section introduces the key concepts of factor analysis and describes the procedures used when applying factor analysis in an exploratory manner.

Introduction to Factor Analysis

With a long history dating back to the early 1900s, factor analysis is a statistical approach pioneered by Charles Spearman and Karl Pearson used to discover the principles of the inheritance of manifest characters (Mulaik, 1972). Spearman's (1904) seminal paper outlining factor analysis to objectively measure intelligence and proved central to establishing psychology as a science in a time where psychology was considered an art. Spearman developed the common factor model that outlined, in a mathematical equation the estimation of latent, underlying factors and reads as follows:

\[ \mathbf{x} = \mathbf{\Lambda} \xi + \mathbf{\delta}, \]

where \( \mathbf{x} \) is a vector of observations on \( p \) variables; \( \xi \) is a column vector of \( k \) common factors; \( \mathbf{\delta} \) is a vector of \( p \) residuals, representing the effects of specific factors in addition to random
measurement error; and \( \Lambda \) is a \( p \times k \) matrix of factor loadings relating the observed \( x \)'s to the latent \( \xi \)'s. Two critical assumptions are made in this model: (a) the residuals, \( \delta \), are uncorrelated with each other; and (b) the residuals, \( \delta \), and the common factors, \( \xi \), are uncorrelated.

The theory behind factor analysis was further developed by L.L. Thurstone during the 1930s with the advent of multiple factor analysis, an approach that assumes there are multiple factors in contrast to the previous model that accounted for one general factor. Factor analysis remains central to summarizing the interrelationships among variables. This makes possible the development of new constructs and theories. It is therefore unsurprising that factor analysis is considered as a powerful statistical technique often employed in the behavioural and social sciences.

Broadly speaking, there are two major types of factor analytic techniques, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA, the most common form, is used when the researcher does not have evidence a priori to form a hypothesis about the nature of the underlying construct. While CFA lies beyond the scope of this paper, it is worthwhile noting that in CFA, the researcher has already formulated and is testing a hypothesis about the number of factors underlying the data.

According to Jöreskog and Sörbom (1979), regression analysis and partial correlation are closely related to the theory of factor analysis. Much of the literature on factor analysis is considered from the perspective of the statistical linear model. Factor analysis is conducted with the assumption that a set of underlying factors exist such that when these factors are separated, the intercorrelations between the observed variables become non-existent (Jöreskog & Sörbom, 1979).

In this linear model, the observed variables (dependent variables) are assumed to be a
weighted combination of a set of factors (independent variables). Many may argue that the use of the linear model to represent the relationships between factors and variables may not always give the best representation of the data – sometimes a nonlinear relationship may exist. However, “the only assumption made in using the linear model is that the variables are linear functions of the factors” and “are not assumed to be linearly related” (Gorsuch, 1983, p. 18).

In the common factor model it is hypothesized that a single variable can be broken down into common variance and unique variance. Common variance is variance shared with other variables included in the model, while unique variance is the proportion of variance unique to a particular variable and also includes the error in the model. In factor analysis only the common variance of the observed variables is analyzed. This is one of the distinguishing features between this approach and a similar procedure known as principal components analysis, which analyses the total variance (common plus unique).

**Exploratory Factor Analysis (EFA)**

**Matrix of Association**

Factor analyses are conducted by computing a correlation or covariance matrix using the raw data. The nature of the observed variables (e.g., categorical) is what determines the type of correlation matrix that is most appropriate. By default, most statistical packages such as SPSS generate product-moment correlations.

**Factor Extraction**

Following the computation of the correlation matrix is the extraction of an initial factor solution. This involves replacing the diagonal of the correlation matrix through the use of communality estimates. The communality of a variable refers to the proportion of a variable’s variance that is associated with the common factors underlying a set of variables.
The literature describes several procedures for determining these estimates such as the use of squared multiple correlations, the largest absolute correlation, unities (1s) as initial estimates or an iterative approach (Gorsuch, 1983; Stevens, 2002). Each approach is associated with one or more extraction procedures, but the iterative method is the most common. The idea is to repeatedly enter communality estimates in the diagonal of the correlation matrix and find new estimates on the basis of a complete factor analysis until the estimates are stable. However, problems may arise when using the iterative process to estimate communalities including failure of the solution to converge and Heywood cases (communality estimates > 1). The first issue can be overcome by increasing the number of iterations and the second by removing the offending variable(s). After the communality estimation process, the appropriate principal factors, that is, those factors accounting for the maximum amount of variance, are then extracted using one of several factor extraction procedures.

**Factor Retention**

Central to exploratory factor analysis is determining the number of factors to retain as the “factor definition/interpretation can vary considerably depending on the number of factors kept for the final solution” (Henson, Capraro & Capraro, 2001, p. 10). There are several criteria one can use to decide on the number of factors to retain.

The Kaiser rule uses the criterion of retaining only those factors with eigenvalues greater than 1. An eigenvalue is the amount of variance in an original data set that is reproduced by a given factor (Kieffer, 1999). It has been found that this “criterion is more accurate when the number of variables is small (10 to 15) or moderate (20 to 30) and the communalities are high (> .70)” (Stevens, 2002, p. 389). The second criterion, the scree plot, is presented in the form of a graph with the eigenvalues plotted on the vertical axis against their ordinal numbers on the
horizontal axis. The plot usually begins with a steep descent because the first eigenvalue tends to be much larger than the subsequent values and then tends to level off. The shape of the plot is often thought to look like where a cliff meets a plain (scree is the debris at the bottom of a cliff). The suggestion is that the number of factors be taken as the number immediately before the plain begins or to the left of the plain.

The third criterion, the percentage of variance rule, seeks to retain as many factors as are required to account for a specified amount of the total variance (Stevens, 1986). There is no set level of variance to be extracted but, in general, investigators are satisfied with 75%, 80%, or 85% of the total variance being accounted for. It is important to note that, given the level of subjectivity involved when using the criteria described above, one should not rely on a single criterion to decide on the number of factors to retain.

*Factor Rotation*

Once the number of factors to extract has been decided, the next step is to determine the method of rotation. Rotation allows the researcher to simplify the factor structure, which results in a more meaningful and interpretable solution. The principle generally used to guide the rotation process is that of simple structure, that is, each variable should have a loading larger than .3 on one and only one factor (Gorsuch, 1983). The rotational strategies employed can be classified into two groups, orthogonal (uncorrelated) and oblique (correlated). An orthogonal rotation yields a new solution (common factor pattern) in which the factors are still uncorrelated, while an oblique rotation yields a common factor pattern, common factor structure and a correlation matrix of the factors (McDonald, 1985). The common factor pattern or pattern matrix consists of coefficients (analogous to beta weights in regression analysis) that indicate the unique contribution of each variable to each factor. The common factor structure or structure matrix
consists of the factor loadings or correlations between the observed variables and each factor.

**Interpretation of Results**

The interpretation of the rotated factors is deduced from the variables that load on – that is, correlate highly with – each factor. One criterion is to consider those factor loadings greater than .3 in absolute value to be significant. This criterion may need to be adjusted as the sample size and the number of variables increases. Typically, at least three items with significant loadings greater than .3 are needed to define a factor. Other criteria, such as whether the loadings on factors share the same theoretical meaning or whether variables loading onto different factors measure distinct constructs must also be taken into consideration.

**Differential Item Functioning and Differential Step Functioning**

One of the key statistical techniques used in this paper is differential item functioning (DIF). This section provides information on DIF and its importance in instrument development and validation.

DIF was first employed in the 1960s to assess the fairness of cognitive ability tests for minority examinees (Angoff, 1993). Such concerns are still present today, from neuropsychological tests such as the Wechsler Adult Intelligence Scale-Fourth Edition (Wechsler, 2008) to Literacy tests such as the Ontario Secondary School Literacy Test. Test developers and test users worry that items may show cultural bias (or bias based on gender or language). That is, the content of some of the items in these may be unfamiliar to minority respondents such that they would be at a greater risk of incorrectly responding to these items despite showing no difference in their overall ability compared to majority respondents. To this end, DIF controls for differences in ability by “matching respondents on an estimate of the latent ability being measured using the score on a matching subtest of items that are hypothesized to not be
functioning differentially” (Walker, 2011; p.365).

The presence of DIF suggests that in addition to the primary dimension the assessment tool was intended to measure, there are additional constructs or abilities or dimensions that are creating a difference in the scores on the original measure between two groups. It follows that DIF is indicative of a lack of measurement invariance because it implies that a scale is not measuring the same construct for all respondents (Walker, 2011). DIF analyses are most commonly conducted on two different groups known as the focal and reference group. Walker (2011) wrote that “the reference group is typically the group that one hypothesizes may have an unfair advantage of obtaining the correct answer to a particular item” (p. 365).

DIF analyses are important in terms of construct validation given that they help determine if an instrument measures only the intended construct and not any additional constructs, abilities, or dimensions. With respect to testing the measurement invariance of school connectedness and school supportiveness, DIF would be a viable methodology.

The conceptual framework of item response theory best illustrates the types of DIF (Walker, 2011). Specifically, the probability of correctly responding to an item is expressed by the following equation:

\[
P(X = 1 | \theta) = c + \frac{1 - c}{1 + e^{-a(\theta - b)}}
\]

Where \( \theta \) is the latent ability of an examinee, \( a \) is the discrimination of a particular item, \( b \) is the difficulty of that particular item, and \( c \) is the lower asymptote of that item. Very briefly, discriminating ability of an item refers to the extent to which an item is able to differentiate those that have some amount of a particular ability from those that do not. The difficulty of an item is indicated by the probability of correctly answering a particular item. Higher \( b \) values are
indicative of higher difficulty items. Theoretically, those with high levels of a particular ability should have a higher probability of correctly answering an item requiring said ability. The lower asymptote, $c$ refers to the probability that extremely low ability respondents correctly answer a test item. As such, it is commonly referred to the guessing parameter.

**Uniform DIF**

Based on the IRT framework, uniform DIF occurs when an item is equally more difficult for one group than another across all levels of ability. That is, respondents in the reference group will differ in exactly the same proportion to their focal group counterparts. According to Walker (2011) “uniform DIF exists if an item is equally discriminating for two groups, yet is more difficult for one group” (p.367).

**Non-uniform DIF**

Non-uniform DIF occurs when an item is unequally difficult for one group than another across different levels of an ability or trait. There are two subtypes of non-uniform DIF, ordinal and disordinal (Walker, Beretvas, & Acerman, 2001). Non-uniform ordinal DIF is similar to uniform DIF with the sole difference being that an item presenting non-uniform ordinal DIF is unequally discriminating for two groups compared to ordinal DIF where the difference remains the same across all ability or trait levels.

Non-uniform disordinal DIF occurs when an item that is initially difficult for one group of examinees compared to another becomes easier for the same group at different ability levels. According to Walker (2011), this type of DIF makes items appear “more difficult for focal group examinees of low ability but easier for focal group examinees of high ability, even though the difficulty parameter is identical for the reference and focal group” (p. 369).

Uniform and non-uniform DIF has been discussed at length in the literature specifically
for dichotomous items. However, Walker (2011) suggested uniform and non-uniform DIF can also be present in polytomous items scored on more than two categories such as Likert-type items. Walker (2011) also suggested that analyzing polytomous items for DIF is akin to analyzing dichotomous items for DIF.

*Procedures for Detecting DIF*

Existing statistical packages such as SPSS can assess for all types of DIF through fitting a series of logistic regression models to items hypothesized to function differently (Swaminathan and Rogers, 1990). As of late, several other statistical packages such as DIFPACK and DIFAS have been developed to test for the various types of DIF for both dichotomous and polytomous items (Penfield, 2007b; Walker, 2001). A set of nested models are compared to determine if either uniform or crossing DIF occurs. The fit of the model which best fits the data is then measured. This procedure was originally proposed by Swaminathan and Rogers (1990) for dichotomous items. However, Miller and Spray (1993) extended this procedure for use with polytomous items.

*Differential Step Functioning*

DIF analyses for polytomous items provide an omnibus measure. That is, DIF analyses tell the researcher that the item(s) show DIF without telling them where the DIF exists. This information is especially useful when examining DIF for Likert-type items. Differential Step Functioning (DSF) provides researchers with a more nuanced idea of exactly at which score level DIF lies (Penfield, 2006). In this regard, DSF analysis permits the researcher to change the slight details of a polytomous item that result in DIF as opposed to completely revising the item.
CHAPTER THREE: METHODS

This chapter first briefly describes the participants involved in this study and provides information about the COMPASS Student Questionnaire (CSQ). The statistical analyses used to explore the dimensionality and validity of the constructs measured by the CSQ are outlined in detail.

The COMPASS study is a longitudinal study spanning 4 years about youth health behaviours. The students participating in this study will be surveyed once annually for 3 or 4 years to observe changes in youth health behaviours over time and to inform school health policies and programs (Church & Leatherdale, 2013). The COMPASS study was supported by a bridge grant from the Canadian Institutes of Health Research (CIHR) Institute of Nutrition, Metabolism and Diabetes (INMD) through the "Obesity - Interventions to Prevent or Treat” priority funding awards (OOP-110788; grant awarded to S. Leatherdale) and an operating grant from the Canadian Institutes of Health Research (CIHR) Institute of Population and Public Health (IPPH) (MOP-114875; grant awarded to S. Leatherdale).

Participants

Data from the CSQ were analyzed for 24,173 students (11,886 boys, 12,076 girls) enrolled in Grades 9, 10, 11, and 12 (6,305 9th graders, 6,179 10th graders, 5,894 11th graders, and 5,699 12th graders). All students attended high schools from several school boards throughout Ontario. Approximately 77% of the participants in this study identified themselves as Caucasian, 6.6% of the participants were Black, 5.4% were Asian students, 3.6% were First Nations students, and 7.8% identified themselves as other or chose not to identify. The students in this study ranged in age from 14 to 18 years. The inclusion criteria included all English-speaking school boards that had secondary schools with grades 9 through 12 and a student population of at least 100 students or
greater per grade level; had schools that operated in a standard school/classroom settings; and permitted the use of active-information passive-consent parental permission protocols (Church & Leatherdale, 2013).

The COMPASS Student Questionnaire

The CSQ is a 30-minute classroom questionnaire that asks participating students questions regarding their height and weight, daily physical activity, daily sedentary activity, eating behaviours, experience with tobacco, alcohol and marijuana use, feelings of connectedness to their school, school supportiveness, academic achievement, and bullying. The analyses described in this paper focus primarily on students’ feelings of connectedness (referred to as closeness) to their school and the extent to which they feel supported by the school as a whole (referred to as supportiveness).

In the CSQ, school connectedness or connectedness describes the extent to which students felt a sense of belonging or closeness to their school. It is hypothesized that Item 52 examined school connectedness and is focused on the relationship between the students and their teachers and peers. Connectedness was measured with 6 sub-items on a 4-point Likert-type scale (1 = strongly agree, 4 = strongly disagree). The sub-items in Item 52 were: I feel close to people at my school; I feel I am a part of my school; I am happy to be at my school; I feel the teachers at my school treat me fairly; I feel safe in my school; and getting good grades is important to me. The closeness scale (Item 52) had good internal consistency reliability ($\alpha = .805$). However, when removing the last sub-item (“getting good grades is important to me”), the internal consistency reliability increased ($\alpha = .825$).

School supportiveness or supportiveness refers to the extent to which students perceived the school to be supportive in regard to promoting a safe and positive school environment. Item
57 is hypothesized to examine supportiveness and focused on the general climate of the school towards promoting physical activity and assistance to keep away from alcohol and drugs. Supportiveness was measured with 5 sub-items on a 4-point Likert-type scale (1 = very supportive, 4 = very unsupportive). The sub-items in Item 57 were: Making sure there are opportunities for students to be physically active; making sure students have access to healthy foods and drinks; making sure no one is bullied at school; giving students the support they need to resist or quit tobacco; and giving students the support they need to resist or quit drugs and/or alcohol. The supportiveness scale (Item 57) had good internal consistency reliability (α = .824).

**Factor Analyses**

*Exploratory Factor Analyses*

To examine the dimensionality of the sub-items pertaining to school connectedness and school supportiveness, Items 52 and 57, respectively, were subjected to exploratory factor analyses (EFA). Prior to conducting these analyses, preliminary reliability analyses were conducted to ensure internal consistency. Given that the focus of this study is on the methodological approach as opposed to the specific items, any item that decreased the overall internal consistency reliability was dropped. Using the statistical program SPSS 16.0 (SPSS Inc., 2007), exploratory factor analyses were performed using principal axis factoring. Given that there is substantial evidence in the literature for correlations between connectedness and supportiveness, Promax rotation with Kaiser Normalization was applied to each factor solution. The number of factors to retain was determined by examining the scree plot. An item was considered to load on a factor if the factor loading exceeded a value of .3.

**Differential Item Functioning and Differential Step Functioning**

DIF and DSF analyses in this study were performed using Penfield’s (2007b) DIFAS 5.0
software program which uses Mantel Chi Square. For a detailed explanation of DIFAS and the mathematical algorithms, refer to Penfield (2007a, 2007b) and Penfield, Gattamorta, and Childs (2009).

To determine if Items 52 and 57 behave differently between Grade 9 students and Grade 12 students, the data were subjected to differential item functioning (DIF) analysis. The Liu-Agresti Cumulative Common Log-odds Ratio ($\alpha_{LA}$) by Penfield (2007a) is an equivalent measure to the Mantel-Haenszel (Penfield, 2007b) and was used to determine the effect size. The rubric for categorizing the level of DIF in polytomous items suggested by Penfield (2007a) was used where $|\alpha_{LA}| < 0.53$ is negligible DIF, $0.53 \leq |\alpha_{LA}| < 0.74$ represents moderate DIF, and $|\alpha_{LA}| \geq 0.74$ is considered large DIF. Items showing moderate or strong DIF were further examined using differential step functioning analysis.

Following DIF analyses, DSF analyses were conducted on each item that exhibited DIF. This analysis showed at which score step the items functioned differentially. The DSF analysis yielded weighted and unweighted estimates of the DSF effect variance (CU-LOR) where $|CU$-LOR| > 0.4 indicates a moderate effect and $|CU$-LOR| > 0.6 represents a large effect.
CHAPTER FOUR: RESULTS AND DISCUSSION

The analyses were conducted in three stages. After computing descriptive statistics for students in Grade 9 and Grade 12, the internal consistency reliability among the sub-items were computed and sub-items that increased the overall internal consistency if removed were removed prior to proceeding with factor analyses. An exploratory factor analysis (EFA) was then conducted for Grades 9 and 12 separately to determine if there exist common, shared variance amongst all sub-items. DIF analyses were then conducted to determine whether individual items within each construct functioned differentially between students in Grade 9 and in Grade 12. Those items showing large DIF were then used to perform DSF analyses. A second DIF analysis was then performed between those Grade 12 students who reported spending their last year in another school and Grade 9 students. Again, for those items that revealed large DIF, additional DSF analyses were performed.

The sub-items measuring closeness were at best poorly correlated with the sub-items examining supportiveness. These modest interrelations suggested that each item on a whole had sufficiently unique variance to justify separate examination. Generally speaking, students tended to endorse feelings of closeness and supportiveness. The data were positively skewed, with the majority of students tending to endorse intermediate responses such as “agree” and “disagree” for Item 52 and “supportive” and “unsupportive” for Item 57.

As previously mentioned, the closeness scale (Item 52) had good internal consistency reliability ($\alpha = .805$). However, when removing the last sub-item (“getting good grades is important to me”), the internal consistency reliability increased ($\alpha = .825$). The supportiveness scale (Item 57) had good internal consistency reliability ($\alpha = .824$). The ensuing factor analyses were conducted using a total of 10 items covering both closeness and supportiveness.
Descriptive Statistics

Summarized in Table 1 are the percentages of Grade 9 and Grade 12 students endorsing the responses “strongly agree” and “very supportive” for items 52 and 57, respectively. Roughly the same percentage of students in both grades responded to Item 52 with “strongly agree”. The disparities in percentages across grades for Item 57 are greater than in Item 52 with Grade 9 students feeling that the school is more supportive with respect to all aspects covered in the sub-items. The least percentage of Grade 9 and Grade 12 students felt that their school was supportive in resisting or quitting tobacco, drugs, and alcohol. Interestingly, the greatest amount of disparity between Grade 9 and Grade 12 students occurred in the sub-items pertaining to resisting or quitting tobacco, drugs, and alcohol. The highest percentage of Grade 9 and Grade 12 students felt that their school provides them with opportunities to be physically active.

Also presented in Table 1 are the results of a chi-square ($\chi^2$) test of independence that indicate, with the exception of sub-items 52A and 52B, significantly different proportions of students in Grades 9 and 12 and their tendency to endorse responses of “strongly agree” and “very supportive” for items 52 and 57, respectively. The proportion of grade 9 students responding with “strongly agree” or “very supportive” is greater for all sub-items except for 52D and 52E.
Table 1. Percentage of Grade 9- and Grade 12- High-School Students Strongly Agreeing and Very Supportive

<table>
<thead>
<tr>
<th>Item</th>
<th>Administration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Grade 9</td>
</tr>
<tr>
<td>52. How strongly do you agree or disagree with each of the following statements?</td>
<td></td>
</tr>
<tr>
<td>a) I feel close to people at my school.</td>
<td>22.0</td>
</tr>
<tr>
<td>b) I feel I am a part of my school.</td>
<td>19.7</td>
</tr>
<tr>
<td>c) I am happy to be at my school.</td>
<td>24.9*</td>
</tr>
<tr>
<td>d) I feel the teachers at my school treat me fairly.</td>
<td>21.4*</td>
</tr>
<tr>
<td>e) I feel safe in my school.</td>
<td>24.0*</td>
</tr>
<tr>
<td>57. How supportive is your school of the following?</td>
<td></td>
</tr>
<tr>
<td>a) Making sure there are opportunities for students to be physically active.</td>
<td>45.4*</td>
</tr>
<tr>
<td>b) Making sure students have access to healthy foods and drinks.</td>
<td>29.7*</td>
</tr>
<tr>
<td>c) Making sure no one is bullied at school</td>
<td>29.0*</td>
</tr>
<tr>
<td>d) Giving students the support they need to resist or quit tobacco.</td>
<td>18.8*</td>
</tr>
<tr>
<td>e) Giving students the support they need to resist or quit drugs and/or alcohol.</td>
<td>20.8*</td>
</tr>
</tbody>
</table>

*p <.01 for chi-squared tests with one degree of freedom comparing the proportions of those strongly agreeing or very supportive against all other responses combined for Grades 9 and 12.
Factor analysis was used to examine the dimensionality or factor structure of the data for items 52 and 57. Exploratory factor analyses revealed a two-factor solution when examining the two sets of sub-items together.

Exploratory Factor Analyses

The minimum amount of data for factor analysis was satisfied, with a final sample size of 21995 (using list wise deletion). The factorability of the final 10 sub-items constituting Items 52 and 57 was examined. Several well-recognized criteria for the factorability were employed. The data showed reasonable factorability with all items correlated at least .3 with at least one other item. Secondly, the Kaiser-Meyer-Olkin measure of sampling adequacy was .823, above the commonly recommended value of .6, and Bartlett’s test of sphericity was significant ($\chi^2$ (45) = 104700, $p < .001$). Lastly, the communality values were all at least above .3 (see Table 2), further confirming that each item shared some common variance with other items. As such, factor analysis was deemed to be suitable for the 10 sub-items.

Table 2 provides the correlations of individual sub-items with each factor, along with their estimates for communality. Examining the sub-items with factor loadings greater than .3 suggests that the first factor pertains to the feelings of connectedness or closeness felt by those students in Grades 9 and 12 and the second factor concerns the extent to which students feel that the school is supportive. Also, the factor structures appear to be consistent across grades.
Table 2.

**Summary of Items and Factor Loadings for Promax Oblique Two-Factor Solution for Items 52 and 57 of the COMPASS Questionnaire-Grades 9 and 12**

<table>
<thead>
<tr>
<th>Item/Sub-item</th>
<th>Grade 9 Factor Loadings</th>
<th>Grade 12 Factor Loadings</th>
<th>Grade 9 Communalities</th>
<th>Grade 12 Communalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>52. How strongly do you agree or disagree with each of the following statements?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) I feel close to people at my school.</td>
<td>.689</td>
<td>.711</td>
<td>.439</td>
<td>.478</td>
</tr>
<tr>
<td>b) I feel I am a part of my school.</td>
<td>.756</td>
<td>.787</td>
<td>.548</td>
<td>.601</td>
</tr>
<tr>
<td>c) I am happy to be at my school.</td>
<td>.774</td>
<td>.811</td>
<td>.611</td>
<td>.650</td>
</tr>
<tr>
<td>d) I feel the teachers at my school treat me fairly.</td>
<td>.522</td>
<td>.634</td>
<td>.354</td>
<td>.349</td>
</tr>
<tr>
<td>e) I feel safe in my school.</td>
<td>.661</td>
<td>.532</td>
<td>.485</td>
<td>.438</td>
</tr>
<tr>
<td>57. How supportive is your school of the following?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a) Making sure there are opportunities for students to be physically active.</td>
<td>.366</td>
<td>.381</td>
<td>.271</td>
<td>.317</td>
</tr>
<tr>
<td>b) Making sure students have access to healthy foods and drinks.</td>
<td>.473</td>
<td>.466</td>
<td>.312</td>
<td>.297</td>
</tr>
<tr>
<td>c) Making sure no one is bullied at school.</td>
<td>.627</td>
<td>.563</td>
<td>.504</td>
<td>.471</td>
</tr>
<tr>
<td>d) Giving students the support they need to resist or quit tobacco.</td>
<td>.941</td>
<td>.927</td>
<td>.812</td>
<td>.786</td>
</tr>
<tr>
<td>e) Giving students the support they need to resist or quit drugs and/or alcohol.</td>
<td>.934</td>
<td>.928</td>
<td>.798</td>
<td>.785</td>
</tr>
</tbody>
</table>

*Note. Factor loadings with absolute value < .3 not included*
Principal axis factoring was employed as the factor extraction method. Initial eigenvalues indicated that the first two factors explained 42.43% and 17.3%, respectively. Solutions for up to ten factors were examined using Promax rotations of the factor loading matrix. The two-factor solution, which explained nearly 59% of the variance was preferred because of: (1) its previous theoretical support; (2) the leveling of eigenvalues on the scree plot after two factors; and (3) the insufficient number of primary loadings for subsequent factors.

No items were eliminated because they adequately contributed to a simple factor structure and met the minimum criteria for having a primary factor loading of .4 or above, and no cross-loading of .3 or above (see Table 2). Overall, the EFA indicated that two distinct factors were underlying students’ responses to items 52 and 57 and that these factors were moderately internally consistent with correlations of 0.434 and 0.430 between closeness and supportiveness, for Grade 9 and 12 students, respectively. This suggests that Grade 9 and Grade 12 students’ feelings of connectedness to their school are only moderately associated with how their perception of school supportiveness.

DIF and DSF Analyses

Table 3 provides the DIF analysis results. The last column indicates the direction of DIF, indicating whether the sub-item favours Grade 9 students or Grade 12 students. In regard to school connectedness, the results provide no evidence of DIF with the exception of moderate DIF in sub-item 52E. The results show that there is no DIF in all the sub-items for school supportiveness.

Table 4 presents the DIF analysis results for Grade 9 students and only those Grade 12 students who attended a different school in the previous year. As in Table 3, the last column indicates the direction of DIF, indicating whether the sub-item favours Grade 9 students or Grade 12 students. Sub-items 52B, 52D, and 52E show DIF, with 52B favouring Grade 9 students and 52D and 52E favouring Grade 12 students.
Table 3. *Differential Item Functioning Results*

<table>
<thead>
<tr>
<th>Item</th>
<th>Mantel</th>
<th>L-A LOR</th>
<th>DIF</th>
<th>Direction of DIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Connectedness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52A</td>
<td>52.887</td>
<td>0.336</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52B</td>
<td>24.609</td>
<td>0.245</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52C</td>
<td>100.815</td>
<td>0.502</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52D</td>
<td>70.792</td>
<td>-0.393</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52E</td>
<td>182.871</td>
<td>-0.704</td>
<td>Moderate</td>
<td>Grade 12 Students</td>
</tr>
<tr>
<td><strong>School- Supportiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57A</td>
<td>1.546</td>
<td>-0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57B</td>
<td>59.167</td>
<td>-0.344</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>57C</td>
<td>42.277</td>
<td>-0.304</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>57D</td>
<td>76.865</td>
<td>0.43</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>57E</td>
<td>78.838</td>
<td>0.434</td>
<td>Negligible</td>
<td>-</td>
</tr>
</tbody>
</table>
### Differential Item Functioning Results - First Year at High School

<table>
<thead>
<tr>
<th>Item</th>
<th>Mantel</th>
<th>L-A LOR</th>
<th>DIF</th>
<th>Direction of DIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>School-Connectedness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>52A</td>
<td>8.01</td>
<td>0.493</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52B</td>
<td>22.98</td>
<td>0.908</td>
<td>Large</td>
<td>Grade 9 Students</td>
</tr>
<tr>
<td>52C</td>
<td>7.613</td>
<td>0.519</td>
<td>Negligible</td>
<td>-</td>
</tr>
<tr>
<td>52D</td>
<td>32.11</td>
<td>-1.155</td>
<td>Large</td>
<td>Grade 12 Students</td>
</tr>
<tr>
<td>52E</td>
<td>15.251</td>
<td>-0.669</td>
<td>Large</td>
<td>Grade 12 Students</td>
</tr>
<tr>
<td><strong>School- Supportiveness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57A</td>
<td>0.002</td>
<td>0.008</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>57B</td>
<td>0.014</td>
<td>-0.021</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>57C</td>
<td>2.549</td>
<td>-0.285</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>57D</td>
<td>1.025</td>
<td>0.185</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>57E</td>
<td>0.815</td>
<td>0.183</td>
<td></td>
<td>-</td>
</tr>
</tbody>
</table>
The moderate DIF exhibited by item 52E (Table 3), suggests that Grade 12 students require less of a feeling of closeness to their school in order to feel as though their teachers treat them fairly compared to their Grade 9 counterparts. This may be in part due to the fact that Grade 12 students are not associating teacher fairness to school closeness. That is, to Grade 12 students, school closeness and teacher fairness are independent. Conversely, Grade 9 students require higher closeness levels compared to Grade 12 students in order to respond to sub-item 52E at the same level. Considering findings from Klaczynski (1997), this finding possibly suggests that Grade 9 students are attributing how close they feel to the school to how they feel about their teachers. This is a plausible explanation as Grade 9 students have only begun to attend a new school with no knowledge about the teachers compared to Grade 12 students who may have had prior experiences with the teachers in question.

Items with moderate or large DIF were further analyzed for DSF. The findings are presented in Table 5. The DSF findings show that some sub-item 52B favours Grade 9 students while sub-items 52D and 52E favours Grade 12 students. Considering the IRT framework suggested by Walker (2011) for interpreting DIF results, there are several interesting findings. Surprisingly, the DIF results show that for sub-item 52B (“I feel like I am a part of my school”), Grade 9 students require less of a feeling of school connectedness compared to Grade 12 students to endorse this sub-item at the same level. One possible explanation for this unexpected finding may be related to Klaczynski’s (1997) and Steinberg’s (2005) findings that suggest that patterns of cognitive development are influenced by the emotional and social context in which reasoning occurs. Arguably, the Grade 9 students in this study that have recently switched schools which usually make an effort to welcome new students through orientation activities involving the new students to integrate with current students, facilitating the move to a new school.

In contrast, DIF results suggest that Grade 12 students require lower levels of school connectedness to endorse sub-items 52D and 52E at the same level as their Grade 9 counterparts. After controlling for familiarity effects by including only those Grade 12 students who attended a different school
last year, the results suggest that these sub-items are measuring a construct aside from school connectedness. Grade 12 students may be answering sub-items 52D and 52E in a way that does not necessarily factor in school connectedness. The findings suggest the need to revise these sub-items or replace them with items that have previously demonstrated good construct validity.

The results show that the bias these sub-items show persist despite controlling for differences in closeness levels by conducting DIF analyses only on those students who reported spending their last year in another school. These results suggest that these sub-items are not necessarily targeting the construct of school closeness. The results imply that these sub-items are measuring dimension(s) in addition to school connectedness. The findings also call into question the stability of the construct of school connectedness in the way it has been operationalized in this study. Consequently, it is important to conduct a qualitative review of the items that were found to exhibit DIF.

Table 5.

<table>
<thead>
<tr>
<th>Item</th>
<th>Step</th>
<th>CU-LOR</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>52B</td>
<td>1</td>
<td>1.105</td>
<td>3.244</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.072</td>
<td>4.326</td>
</tr>
<tr>
<td>52D</td>
<td>1</td>
<td>-1.148</td>
<td>-2.226</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.731</td>
<td>-5.164</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.712</td>
<td>-3.078</td>
</tr>
<tr>
<td>52E</td>
<td>2</td>
<td>-0.888</td>
<td>-3.085</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.747</td>
<td>-3.183</td>
</tr>
</tbody>
</table>
CHAPTER FIVE: CONCLUSION

Taken together, the results of this study hold notable implications for the use of the CSQ given that it will be used in the larger COMPASS study to measure changes in school connectedness and school supportiveness. Specifically, this study shows that the construct(s) that are intended to be measured by an instrument should be stable across time and across different groups. Moreover, it is important to establish measurement invariance in order to make valid inferences.

Limitations

A few limitations of this study impact the interpretation of the results. The results are based on a cross-sectional design which may differ had a longitudinal study been conducted. Given that the COMPASS project is a longitudinal project, a follow-up study should be conducted with longitudinal data to see if the trends observed in this study persist or if new findings emerge that can potentially inform the use of the CSQ in tracking students’ health behaviours over time.

This study conducted DIF analyses without first conducting CFA which provides a robust test of dimensionality. Other models may exist that describe the factor structure or dimensionality of the data better than the two-factor model found by EFA.

Future Directions

Given the results from this study, it would be important to replicate this study on a longitudinal data set to truly (so that we actually can observe effects of time on these constructs thus be able to say something about the measurement invariance since we’re doing the tests on the same people). Also, need to run confirmatory factor analysis to ascertain the dimensionality of the data. Given that the objective of this longitudinal study is to track changes in students’ attitudes over time, it is important to address the issues related to measurement invariance. Moreover, future data should be subjected to con-
firmatory factor analyses to validate the results of the exploratory factor analysis conducted in this study. This would provide researchers with a clearer idea whether the CSQ is measuring the intended constructs.

Lastly, once the relevant sections of the CSQ have been validated, it should be administered on several occasions, and the factor structures should be assessed to ensure that meaningful and valid comparisons can be made about students’ feelings of school connectedness, school supportiveness, and overall health behaviours over time.
REFERENCES


