A COMPARISON OF CFA AND ESEM APPROACHES USING TIMSS SCIENCE ATTITUDES ITEMS: EVIDENCE FROM FACTOR STRUCTURE AND MEASUREMENT INVARIANCE

by

Ji Yoon Jung

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Education

Department of Educational Studies

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Anne Traynor

Department of Educational Studies

Dr. Yukiko Maeda

Department of Educational Studies

Dr. Susan South

Department of Psychological Studies

Dr. Selcen Guzey

Department of Curriculum and Instruction

Approved by:

Dr. Richard Olenchak

Head of the Graduate Program

ACKNOWLEDGMENTS

I would like to thank my academic advisor, Dr. Anne Traynor, for her sincere support and guidance throughout my master's program. I truly enjoyed my academic journey at Purdue, which would not be possible without her invaluable mentorship. I would also like to thank the rest of my committee members, Dr. Yukiko Maeda and Dr. Susan South, for contributing to the development of my research.

I am deeply grateful to Dr. Selcen Guzey and Dr. Lynn Bryan for providing me the wonderful opportunity to join in the PULSE project. I have been fortunate to have worked with the PULSE colleagues, who are always professional and supportive. Working for this project has inspired my research interest and enabled me to balance academic work and parenting with a sense of happiness.

I want to express many thanks to my husband, Ikbeom Jang, for his unwavering encouragement and love throughout my graduate studies. He has been here every step of the way. I also express my deep gratitude to my parents and parents-in-law. Last but not least, I would like to thank my precious little one, Evelyn (Sunyu), for coming into my life, and more importantly, for making me her mother.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Jung, Ji Yoon. MSED
Institution: Purdue University
Degree Received: May 2019
Title: A Comparison of CFA and ESEM Approaches Using TIMSS Science Attitudes Items: Evidence from Factor Structure and Measurement Invariance
Committee Chair: Dr. Anne Traynor

The power of positive attitudes toward science is that they influence science achievement by reinforcing higher performance. Interestingly, there continue to be gender disparities in attitudes toward science across many countries. Males generally have more positive attitudes toward science than females. Although most research related to attitudes toward science have been based on the Trends in International Mathematics and Science Study (TIMSS) Student Questionnaire, there remains a dearth of evidence validating the TIMSS science attitudes items and measurement equivalence across genders.

The goals of this research were as follows: (1) to build support for the structural validity of the TIMSS items, and (2) to investigate whether the instrument measures the same latent construct (attitudes toward science) across genders. The present study followed two steps of statistical analyses. As a first step, two modeling methods (confirmatory factor analysis and exploratory structural equation modeling) were conducted to identify the best-fitting model for the instrument. Second, after determining the model of choice, we tested several nested invariance models progressively.

This study found (1) the latent factor structure of the TIMSS items and (2) strong measurement invariance across genders. This result indicated that the instrument is well designed by the *a priori* specification and measures the same latent variable for both female and male students. This study provides support for the multidimensional approach to measuring science

attitudes and shows the flexibility of ESEM over CFA by demonstrating that the ESEM approach provided better representation of the underlying factor structure.

CHAPTER 1. INTRODUCTION

Problem Statement

There has been an increasing demand on the Science, Technology, Engineering, and Mathematics (STEM) workforce in the United States to maintain its leadership in the world economy (Banning & Folkestad, 2012; Hossain, 2012; Langdon, McKittrick, Beede, Khan, & Doms, 2011). Despite the growing need, the majority of STEM-related jobs are expected to be unfilled due to the lack of quality workers (Atkinson, 2013; Guzey, Harwell, & Moore, 2014). Although the United States has experienced steady growth in the number of STEM-related workers, it still lags behind the fast growth of European and Asian competitors (National Science Board, 2018). Hence, attracting more students to be engaged in STEM fields should be a top priority in STEM education in order to maintain the competitiveness of the United States.

With respect to attitudes toward science, there is a considerable agreement on the necessity of developing a valid measure of attitudes toward science (Osborne, Simon, & Collins, 2003). Nevertheless, existing instruments designed to measure students' attitudes toward science tend to have problems, such as a lack of a theoretical framework and empirical evidence to support construct validity (Bennett, 2001; Cheung, 2009; Wang & Berlin, 2010). Furthermore, although there have been gender disparities in attitudes toward science in the United States, with male students expressing more favorable attitudes (Brotman & Moore, 2008; Smith, Pasero, & McKenna, 2014), few instruments have had their measurement invariance across genders tested. For example, although the Test of Science-Related Attitudes (TORSA; Fraser, 1981) has been extensively used in science education research (Lang, Wong & Fraser, 2005), Cheung (2009) stated that the factor analysis failed to support the distinctiveness of the seven subscales used in the study: Social Implications of Science, Normality of Science, Attitude to Scientific Inquiry, Adoption of Scientific Attitudes, Enjoyment of Science Lessons, Leisure Interest in Science, and Career Interest in Science. Similarly, the developers of the Scientific Attitude Inventory (Moore & Sutman, 1970) did not perform thorough validity tests, such as confirmatory factor analysis or item response analysis, which means that the validity data accumulated for the instrument is very weak (Aydeniz & Kotowski, 2014). With respect to the Colorado Learning Attitudes About Science Survey (Adams, Perkins, Dubson, Finkelstein, & Wieman, 2005), it did not employ a conceptual framework for defining attitudes toward science, it and included several items that did not clearly measure the target construct – attitudes (Aydeniz, & Kotowski, 2014).

For the Trends in International Mathematics and Science Study (TIMSS; Hooper, Mullis, & Martin, 2015) context questionnaire, three scales were constructed to measure students' attitudes: positive affect, self-confidence, and valuing science (Martin & Preuschoff, 2007; Martin, Mullis, & Hooper, 2016). Although the structural construct validity of the most recent TIMSS background survey from 2015 has been examined with principal components analysis (Martin, Mullis, & Hooper, 2016), studies analyzing its factor structure with alternate models are scarce. Given that principal components are not latent variables (Fabrigar, Weagener, MacCallum, & Strahan, 1999), it is instructive to examine the underlying structure with other validity tests, such as factor analysis or structural equation modeling, to examine whether the items apparently measure their intended factors. Also, science attitudes instruments have been critiqued for their weakness in justifying validity (Wang & Berlin, 2010). Although it seems unfeasible to agree upon an accurate technical definition for validity (Camargo, Herrera, & Traynor, 2018; Newton & Shaw, 2013), this study adapted the following definition: validity is the degree to which evidence and theory support the interpretations of test scores for the intended uses of tests (AERA, APA, & NCME, 2014). To summarize, the current study contributes to the evidence on the validity of the TIMSS science attitudes items.

Purpose of the Study

The purpose of this research was to investigate (1) the factor structure of the TIMSS 2015 science attitudes items, and (2) its measurement invariance for female and male students. Specifically, two major research questions are:

Research Question 1. Do the TIMSS 2015 attitudes toward science items measure what they are designed to measure? Do they measure their intended factors (*Students Like Learning Science*, *Students Confident in Science*, and *Students Value Science*) according to the priori specification?

Research Questions 2. Do the TIMSS 2015 science attitudes items measure the same trait across gender groups? Do any observed gender differences in attitudes toward science reflect true differences in attitudes?

Significance of the Study

Given that the STEM workforce is an essential driver of the US economy (Xue & Larson, 2015), reinforcing students' attitudes toward science has shown to be important in science education. Fostering students' positive attitudes toward science is the first step to attracting more students to the STEM fields, so it is important to measure students' attitudes with a valid instrument. A body of research studying US students' attitudes about studying science has relied on TIMSS results (House & Telese, 2008; Papanastasiou & Papanastasiou, 2004; Smith, Pasero, & McKenna, 2014; Stiles, Adkisson, Sebben, & Tamashiro, 2008). Despite the frequent use of the TIMSS attitudes scale, few studies have explored and tested the validity of the items. Thus, it is necessary to validate the TIMSS science attitudes items by investigating (1) its latent factor structure and (2) the existence of measurement invariance across genders.

CHAPTER 2. LITERATURE REVIEW

Students' Attitudes toward Science

Definition of Attitudes toward Science

The definition of attitudes as they relate to science is inconsistent and often poorly articulated (Blosser, 1984; Germann, 1998; Osborne, Simon & Collins, 2003; Scantlebury, Tal, & Rahm, 2007). Thurstone (1931), Fishbein (1967), and Mueller (1986) favored a unidimensional concept of attitudes, defining the construct as the affect for or against a psychological object. By accepting this unitary view, attitudes toward science can be defined as a general and enduring feeling about science, and a predisposition to learn science (Koballa & Crawley, 1985; Lovelace & Brickman, 2013). Similarly, George (2006) described attitudes toward science as positive or negative feelings about science, especially science classes. On the other hand, according to psychologists supporting a multidimensional conception, attitudes include a myriad of variables such as judgments of personal ability in science, the value of science to the individual, the value of science to society, attitudes toward methods of teaching science and attitudes toward scientific interests (Germann, 1988). Hassan (2008) mentioned that attitudes toward science include several subconstructs: motivation for science, lack of anxiety, the usefulness of science, self-concept of ability, and ability to make choices and career interest. In summary, studies advocating a multidimensional perspective have incorporated a range of subconstructs as follow: the value of science, self-esteem about science, motivation toward science, enjoyment of science, attitudes of parents toward science and so forth (Osborne et al, 2003).

In the case of the TIMSS science attitudes items, they are not based on a specific definition of attitudes toward science. However, they were designed based on a multidimensional conception of attitudes with three factors: *Students Like Learning Science* (SLS), *Students Confident in Science* (SCS), and *Students Value Science* (SVS) (Martin, Mullis & Hooper, 2016; Mullis & Martin, 2017). The SLS scale is intended to measure a student's intrinsic motivation to learn science (Mullis & Martin, 2017). Deci and Ryan (1985) defined intrinsic motivation as the energizer of behavior. Specifically, academic intrinsic motivation is an orientation toward learning challenging, difficult, and novel tasks (Gottfried, 1990). The SCS scale is meant to measure a student's academic self-concept (Mullis & Martin, 2017). According to Bong & Skaalvik (2003), academic self-concept pertains to individuals' knowledge and perceptions about themselves with respect to an achievement situation. The SVS scale assesses a student's extrinsic motivation to learn science, which is driven by external rewards such as praise or career success (Mullis & Martin, 2017). Furthermore, TIMSS results have consistently demonstrated that SLS, SCS, and SVS scores have a strong relationship with students' academic performance (Mullis & Martin, 2017).

Although TIMSS has been extensively used for measuring students' attitudes toward science with the three factors aforementioned, few studies have investigated its latent factor structure. It is crucial to examine the underlying structure of the TIMSS attitudes items because there remains a dearth of evidence that the items targeted for each scale (SLS, SCS, and SVS) actually measure what they are supposed to measure. As the items were constructed without a specific theory of attitudes, exploring and supporting the factor structure is important for building on the validity of the instrument. Even though Smith, Pasero, and McKenna (2014) analyzed attitudes toward science using the three factors of SLS, SCS, and SVS, they did not examine the underlying structure of the TIMSS attitudes items. Furthermore, the latent structure of the items should be determined in advance in order to investigate gender disparity in attitudes toward science.

Given that previous studies have noted a gender gap in science attitudes (Blickenstaff, 2005; Christidou, 2011 Hayes & Tariq, 2000), it is critical to examine whether such differences represent true disparity in attitudes or derive from measurement variance across genders.

Importance of Fostering Students' Attitudes toward Science

The promotion of attitudes toward science is an issue that has received long standing attention in science education (Barmby, Kind, & Jones, 2008). Osborne and colleagues (2003) characterized developing positive attitudes toward science as a critical agenda for research.

Improving positive attitudes is particularly important in science education mainly for two reasons. Above all, attitudes heavily influence students' science achievement by reinforcing higher or lower performance (Cannon & Simpson, 1985; Papanastasiou & Papanastasiou, 2004; Papanastasiou & Zembylas, 2002; Ozel, Caglak, & Erdogan, 2013; Reynolds & Walberg, 1992). As low levels of positive attitudes toward science are likely to lead to students' apathy toward science or dropping from advanced science classes (Nieswandt, 2007), Hong and Lin (2011) mentioned that science educators should pay more attention to attitudes toward science. Singh, Granville, and Dika (2002) examined the effects of motivation, attitudes, and academic engagement on students' achievement in mathematics and science. They found that the attitudes toward science had the second-largest effect on science learning, following the academic time factor. Oliver and Simpson (1985) also revealed that attitudes are not only a substantial predictor of science achievement, but they also explain a large portion of the variance in achievement. Specifically, their attitudes variables accounted for approximately 20% of the variance in chemistry achievement.

Moreover, the promotion of positive attitudes toward science can encourage science-related careers (Becker & Park, 2011; George, 2006; Wang & Staver, 2001; Ware & Lee, 1988). Previous studies revealed that the one of the strongest determinants of pursuing a STEM career is students' attitudes toward science and mathematics in their adolescence (Correll, 2001; Maltese & Tai, 2011; Maple & Stage, 1991). Similarly, Sadler, Sonnert, Hazari, and Tai (2012) stated that experiences and attitudes developed prior to high school vastly contribute to an interest in STEM careers. Riegle-Crumb and colleagues (2011) examined the relationship between students' career aspirations and attitudes toward science and mathematics, which were represented by intrinsic interest and self-concept. They found that while positive attitudes are not necessarily the strongest predictors of how well students perform on standardized tests, nevertheless attitudes may still be important for keeping students interested in pursuing a STEM-related career in the future. In fact, according to social cognitive models, attitudes are one of the important constructs that exert substantial influence on students' pursuit of STEM courses and careers (Rice, Barth, Guadagno, Smith, & McCallum, 2013). Specifically, Rice and colleagues (2013) illustrated that students with greater support from parents, teachers, and peers were likely to have more positive attitudes toward math and science, and thereby they reported higher competence in these subjects. Examining the effect of attitudes from the precollege perspective, Wang (2013) revealed that an intent to major in STEM is subject to early attitudes toward STEM fields.

Confirmatory Factor Analysis and Exploratory Structural Equation Modeling

Confirmatory factor analysis (CFA) has been extensively used in the construction and development of psychoeducational instruments (DiStefano & Hess, 2005). However, Sass and Schmitt (2011) proposed not assuming CFA is the most suitable statistical model simply due to its popularity. They said that it is especially true when possible cross-loadings on factors have not been explored before. In fact, the typical independent clusters model of CFA (ICM-CFA) has been criticized because the assumption of CFA – that all items should load on only one factor without any cross-loadings – may be overly restrictive for most multidimensional instruments (Arens & Morin, 2016; Marsh et al, 2009). Such a restrictive requirement generally leads to inflated factor correlations, and thereby structural relations between factors can be distorted (Asparouhov & Muthen, 2009; Marsh et al, 2009; Morin, Arens, Tran, & Caci, 2016; Schmitt, 2011). Also, the exclusion of nonzero cross-loadings in CFA can distort the size of relations among the factors, as well as result in a poor fit to the data (Marsh et al, 2009). Given that the TIMSS science attitudes scale regards attitudes as a multidimensional concept with correlated dimensions, it is instructive to investigate its latent structure outside of the ICM-CFA approach.

Marsh and colleagues (2014) suggested exploratory structural equation modeling (ESEM) as a more preferable model to ICM-CFA due to its remarkable flexibility. While new, ESEM is a new modeling framework that incorporates traditional exploratory factor analysis (EFA), CFA, and structural equation modeling. Although ESEM factors are basically EFA factors, ESEM can be viewed as a primarily confirmatory approach with the use of target rotation, because target rotation assumes *a priori* latent factor structure like CFA does (Browne, 2001). Meanwhile, the major difference between CFA and ESEM lies in the incorporation of all possible cross-loadings in the ESEM model. The cross-loadings that are constrained to zero in CFA are freely estimated in ESEM (Marsh, Morin, Parker, & Kaur, 2014). One of the advantages of considering possible cross-loadings is that it provides more accurate factor correlations. Also, estimating the cross-loadings in the scale development could suggest more accurate information about what factors each item is primarily measuring. In other words, cross-loadings address how items might be fallible indicators of the factor they are intended to measure and represent at least some degree of relationship with other factors (Arens & Morin, 2016). As the underlying structure of psychoeducational assessments measuring interrelated constructs may often include cross-loadings, the fit of ESEM is expected to be better than that of independent clusters of CFA models to the science attitudes items. In general, ESEM might be viewed as the model of choice if it fits the data better than a corresponding independent clusters CFA model does (Marsh, Morin, Parker, & Kaur, 2014).

Recently, ESEM has been applied in educational and psychological research to evaluate various multidimensional constructs (Morin, Marsh, & Nagengast, 2013). For instance, Caro, Sandoval-Hernández, and Lüdtke (2014) compared the model fit of CFA and ESEM using international large-scale assessment data from the Progress in International Reading Literacy Study (PIRLS) 2006 and the Programme for International Student Assessment (PISA) 2009. They evaluated and compared factor structure of response data from item sets meant to measure cultural, economic, and social capital. For both PIRLS and PISA, an ESEM solution provided more acceptable fit indices as well as stronger support for discriminant validity, than CFA. Guay, Morin, Valois, and Vallernad (2015) examined the construct validity of scores from the Academic Motivation Scale using CFA and ESEM. They found that ESEM yielded a better fit to the data, and the pattern of factor correlations from ESEM was more aligned with their theoretical framework than that from CFA. Also, Joshanloo and Lamers (2016) conducted CFA and ESEM with a mental well-being assessment and found that ESEM outperformed CFA in capturing the factor structure. With ESEM, they successfully distinguished two dimensions of well-being, which were not empirically distinct in CFA. In summary, ESEM contributes to testing and examining the latent structure of multidimensional measures by overcoming some of the limitations of CFA. This advantage over CFA may enable a more thorough investigation of the factor structure of TIMSS attitudes items, which were based on a multidimensional conceptualization.

Measurement Invariance across Genders

In the present study, measurement invariance is examined to test if any observed gender differences in science attitudes reflect true differences in attitudes, or a failure of the instrument to measure the trait equivalently across gender groups. Studies have reported that boys and girls show different attitudes toward science, with boys expressing more positive attitudes (Blickenstaff, 2005; Christidou, 2011; Osborne, Simon, & Collins, 2003; Weinburgh, 1995). According to Smith, Pasero, and McKenna (2014), gender disparity in science attitudes was found in the 2011 TIMSS results. They found that fourth and eighth grade boys in the United States. showed more confidence in science than girls, and such discrepancy across genders becomes more generalized as students develop. In a similar vein, Brotman and Moore (2008) revealed that a body of large-scale studies reported continued inequities in attitudes toward science; female students' overall attitudes toward science are either less positive than male students' in cross-sectional data or decrease more substantially with age in longitudinal data.

In spite of these predominant trends, a few studies demonstrated variations in attitudes toward science. According to Anwer, Iqbal, and Harrison's study of Pakistan (2012), tenth grade girls showed significantly more positive attitudes toward science than boys. Girls expressed more favorable attitudes on five subscales of the social implications of science, attitudes to scientific inquiry, adoption of scientific attitudes, enjoyment of science lessons, and leisure interest in science. Although boys showed more career interest in science, the difference was insignificant. Similarly, some research has revealed that girls tend to be positive and confident in learning science and support women in science. Baker and Leary (1995) used qualitative methods such as interviews to represent girls' attitudes toward science in more detail. They found that second, fifth, eighth, and eleventh grade girls in the United States enjoyed learning science, and were confident in their ability to do well in science rather than avoiding it. According to Buck and colleagues (2009), the majority of fourth, fifth, and sixth grade African American girls expressed a high level of confidence, desire, and value to learn science. On the other hand, several studies reported that no gender differences were found in attitudes toward science (Dhindsa & Chung, 2003; Scantlebury, Baker, Sugi, Yoshida, & Uysal, 2007). Also, Zeidan and Jayosi (2015) found that there were no significant differences in attitudes toward science due to gender. In summary, although there exist overall trends in attitudes toward science across genders, the literature is inconclusive regarding gender differences in science attitudes. Hence, a first step may be to investigate whether or not science attitude items measure the same latent construct across genders to examine the true relationship between gender and attitudes toward science.

CHAPTER 3. METHODS

Data Source

The data used in the current study was drawn from TIMSS 2015, directed by the International Association for the Evaluation of Educational Achievement (IEA). The U.S. sample consisted of 246 schools and 10,221 eighth graders (50.1% female, 49.9% male). TIMSS 2015 followed a stratified two-stage sample design; with the first stage was a sample of schools, and the second stage was a sample of classes from each school (Averett, Ferraro, Tang, Erberber, & Stearns, 2017). Stratification was used to increase sample efficiency and consistency of the sample design.

The U.S sample was based on three explicit stratification variables (poverty level, school type, and census region) and two implicit stratification variables (urbanization and ethnicity status). The present study considered the four variables except census region as it was not included in either the public-use or restricted-use data file. Participants responded to a Student Questionnaire that provides information about students' home and school lives (Foy, 2017). The questions that have been used for this study were designed to assess students' attitudes towards science. TIMSS examinees responded to 26 categorical and polytomous items with 9 items intended to measure *Students Enjoy Learning Science* (SES), 8 items intended to measure *Students Confidence in Science* (SCS), and 9 items intended to measure *Students Perceived Value Learning Science* (SVS) (Smith, Pasero, & McKenna, 2014). The items of interest were on 4-point Likert scales that ranged from *agree a lot* (1) to *disagree a lot* (4). Negatively worded questions were reverse coded.

Statistical Analyses

The TIMSS science attitudes items were designed to measure three distinct scales: Students Like Learning Science (SLS), Students Confident in Science (SCS), and Students Value Science (SVS) (Martin, Mullis, & Hooper, 2016). Thus, the current study investigated a latent structure and measurement invariance of TIMSS science attitudes items with the assumption of the three factors of SLS, SCS, and SVS.

Investigation of Factorial Structure in TIMSS Science Attitudes Items

The main analysis of this study was to compare 3-factor CFA and ESEM solutions. Because we were seeking to identify the best model for these attitudinal items, if the *a priori* confirmatory model was not the best-fitting, we also intended to test ESEM models with different numbers of factors to examine if these models could provide more appropriate representation of the factor structure. In traditional CFA models, each of the items was allowed to load on the latent factor that it was supposed to measure, and no cross-loadings were allowed. Results of the CFA were compared with those from ESEM, a new approach that integrates the benefits of CFA, SEM, and EFA (Marsh, Liem, Martin, Morin, & Nagengast, 2011). ESEM differs from CFA in its inclusion of cross-loadings for all items on all factors in a single step (Asparouhov & Muthén, 2009). The ESEM model was estimated with oblique target rotation according to the specification of Marsh et al. (2010). Target rotation was suitable for this study because there was a *priori* specification of the expected latent factor structure (Marsh et al. 2014). All analyses were conducted in Mplus 7.31 (Muthén & Muthén, 1998-2015) using the maximum likelihood robust (MLR) estimator that is robust to violation of multivariate normality, possibly induced by the Likert nature of the items (Morin, Arens, & Marsh, 2016).

Multiple fit indices were examined to evaluate the model fit of the CFA/ESEM models. Given the oversensitivity of the chi-square test of exact fit to a sufficiently large sample size (Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004), we also considered approximate fit indices that are less dependent on sample size: comparative fit index (CFI), Tucker-Lewis index (TLI), root-mean square error of approximation (RMSEA), and standardized root-mean square residual (SRMR). CFI and TLI values in the range of .90–.95 are acceptable, whereas those greater than .95 are indicative of an excellent fit (Bentler, 1990). RMSEA estimates smaller than .08 and .05 suggest, respectively, reasonable and good model fit (Brown & Cudeck, 1993). MacCallum et al. (1996) proposed that RMSEA with the value of .1 or above should be rejected. SRMR values equal to or below .08 support an adequate model fit (Hu & Bentler, 1999).

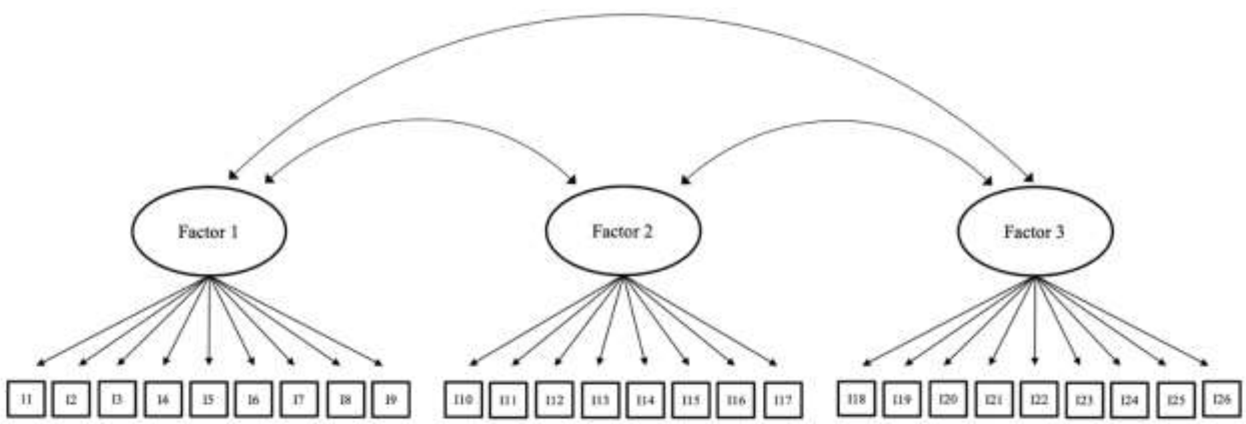

Figure 1. Path Diagram of ICM-CFA Model

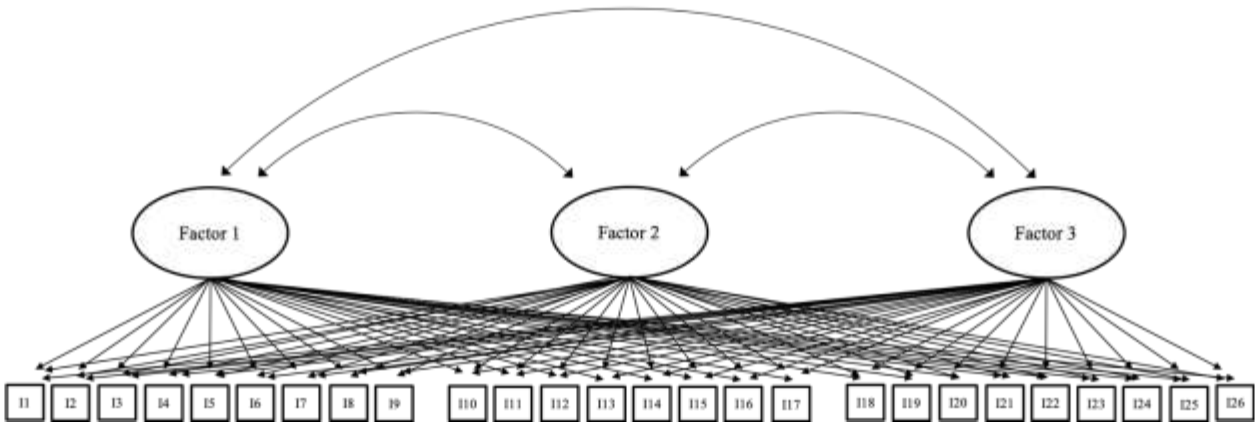

Figure 2. Path Diagram of ESEM Model

Measurement Invariance across Genders

After determining the model of choice, the equality of model parameters across student gender groups was investigated. We compared several nested models, testing progressively more restrictive models that specified measurement invariance across genders (Meredith, 1993; Samuel, South, & Griffin, 2015). We first conducted simultaneous analyses in which the measurement parameters in both groups were freely estimated (Levesque, Zuehlke, Stanek, & Ryan, 2004). The estimated parameters were factor loadings, error variances, and covariance among latent factors. The next model of configural invariance assumes an equivalent factor structure across groups, demonstrating that different groups conceptualize the underlying construct in the same way (Milfont & Fischer, 2010). Models of metric invariance imposed equality constraints on all factor loadings and cross-loadings across groups under the factor pattern identified previously. In a test of scalar invariance, factor structure, factor loadings, and item intercepts were set to be invariant across gender groups (Arens & Morin, 2016).

We relied on changes in goodness-of-fit indices to make a comparison between nested models; since models become progressively more constrained, each new invariance model is nested in the previous model Because, as like the chi-square test of exact fit, chi-square difference tests are sensitive to sample size, only a trivial difference may result in the rejection of the null model with large samples (Bollen, 1989; Brannick, 1995; Cheung & Rensvold, 2002; Tucker & Lewis, 1973). Chen (2007) provided guidelines that a difference in CFI (ΔCFI) less than .010 and difference in RMSEA less than .015 could be interpreted as support for measurement invariance across groups. Although these guidelines were based on maximum likelihood (ML) estimation, Sass, Schmitt, and Marsh (2014) found that the MLR scaling correction had comparatively little impact on the changes in approximate fit indices and resulted in similar values to those obtained by ML. In addition, the Bayesian information criterion (BIC) was used when comparing fit of nested and nonnested models. Models with lower BIC values are considered superior in terms of fit and parsimony (Kramer, Krueger, & Robert, 2008). If the BIC difference for two models is 0–2, 2–6, 6–10, or > 10, it respectively indicates weak, positive, strong, or very strong evidence for a preferred model (Raftery, 1995). We used both the absolute and relative fit indices to test nested invariance models across gender groups.

CHAPTER 4. RESULTS

Descriptive Statistics

Descriptive statistics for the TIMSS science attitudes items are provided in Table 1 for the total, female, and male samples. The possible total score ranged from 26 to 104, with a lower score indicating more positive attitudes toward science. Specifically, SLS and SVS total scores ranged from 9 to 36, and the SCS total score from 8 to 32, with a lower score corresponding to more favorable attitudes toward science. According to independent *t*-tests, statistically significant gender differences were found in SLS, SCS, and SVS scores ($p < .05$), with male students showing lower scores than female students on all three scales (i.e., males tend to report enjoying learning science more, being more confident in science, and valuing science more). SLS8, SCS6, and SVS9 showed particularly small total samples means of 1.64, 1.63, and 1.55, respectively. This reflects that the majority of students enjoy science experiments, think their parents value science highly, and agree that it is important to do well in science classes. On the other hand, SCS6 and SVS5 were very high in total sample means with 2.25 and 2.29, which represents that many students disagree that they are good at science, and they would like a job involving science.

Table 1. *Item Means and Standard Deviations for SLS, SCS, and SVS scales*

Items		Total		Female		Male	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
Students Like Learning Science (SLS)		17.11	6.77	17.65	6.91	16.56	6.57
1	I enjoy learning science	1.87	0.94	1.95	0.96	1.78	0.92
2	I wish I did not have to study science*	2.23	1.04	2.24	1.03	2.22	1.05
3	Science is boring*	2.14	1.03	2.17	1.03	2.10	1.03
4	I learn many interesting things in science	1.65	0.84	1.68	0.85	1.62	0.83
5	I like science	1.87	0.96	1.95	0.98	1.78	0.92
6	I look forward to learning science in school	2.02	0.99	2.11	1.02	1.93	0.96
7	Science teaches me how things in the world work	1.67	0.84	1.70	0.85	1.65	0.83
8	I like to conduct science experiments	1.64	0.89	1.68	0.92	1.61	0.87
9	Science is one of my favorite subjects	2.12	1.08	2.25	1.10	1.98	1.03
Students Confident in Science (SCS)		16.15	5.60	16.61	5.77	15.67	5.38
1	I usually do well in science	1.67	0.79	1.70	0.81	1.64	0.77
2	Science is more difficult for me than for many of my classmates*	1.95	0.94	1.96	0.94	1.95	0.95
3	Science is not one of my strength*	2.19	1.04	2.24	1.04	2.14	1.04
4	I learn things quickly in science	1.96	0.90	2.05	0.92	1.86	0.87
5	I am good at working out difficult science problems	2.21	0.97	2.35	0.98	2.06	0.94
6	My teacher tells me I am good at science	2.25	1.01	2.34	1.01	2.17	0.99
7	Science is harder for me than any other subject*	1.91	0.97	1.93	0.97	1.90	0.98
8	Science makes me confused*	2.10	1.00	2.13	1.00	2.06	1.00
Students Value Science (SVS)		16.86	6.71	17.02	6.70	16.69	6.72
1	I think learning science will help me in my daily life	1.89	0.92	1.92	0.93	1.85	0.92
2	I need science to learn other school subjects	2.16	0.99	2.23	0.99	2.10	0.99
3	I need to do well in science to get into the <university> of my choice	1.76	0.91	1.76	0.92	1.76	0.90
4	I need to do well in science to get the job I want	1.99	1.04	1.99	1.05	2.00	1.02
5	I would like a job that involves using science	2.29	1.11	2.35	1.14	2.22	1.08
6	It is important to learn about science to get ahead in the world	1.87	0.93	1.88	0.94	1.85	0.92
7	Learning science will give me more job opportunities when I am an adult	1.79	0.93	1.82	0.94	1.77	0.91

Table 1 continued

8	My parents think that it is important that I do well in science	1.63	0.84	1.61	0.84	1.66	0.84
9	It is important to do well in science	1.55	0.80	1.52	0.79	1.58	0.81

**Note.* Reverse-coded; SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science

Factor Structure of TIMSS 2015 Science Attitudes Items

The ESEM model seemed to be superior to the CFA model in terms of fit indices, parameter estimates, factor correlations, and model interpretability. As can be seen in Table 2, the ESEM model provided better fit indices than the CFA model. The CFA solution provided a reasonable fit to the data (CFI = .959, TLI = .955, RMSEA = .042, SRMR = .033), while the ESEM had a slightly better fit according to multiple fit indices (CFI = .971, TLI = .963, RMSEA = .038, SRMR = .018). The model fit differences were as follow: ΔCFI = +.012, ΔTLI = +.008, ΔRMSEA = –.004, and ΔSRMR = –.015. According to BIC, the significant decrease by 4,478 supported that the ESEM solution was superior to the CFA model. The chi-square difference test also supported that the ESEM approach was significantly better than the CFA approach ($p < .001$). This means that the CFA model in which all cross-loadings are constrained to be zero is too restrictive even though it is more parsimonious (Marsh, Liem, Martin, Morin, & Nagengast, 2011).

Next, a detailed evaluation of parameter estimates shows that both CFA and ESEM models yielded reasonable factor loadings. With respect to ESEM, the 26 target factor loadings were within a preferred range as shown in Table 3. The items' factor loadings ranged from .675 to .953 for SLS (M = .837), .407 to 1.004 for SCS (M = .741), and .736 to .925 for SVS (M = .838). The CFA model also provided acceptable factor loadings from .756 to .911 for SLS (M = .848), from .784 to .887 for SCS (M = .843), and from .832 to .922 for SVS (M = .886). Although the CFA provided slightly better factor loadings than the ESEM solution, there were only small differences and the factor loading patterns were very similar. Given both target and nontarget factors, the CFA and ESEM solutions showed a very similar pattern with a profile similarity index (PSI) of .956. The PSI was used to evaluate the similarity of the patterns of parameters, which was to examine whether the same item has a relatively high or low factor loading across models (Marsh et al, 2010). Hence, the PSI in this study is the correlation between 78 ESEM factor loadings and the corresponding CFA parameters. The PSI of .956 indicated the pattern of factor loadings for the two models was very similar.

However, a detailed observation of the cross-loadings supported the usefulness of the ESEM model. For instance, SLS2 (main loading = .675; cross-loading = .248) and SLS3 (main loading = .700; cross-loading = .202) showed substantial loadings on the SLS with displaying moderate cross-loadings on the SCS scale. Also, SCS1 (main-loading = .598; cross-loading = .222), SCS5 (main-loading = .541; cross-loading = .222), and SCS6 (main-loading = .407; cross-loading = .286) had noticeable cross-loadings on the SVS scale. These results suggest that there exist multiple meaningful cross-loadings, and such conceptual overlap should not be ignored as in the CFA solution. The consideration of cross-loadings is important because the omission of even a few small cross-loadings has possibly resulted in biased estimates of factor correlations (Morin et al, 2013; Schmitt & Sass, 2011).

The CFA and ESEM solutions differed in terms of factor correlations. The ESEM factor correlations were lower than those of CFA. Regarding the ESEM solution, the correlations of SLS/SCS, SLS/SVS, and SCS/SVS were .761, .731, and .762, respectively. The corresponding CFA factor correlations were .789, .742, and .823. This showed that the CFA factor correlations appeared to be larger than the ESEM correlations, demonstrating that the ESEM more clearly distinguished the factors. The clear differentiation of factors was more consistent with *a priori* specification of the Methods and Procedures in TIMSS 2015 (Martin, Mullis, & Hooper, 2016) as it stated that the TIMSS science attitudes items were designed to measure the three distinct factors of SLS, SVS, and SCS.

Table 2. *Goodness-of-fit indices for the CFA and ESEM models*

	χ^2	*df*	SCF	RMSEA	RMSEA 90% CI	CFI	TLI	SRMR	BIC
CFA	5,748	296	3.14	.042	.041 to .043	.959	.955	.033	683,978
ESEM	4,029	250	3.26	.038	.037 to .040	.971	.963	.018	679,500

Note. χ^2 = adjusted chi-square fit statistic with robust standard errors; *df* = degrees of freedom; SCF = Scale correction factor; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean residual; BIC = Bayesian information correction.

Table 3. *Standardized ESEM and CFA Factor Loadings for TIMSS Science Attitudes Items*

		ESEM Factor 1	ESEM Factor 2	ESEM Factor 3	CFA factor loadings
SLS	1	**0.953**	-0.011	-0.043	0.911
	2	**0.675**	0.248	-0.131	0.772
	3	**0.700**	0.202	-0.140	0.756
	4	**0.838**	-0.063	0.051	0.829
	5	**0.924**	-0.020	-0.026	0.889
	6	**0.947**	-0.046	-0.013	0.900
	7	**0.779**	-0.086	0.182	0.849
	8	**0.777**	-0.052	0.103	0.816
	9	**0.941**	0.012	-0.054	0.908
SCS	1	0.113	**0.598**	0.219	0.887
	2	-0.058	**0.968**	-0.058	0.848
	3	0.011	**0.893**	-0.069	0.828
	4	0.138	**0.562**	0.179	0.845
	5	0.126	**0.541**	0.222	0.850
	6	0.144	**0.407**	0.286	0.784
	7	-0.064	**1.004**	-0.080	0.854
	8	-0.031	**0.952**	-0.061	0.851
SVS	1	0.094	0.069	**0.747**	0.873
	2	0.068	-0.012	**0.821**	0.862
	3	-0.030	0.055	**0.884**	0.904
	4	-0.002	-0.020	**0.906**	0.887

Table 3 continued

5	0.108	0.019	**0.736**	0.832
6	0.005	0.001	**0.912**	0.915
7	-0.018	0.015	**0.925**	0.922
8	-0.062	0.143	**0.804**	0.871
9	-0.013	0.136	**0.806**	0.904

Note. The target loadings for each factor are in bold. ESEM = Exploratory Structural Equation Model; CFA = Confirmatory Factor Analysis; SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science.

The ESEM solution supported the intended interpretation of the TIMSS science attitudes items. First, the model yielded excellent model fit and reasonable factor loadings in accordance with the a *priori* factor structure. The Methods and Procedures in TIMSS 2015 (Martin, Mullis, & Hooper, 2016) mentioned that the instrument was constructed to assess three distinct factors of SLS, SCS, and SVS. The result with the ESEM model supported that these factors were clearly measured by nine, eight, and nine items, respectively. Second, the model appropriately differentiated and related the three factors. By covering and relating all the domains under attitudes toward science, the model provided evidence of content validity. In a similar vein, the CFA solution fit well, provided reasonable factor loadings according to *a priori* structure, and correlated the three factors. However, its highly correlated factors may raise a question of whether the factors are distinguishable. Indeed, the extremely high correlation of .823—the correlation between SCS and SVS—can be problematic, which was the result of a restrictive condition that nontarget factor loadings should be fixed to be zero (Marsh et al., 2011). The excessively high factor correlation between SCS and SVS may decrease the discriminant validity of factors and it makes it difficult to differentiate the 3-factor CFA model from a 2-factor CFA solution. Therefore, we concluded that the ESEM model provided more appropriate representation of the underlying factor structure of the TIMSS items than the CFA model did.

After determining that the ESEM solution as a model of choice, we explored 2-factor and 4-factor ESEM solutions to see if they could be alternative models. As more parsimonious models are preferred to complicated models (Hooper, Coughlan, & Mullen, 2008), it is meaningful to investigate a 2-factor ESEM solution. Also, a 4-factor solution was examined to investigate if there could be another possible factor affecting students' attitudes toward science. These models were tested using an oblique goemin rotation. The 3-ESEM model yielded a superior representation of the latent structure of the TIMSS science attitudes items than the 2-ESEM, and 4-ESEM. The 2-ESEM provided a modest model fit (CFI = .919, TLI = .904, RMSEA = .062, RMSEA = .062) and acceptable factor loadings. As seen in Table 4, the main factor loading for the first factor ranged from .714 to .918 (*M* = .404), and for the second factor from .394 to .931 (*M* = .453).

Nevertheless, the two factors were very highly correlated with a factor correlation of .717, which can be potentially problematic. Also, while the first factor was the same as the SLS in the 3-ESEM model, the second factor was the simple combination of the SCS and SVS in the 3-ESEM. The second factor is not well defined since SCS and SVS are distinct factors according to the Methods and Procedures in TIMSS 2015 (Martin, Mullis, & Hooper, 2016); SCS and SVS are designed to measure students' academic self-concept and extrinsic motivation, respectively. Although the SCS and SVS were highly related with the factor correlation of .762, it is not reasonable to integrate these two distinguishable factors into a single construct. Moreover, multiple items mainly loading on the second factor demonstrated substantial cross-loadings on the first factor. The following items did not show much difference between main loadings and cross-loadings: SCS3 (main loading = .394; cross-loading .390), SCS7 (main loading = .436; cross-loading = .364), and SCS8 (main loading = .430; cross-loading = .373). These items displayed relatively low loading on their primary factor while showing considerable cross-loadings on the nontarget factors. These findings call into the question whether the second factor in the 2-ESEM can be justified.

In the case of the 4-factor ESEM model (4-ESEM), it showed problems with regard to its fourth factor and the factor correlation. The model yielded the best fit indices (CFI = .982, TLI = .974, RMSEA = .032, SRMR = .013) and its factor loadings were within preferred ranged as can be seen in Table 5. The items' factor loadings for the first factor ranged from .701 to .874 (*M* = .785), for the second factor from .635 to .888 (*M* = .729), for the third factor from .713 to .963 (*M* = .846), and for the fourth factor from .414 to .481 (*M* = .448). The factor correlations ranged from .238 to .809 with a mean of .502.

However, there were no items that showed primary factor loadings on the fourth factor. Although the four items showed considerable cross-loadings on the fourth factor but their primary factor was the second factor: SCS2 (main loading = .660; cross-loading = .421), SCS3 (main loading = .666; cross-loading = .354), SCS7 (main loading = .635, cross-loading = .481), and SCS8 (main loading = .658; cross-loading = .414). These items commonly measured the degree to which students feel difficulty in learning science, but it is unreasonable to define a latent construct simply based on cross-loadings. Considering that ideally from a minimum 2-3 up to a maximum of 4–6 indicators are required to provide minimum coverage of the construct's theoretical domain (Hair, Black, Babin, & Anderson, 2010), it is difficult to define the fourth factor as a distinct latent construct. Next, a factor correlation between the second and fourth factor was .809, which reflects that these two factors are excessively correlated and thereby it is difficult to distinguish between the two. In other words, such high factor correlation undermines discriminant validity, which ensures that a construct is empirically unique and reflects phenomena of interest that other measures do not represent (Hair et al., 2010). Discriminant validity is a necessary condition for establishing a construct validity (Watson, Weber, Assenheimer, Clark, Strauss, & McCormick, 1995) and for examining the relationship between latent variables (Henseler, Ringle, & Sarstedt, 2015). In this vein, the extremely high factor correlation is not desirable. These findings corroborate the superiority of 3-ESEM, which supported three main factors of SLS, SCS, and SVS. Hence, we concluded that the 3-ESEM should be the model of choice among the CFA, 2-ESEM, and 4-ESEM solutions.

Table 4. *Standardized 2-ESEM Factor Loadings for TIMSS Science Attitudes Items*

		ESEM Factor 1	ESEM Factor 2
SLS	1	0.918	-0.013
	2	0.763	0.020
	3	0.769	-0.012
	4	0.783	0.051
	5	0.886	-0.001
	6	0.896	0.000
	7	0.714	0.167
	8	0.729	0.106
	9	0.917	-0.013
SCS	1	0.363	0.533
	2	0.354	0.440
	3	0.390	0.394
	4	0.372	0.476
	5	0.351	0.507
	6	0.312	0.503
	7	0.364	0.436
	8	0.373	0.430
SVS	1	0.113	0.788
	2	0.053	0.818
	3	-0.015	0.912
	4	-0.018	0.893
	5	0.106	0.749
	6	-0.003	0.911
	7	-0.019	0.931
	8	-0.011	0.881

Table 4 continued

9	0.034	0.880

Note. SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science.

Table 5. *Standardized 4-ESEM Factor Loadings for TIMSS Science Attitudes Items*

		ESEM Factor 1	ESEM Factor 2	ESEM Factor 3	ESEM Factor 4
SLS	1	0.874	0.057	-0.010	0.008
	2	0.701	-0.002	0.014	0.246
	3	0.719	-0.023	-0.002	0.222
	4	0.778	-0.026	0.096	0.006
	5	0.846	0.055	0.004	0.000
	6	0.856	0.075	-0.003	-0.034
	7	0.729	-0.070	0.235	0.006
	8	0.712	0.013	0.130	-0.012
	9	0.849	0.129	-0.047	-0.015
SCS	1	0.037	0.748	0.111	0.070
	2	-0.007	0.660	0.018	0.421
	3	0.039	0.666	-0.023	0.354
	4	0.017	0.865	0.000	-0.019
	5	-0.008	0.888	0.023	-0.050
	6	0.029	0.714	0.119	-0.070
	7	0.001	0.635	0.020	0.481
	8	0.015	0.658	0.009	0.414
SVS	1	0.073	0.115	0.728	-0.018
	2	0.046	0.056	0.796	-0.055
	3	-0.021	0.025	0.895	0.014
	4	-0.001	-0.015	0.907	-0.024
	5	0.083	0.082	0.713	-0.040
	6	0.016	-0.041	0.937	0.008
	7	0.001	-0.055	0.963	0.026
	8	-0.040	0.058	0.836	0.068
	9	0.004	0.055	0.840	0.066

Note. SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science.

Measurement Invariance across Genders

Having determined that the 3-factor ESEM model was best-fitting, we examined the model fit for each gender group (see Table 6). These ESEM models without any equality constraints fit well for both female and male students. Next, the nested invariance models (configural, metric, and scalar) were tested according to the guidelines set by Meredith (1993), with some modifications for ESEM models (Asparouhov & Muthén, 2009).

Table 6. *Goodness-of-Fit Indices for the Baseline Model across Genders*

	χ^2	*df*	SCF	RMSEA	RMSEA 90% CI	CFI	TLI	SRMR	BIC
Female	2,426	250	2.605	.041	.040 to .043	.948	.933	.023	332,492
Male	2,049	250	3.442	.038	.036 to .039	.965	.954	.019	340,711

Note. χ^2 = adjusted chi-square fit statistic with robust standard errors; *df* = degrees of freedom; SCF = Scale correction factor; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean residual; BIC = Bayesian information correction.

The first invariance model, configural invariance (M2), assumes the same factor structure, and similar association of indicators to factors, across groups. When testing invariance with ESEM models, the configural model imposes equality restrictions on the number of factors and pattern of factor loadings, and allows all cross-loadings; this model was constructed to measure three factors of SLS, SCS, and SVS, with nine, eight, and nine items loading on each factor, respectively. Factor means were fixed to 0, while factor variances were set to 1.0 across gender groups. As can be seen in Table 10, the configural model displayed an acceptable fit (χ^2 = 4,422, RMSEA = .039, CFI = .958, TLI = .946, SRMR = .021). Also, the chi-square difference between the baseline model (M1) and M2 was insignificant ($\Delta\chi^2$ (250) = 80, p > .999), and an increase in fit statistics due to model parsimony was observed (ΔRMSEA = +.001, ΔCFI = –.013, ΔTLI = –.017). The pattern of factor loadings for each gender group is shown in Table 7. The pattern of M1 and M2 models were very similar with PSI of .991, .984, and .991 for the SLS, SCS, and SVS scales, respectively. Altogether, we concluded that configural invariance of the TIMSS science attitude items is tenable across the female and male student groups.

Table 7. *Unstandardized Factor Loadings for Configural Invariance Model*

		SLS		SCS		SVS	
		Female	Male	Female	Male	Female	Male
SLS	1	1.053	1.334	0.021	-0.049	-0.061	-0.042
	2	0.830	0.943	0.199	0.431	-0.141	-0.233
	3	0.907	1.068	0.157	0.402	-0.142	-0.276
	4	0.927	1.266	-0.063	-0.116	0.064	0.052
	5	1.119	1.418	0.012	-0.048	-0.013	-0.038
	6	1.150	1.388	-0.032	-0.078	-0.012	0.002
	7	0.782	1.189	-0.133	-0.092	0.264	0.165
	8	0.830	1.188	-0.076	-0.078	0.118	0.12
	9	1.166	1.410	0.05	-0.005	-0.081	-0.03
SCS	1	0.065	0.195	0.748	0.741	0.144	0.411
	2	-0.074	-0.073	1.129	1.473	-0.083	-0.088
	3	0.032	0.016	1.108	1.421	-0.085	-0.094
	4	0.132	0.244	0.787	0.798	0.153	0.392
	5	0.120	0.198	0.717	0.755	0.211	0.434
	6	0.114	0.268	0.577	0.568	0.326	0.501
	7	-0.081	-0.078	1.196	1.562	-0.091	-0.137
	8	-0.037	-0.030	1.144	1.467	-0.077	-0.086
SVS	1	0.124	0.127	0.065	0.121	0.846	1.183
	2	0.094	0.084	-0.001	-0.029	0.933	1.343
	3	-0.056	-0.023	0.04	0.121	1.079	1.374
	4	-0.015	0.020	-0.028	-0.015	1.166	1.465
	5	0.194	0.136	0.028	0.048	0.95	1.255
	6	0.037	-0.014	-0.022	0.038	1.108	1.441
	7	-0.029	-0.013	-0.012	0.072	1.148	1.435
	8	-0.100	-0.073	0.134	0.244	0.921	1.249
	9	-0.045	0.013	0.151	0.219	0.929	1.244

After finding that configural invariance holds, we tested metric invariance (M3) across genders. This invariance imposes equality constraints on the factor structure and factor loading values for each item. Factor means were set to 0 in both groups, while factor variances were respectively set to 1.0 and free in female and male. The model fit of metric invariance models was excellent according to various fit indices: χ^2 = 4,796, RMSEA = .038, CFI = .955, TLI = .949, SRMR = .025. The changes in approximate fit indices across groups (ΔRMSEA = –.001, ΔCFI = –.003, ΔTLI = +.003) were strong evidence in support of the metric invariance. Also, the considerable decrease in BIC (11) clearly supported the metric invariance of the ESEM model. Hence, there is invariance of the factor loadings across genders.

Table 8. *Unstandardized Factor Loadings for Metric Invariance Model*

		SLS	SCS	SVS
SLS	1	1.061	-0.013	-0.045
	2	0.785	0.269	-0.153
	3	0.871	0.242	-0.171
	4	0.977	-0.084	0.055
	5	1.129	-0.021	-0.023
	6	1.126	-0.051	-0.005
	7	0.880	-0.106	0.193
	8	0.899	-0.072	0.113
	9	1.143	0.021	-0.051
SCS	1	0.113	0.659	0.227
	2	-0.073	1.140	-0.078
	3	0.015	1.109	-0.086
	4	0.166	0.696	0.225
	5	0.139	0.648	0.266
	6	0.168	0.500	0.348
	7	-0.078	1.207	-0.104
	8	-0.037	1.145	-0.077
SVS	1	0.101	0.077	0.875
	2	0.070	-0.020	0.980
	3	-0.044	0.059	1.059
	4	-0.005	-0.032	1.130
	5	0.138	0.026	0.941
	6	0.002	-0.006	1.097
	7	-0.029	0.011	1.116
	8	-0.086	0.158	0.941

Table 8 continued

	9	-0.024	0.152	0.940

Note. SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science.

Then we turned to the final type of invariance, scalar invariance (M4). This invariance model presumes that groups have the same factor structure, the same pattern of factor loadings, and equal item intercepts. The scalar invariance model produced excellent fit (χ^2 = 5,126, RMSEA = .039, CFI = .952, TLI = .947, SRMR = .026). Although BIC slightly increased, the changes in fit indices were within the acceptable range (ΔCFI = –.003, ΔTLI = –.002, ΔRMSEA = +.001). Based on these results, we can conclude that TIMSS science attitudes items allow the comparison of group means on factors, indicators, and the total score.

Table 9. *Unstandardized Factor Loadings for Scalar Invariance Model*

		SLS	SCS	SVS
SLS	1	1.062	-0.012	-0.046
	2	0.776	0.272	-0.146
	3	0.865	0.245	-0.167
	4	0.972	-0.082	0.058
	5	1.129	-0.02	-0.024
	6	1.128	-0.05	-0.008
	7	0.874	-0.103	0.197
	8	0.895	-0.07	0.116
	9	1.148	0.022	-0.057
SCS	1	0.112	0.657	0.229
	2	-0.078	1.14	-0.073
	3	0.016	1.108	-0.086
	4	0.174	0.692	0.22
	5	0.152	0.643	0.256
	6	0.174	0.497	0.344
	7	-0.082	1.208	-0.1
	8	-0.039	1.145	-0.075
SVS	1	0.103	0.075	0.875
	2	0.075	-0.022	0.976
	3	-0.046	0.058	1.061
	4	-0.007	-0.033	1.132

Table 9 continued

5	0.143	0.023	0.937
6	0.001	-0.007	1.099
7	-0.03	0.01	1.117
8	-0.093	0.157	0.947
9	-0.031	0.152	0.946

Note. SLS = Students Like Learning Science; SCS = Students Confident in Science; SVS = Students Value Science.

Table 10. *Goodness-of-Fit Indices for Measurement Invariance across Genders*

	χ^2	*df*	SCF	RMSEA	RMSEA 90% CI	CFI	TLI	SRMR	BIC	*p*	ΔCFI	ΔRMSEA
M1	4,029	250	3.264	.038	.037 to .040	.971	.963	.018	679,500			
M2	4,422	500	3.024	.039	.038 to .040	.958	.946	.021	673,378	0	–.013	+.001
M3	4,796	569	2.918	.038	.037 to .039	.955	.949	.025	673,367	0	–.003	–.001
M4	5,126	592	2.848	.039	.038 to .040	.952	.947	.026	673,759	0	–.003	+.001

Note. χ^2 = adjusted chi-square fit statistic with robust standard errors; *df* = degrees of freedom; SCF = Scale correction factor; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean residual; BIC = Bayesian information correction; M1 = baseline model (no invariance imposed); M2 = configural invariance; M3 = metric invariance; M4 = scalar invariance.

CHAPTER 5. DISCUSSION

The definition of attitudes toward science is inconsistent—with unidimensional and multidimensional approache —and thereby, science attitudes instruments have been criticized for their weakness in supporting validity (Wang & Berlin, 2010). In particular, the TIMSS students' attitudes toward science items have been used widely for studying attitudes, but there exists only limited validity evidence for these items. Justifying the validity of the TIMSS items is important not only because it contributes to identifying whether the items measure what they are supposed to measure, but it also enables researchers to test whether the items assess the same trait for both female and male students. The purpose of this study was to examine (1) the underlying factor structure of the TIMSS science attitudes items and (2) measurement equivalence across gender groups. The results show that the ESEM model with three factors of SLS, SCS, and SVS provided more appropriate representation of the factor structure of the TIMSS science attitudes scales, and the items measured the same trait for both female and male students. The implication of these results from both practical and methodological perspectives are addressed in this chapter.

Students' Attitudes toward Science with TIMSS Items

Results from the present study support that the TIMSS attitudes items consist of three distinct factors of SLS, SCS and SVS, and are well designed by the *priori* specification, which indicates that the evidence supports the intended interpretations of total scores (subscores) from the TIMSS science attitudes scale. We tested and found that the TIMSS science attitudes items assess three latent factors of SLS, SCS, and SVS, which relate to intrinsic motivation, academic self-concept, and extrinsic motivation, respectively, in the science domains. Given that there has been no consensus on the theoretical conceptualization of attitudes toward science, this study contributes to the science education literature by providing support for the multidimensional approach to measuring science attitudes. Specifically, the moderately high positive factor correlations show that these factors are associated with, but distinguished from, one another with pairwise correlations of about .75. The result is consistent with previous studies that suggest motivation and academic self-concept are closely related (Bong & Clarks, 1999; Green, Nelson, Martin, & Marsh, 2006; Pajares & Schunk, 2001).

Based on these results, future researchers can perform more in-depth analyses. For example, researchers can examine the degree of each factor's contribution to predicting science attitudes with the use of multiple regression. Also, science educators could employ qualitative approaches such as proposing a multidimensional definition of attitudes toward science based on the three factors (SLS, SVS, and SCS) or investigating the property of conceptual overlap between factors. These future studies will lead to deeper knowledge in the underlying factors, and they should ultimately allow researchers and science educators to achieve better insight into the differences in attitudes toward science: Why do some students have more favorable science attitudes than others? What factors cause such differences between students? How could we improve each student's enjoyment in learning science (SLS), confidence in learning science (SCS), value in science (SVS), and attitudes toward science in the end?

Next, this study reveals that there exists measurement invariance of the TIMSS science attitudes items across genders. The dimensional, configural, metric, and scalar invariance models were all tenable because each model fit the data very well according to the multiple fit indices. The TIMSS science attitudes items have the same factor structure for both female and male students (configural invariance), there were no gender differences in the factor loadings (metric invariance), and the item intercepts were not significantly different across genders (scalar invariance). These results indicate that the items measure the same latent construct—students' attitudes toward science—across genders. Thus, any observed discrepancy in science attitudes may reflect true gender differences rather than measurement errors or item bias. In other words, the TIMSS science attitudes items can be used safely when inspecting the effect of gender on issues related to science attitudes.

The Flexibility of ESEM over CFA

As expected, this research shows the flexibility of ESEM over CFA by illustrating that ESEM provides superior representation of the factor structure. The 3-ESEM solution was determined to be the model of choice for the TIMSS 2015 science attitudes items in terms of model fit, factor loadings, factor correlations, and model interpretability. One potential problem with ICM-CFA model lies in its exclusion of nonzero cross-loadings. The ignorance of meaningful cross-loadings could not only result in a poor model fit but also might distort the observed relations among the factors (Marsh et al., 2009). Although Stromeyer, Miller, Sriramachandramurthy, and DeMartino (2015) maintained that modeling cross-loadings is similar to modeling "noise," small cross-loadings allow the constructs to be estimated based on all of the relevant information present at the indicator level rather than tainting the constructs (Asparouhov, Muthén, & Morin, 2015). Marsh and colleagues (2009) also mentioned that when many small cross-loadings are restricted to be zero, the only way of accounting for these cross-loadings in the estimation results is by inflating the factor correlations.

In the case of the TIMSS science attitudes items, there exist many small cross-loadings across factors, and the five items (SLS2–*I wish I did not have to study science*; SLS3–*Science is boring*; SCS1–*I usually do well in science*; SCS5–*I am good at working out difficult science problems*; and SCS6–*My teacher tells me I am good at science*) were with noticeable cross-loadings on the nontarget factors. These minor cross-loadings could not be captured within the ICM-CFA framework, and thereby biased factor correlations may have resulted. This result is consistent with previous studies that maintained that the assumption of ICM-CFA may be too restrictive for many multidimensional and complex instruments (Arens & Morin, 2016; Marsh et al, 2009). Although CFA solutions can allow a limited number of possible cross-loadings, such inclusion is only possible with *a priori* latent structure (Arens & Morin, 2016). Furthermore, since SLS, SCS, and SVS are closely related at the theoretical level as mentioned previously, many possible cross-loadings are expected for the items. Thus, it is recommended to investigate the latent structure of the TIMSS science attitudes items with an ESEM approach.

Limitations and Future Study

One limitation of this study is that the TIMSS science attitudes items may need to be tested with additional, various data sets. The current study is only based on the US eighth grade students, so it is recommended to investigate a latent structure of the items and measurement invariance across genders using different countries or other grades of students to substantiate the present results. Relying on the guidelines used in CFA (Marsh et al, 2009) for examining model fit is another limitation of this study. As Joshanloo and Lamers (2016) stated, more detailed guidelines for the evaluation of model fit in ESEM is required as more parameters are estimated in ESEM solutions. Future studies should be encouraged to test the structural validity of the TIMSS science attitudes items with samples of students from other cultures such as Asia or Europe in order to generalize the results. Additionally, since the items were targeted for only eighth grade students, it will be instructive to test if any items may need to be modified or removed depending on students' grade level. Despite the limitations, this study contributes to revealing more appropriate representation of the underlying structure of the TIMSS science attitudes items and allows researchers to use the items safely when analyzing attitudes toward science and examining gender differences in attitudes toward science.

REFERENCES

Adams, W. K., Perkins, K. K., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2005). The design and validation of the Colorado Learning Attitudes about Science Survey. *AIP Conference Proceedings*, *790*(1), 45–48. https://doi.org/10.1063/1.2084697

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (AERA, APA, & NCME). (2014). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Anwer, M., Iqbal, H. M., & Harrison, C. (2012). Students' attitude towards science: A case of Pakistan. *Pakistan Journal of Social and Clinical Psychology*, *9*(2), 3–9.

Arens, A. K., & Morin, A. J. (2016). Relations between teachers' emotional exhaustion and students' educational outcomes. *Journal of Educational Psychology, 108*(6), 800–813. http://dx.doi.org/10.1037/edu0000105

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438.

Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*, 1561–1577.

Atkinson, R. (2013). A short and long-term solution to America's STEM crisis. *The Hill*. Retrieved from http://thehill.com/blogs/congress-blog/technology/287435-a-short-andlong-term-solution-to-americas- stem-crisis

Averett, C., Ferraro, D., Tang, J., Erberber, E., & Stearns P. (2017). *Trends in International Mathematics and Science Study (TIMSS), U.S. TIMSS 2015 and TIMSS Advanced 1995 & 2015 Technical Report and User's Guide* (NCES 2018-020). U.S. Department of Education. Washington, DC: National Center for Education Statistics, Institute of Education Sciences. Retrieved April 1, 2017, from https://nces.ed.gov/pubsearch

Aydeniz, M., & Kotowski, M. R. (2014). Conceptual and methodological issues in the measurement of attitudes towards science. *Electronic Journal of Science Education*, *18*(3), 1–24. Retrieved from http://ejse.southwestern.edu

Baker, D., & Leary, R. (2003). Letting Girls Speak Out about Science. *Journal of Research in Science Teaching*, 40, (Suppl.) S176–S200.

Banning, J., & Folkestad, J. E. (2012). STEM education related dissertation abstracts: A bounded qualitative meta-study. *Journal of Science Education and Technology*, *21*(6), 730–741.

Barmby, P., Kind, P. M., & Jones, K. (2008). Examining changing attitudes in secondary school science. *International Journal of Science Education*, *30*(8), 1075–1093.

Becker, K., & Park, K. (2011). Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: A preliminary meta-analysis. *Journal of STEM Education: Innovations & Research*, *12*(5/6), 23–37.

Bennett, J. (2001). The development and use of an instrument to assess students' attitude to the study of chemistry. *International Journal of Science Education*, *23*(8), 833–845.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246.

Blosser, P. E. (1984). *Attitude research in science education*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics and Environmental Education.

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303–316.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really?. *Educational psychology review*, *15*(1), 1–40.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*(3), 201–213.

Brotman, J. S., & Moore, F. M. (2008). Girls and science: A review of four themes in the science education literature. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *45*(9), 971–1002.

Brown, M., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150.

Buck, G., Cook, K., Quigley, C., Eastwood, J., & Lucas, Y. (2009). Profiles of urban, low SES, African American girls' attitudes toward science: A sequential explanatory mixed methods study. *Journal of Mixed Methods Research*, *3*(4), 386–410.

Camargo, S. L., Herrera, A. N., & Traynor, A. (2018). Looking for a consensus in the discussion about the concept of validity. *Methodology*, *14*(4), 146–155.

Cannon Jr, R. K., & Simpson, R. D. (1985). Relationships among attitude, motivation, and achievement of ability grouped, seventh- grade, life science students. *Science Education*, *69*(2), 121–138.

Caro, D. H., Sandoval-Hernández, A., & Lüdtke, O. (2014). Cultural, social, and economic capital constructs in international assessments: An evaluation using exploratory structural equation modeling. *School Effectiveness and School Improvement*, *25*(3), 433–450.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Cheung, D. (2009). Developing a scale to measure students' attitudes toward chemistry lessons. *International Journal of Science Education*, *31*(16), 2185–2203.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Christidou, V. (2011). Interest, attitudes and images related to science: Combining students' voices with the voices of school science, teachers, and popular science. *International Journal of Environmental and Science Education*, *6*(2), 141–159.

Blickenstaff, J. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education*, *17*(4), 369–386.

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology*, *106*(6), 1691–1730.

Coulson, R. (1992). Development of an instrument for measuring attitudes of early childhood educators towards science. *Research in Science Education*, *22*(1), 101–105.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134.

Dhindsa, H. S., & Chung, G. (2003). Attitudes and achievement of Bruneian science students. *International Journal of Science Education*, *25*(8), 907–922.

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, *23*(3), 225–241.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299.

Fishbein, M. E. (1967). *Readings in attitude theory and measurement*, New York, NY: John Wiley & Sons.

Foy, P. (2017). *TIMSS 2015 User Guide for the International Database*. International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Fraser, B. J. (1981). *Test of Science-Related Attitudes (TOSRA)*. Melbourne, Victoria: Australian Council for Educational Research.

George, R. (2006). A cross- domain analysis of change in students' attitudes toward science and attitudes about the utility of science. *International Journal of Science Education*, *28*(6), 571–589.

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching*, *25*(8), 689–703.

Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, *82*(3), 525–538.

Green, J., Nelson, G., Martin, A. J., & Marsh, H. (2006). The causal ordering of self-concept and academic motivation and its effect on academic achievement. *International Education Journal*, *7*(4), 534–546.

Guay, F., Morin, A. J., Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education*, *83*(1), 51–82.

Guzey, S. S., Harwell, M., & Moore, T. (2014). Development of an instrument to assess attitudes toward science, technology, engineering, and mathematics (STEM). *School Science and Mathematics*, *114*(6), 271–279.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). New Jersey, NJ: Prentice Hall.

Hassan, G. (2008). Attitudes toward science among Australian tertiary and secondary school students. *Research in Science & Technological Education*, *26*(2), 129–147.

Hayes, B. C., & Tariq, V. N. (2000). Gender differences in scientific knowledge and attitudes toward science: a comparative study of four Anglo-American nations. *Public Understanding of Science, 9*(4), 433–447. https://doi.org/10.1088/0963-6625/9/4/306

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, *43*(1), 115–135.

Hong, Z. R., & Lin, H. S. (2011). An investigation of students' personality traits and attitudes toward science. *International Journal of Science Education*, *33*(7), 1001–1028.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electrical Journal of Business Research Methods, 6*(1), 53–60.

Hooper, M., Mullis, I. V., & Martin, M. O. (2015). Chapter 3: TIMSS 2015 context questionnaire framework. In I.V.S. Mullis & M.O. Martin (Eds.), *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Hossain, M., & Robinson, M. (2012). How to motivate us students to pursue STEM (science, technology, engineering and mathematics) careers. *US-China Education Review, 4*(A), 442–451.

House, J. D., & Telese, J. A. (2008). Relationships between student and instructional factors and algebra achievement of students in the United States and Japan: An analysis of TIMSS 2003 data. *Educational Research and Evaluation*, *14*(1), 101–112.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Joshanloo, M., & Lamers, S. M. (2016). Reinvestigation of the factor structure of the MHC-SF in the Netherlands: Contributions of exploratory structural equation modeling. *Personality and Individual Differences*, *97*, 8–12.

Kadijevich, D. (2006). Achieving educational technology standards: the relationship between student teacher's interest and institutional support offered. *Journal of Computer Assisted Learning*, *22*(6), 437–443.

Kadijevich, D., & Haapasalo, L. (2008). Factors that influence student teacher's interest to achieve educational technology standards. *Computers & Education*, *50*(1), 262–270.

Khine, M. S., Al-Mutawah, M., & Afari, E. (2015). Determinants of affective factors in mathematics achievement: Structural equation modeling approach. *Journal of Studies in Education*, *5*(2), 199–211.

Khine, M. S., Al-Mutawah, M., & Afari, E. (2015). Determinants of affective factors in mathematics achievement: Structural equation modeling approach. *Journal of Studies in Education*, *5*(2), 199–211.

Koballa Jr, T. R., & Crawley, F. E. (1985). The influence of attitude on science teaching and learning. *School Science and Mathematics*, *85*(3), 222–232.

Kramer, M. D., Krueger, R. F., & Hicks, B. M. (2008). The role of internalizing and externalizing liability factors in accounting for gender differences in the prevalence of common psychopathological syndromes. *Psychological Medicine*, *38*(1), 51–61.

Lang, Q. C., Wong, A. F., & Fraser, B. J. (2005). Teacher-student interaction and gifted students' attitudes toward chemistry in laboratory classrooms in Singapore. *Journal of Classroom Interaction*, *40*, 18–28.

Langdon, D., McKittrick, G., Beede, D., Khan, B., & Doms, M. (2011). *STEM: Good jobs now and for the future*. Washington, DC: U.S. Department of Commerce Economics and Statistics Administration. Retrieved from http://www.esa.doc.gov/sites/default/files/reports/documents/stemfinalyjuly14_1.pdf

Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Educational Psychology*, *96*(1), 68–84.

Lovelace, M., & Brickman, P. (2013). Best practices for measuring students' attitudes toward learning science. *CBE—Life Sciences Education*, *12*(4), 606–617.

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.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among US students. *Science Education*, *95*(5), 877–907.

Maple, S. A., & Stage, F. K. (1991). Influences on the choice of math/science major by gender and ethnicity. *American Educational Research Journal*, *28*(1), 37–60.

Marsh, H. W., Hau, K. T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. McCardle (Eds.), *Psychometrics: A festschrift to Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*(3), 320–341.

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation modeling (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, *29*(4), 322–346.

Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85–110.

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 439–476.

Martin, M. O., & Preuschoff, C. (2007). Creating the TIMSS 2007 background indices. In J. F. Olson, M. O. Martin I. V. S. Mullis (Eds.), *TIMSS 2007 technical report* (pp. 281–338). Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Martin, M. O., Mullis, I. V. S., & Hooper, M. (Eds.). (2016). *Methods and Procedures in TIMSS 2015*. Retrieved from Boston College, TIMSS & PIRLS International Study Retrieved from http://timssandpirls.bc.edu/publications/timss/2015-methods.html

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111–130.

Moore, R. W., & Sutman, F. X. (1970). The development, field test and validation of an inventory of scientific attitudes. *Journal of Research in Science Teaching*, *7*(2), 85–94.

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116–139.

Morin, A. J., Arens, A. K., Tran, A., & Caci, H. (2016). Exploring sources of construct- relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research*, *25*(4), 277–288.

Mueller, D. J. (1986). *Measuring social attitudes: A handbook for researchers and practitioners*. New York, NY: Teachers College Press.

Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Retrieved from http://timssandpirls.bc.edu/timss2019/frameworks/

Mullis, I.V.S. & Martin, M.O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks.* Retrieved from http://timssandpirls.bc.edu/timss2015/frameworks.html

National Science Board. (2018). *Science and Engineering Indicators 2018* (NSB-2018-1). Alexandria, VA: National Science Foundation. Retrieved from https://www.nsf.gov/statistics/indicators/

Newton, P. E., & Shaw, S. D. (2013). Standards for talking and thinking about validity. *Psychological Methods*, *18*(3), 301–319.

Nieswandt, M. (2007). Student affect and conceptual understanding in learning chemistry. *Journal of Research in Science Teaching*, *44*(7), 908–937.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Ozel, M., Caglak, S., & Erdogan, M. (2013). Are affective factors a good predictor of science achievement? Examining the role of affective factors based on PISA 2006. *Learning and Individual Differences*, *24*, 73–82.

Papanastasiou, C., & Papanastasiou, E. C. (2004). Major influences on attitudes toward science. *Educational Research and Evaluation*, *10*(3), 239–257.

Papanastasiou, E. C., & Zembylas, M. (2002). The effect of attitudes on science achievement: A study conducted among high school pupils in Cyprus. *International Review of Education*, *48*(6), 469–484.

Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. *Perception*, 11, 239–266.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Reynolds, A. J., & Walberg, H. J. (1992). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology*, *84*(3), 371–382.

Rice, L., Barth, J. M., Guadagno, R. E., Smith, G. P., & McCallum, D. M. (2013). The role of social support in students' perceived abilities and attitudes toward math and science. *Journal of Youth and Adolescence*, *42*(7), 1028–1040.

Riegle-Crumb, C., Moore, C., & Ramos- Wada, A. (2011). Who wants to have a career in science or math? Exploring adolescents' future aspirations by gender and race/ethnicity. *Science Education*, *95*(3), 458–476.

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education*, *96*(3), 411–427.

Samuel, D. B., South, S. C., & Griffin, S. A. (2015). Factorial invariance of the five-factor model rating form across gender. *Assessment*, *22*(1), 65–75.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, *21*(2), 167–180.

Scantlebury, K., Baker, D., Sugi, A., Yoshida, A., & Uysal, S. (2007). Avoiding the issue of gender in Japanese science education. *International Journal of Science and Mathematics Education*, *5*(3), 415–438.

Scantlebury, K., Tal, T., & Rahm, J. (2007). "That don't look like me." Stereotypic images of science: Where do they come from and what can we do with them? *Cultural Studies of Science Education*, *1*(3), 545–558.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321.

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, *71*(1), 95–113.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research*, *95*(6), 323–332.

Smith, T. J., Pasero, S. L., & McKenna, C. M. (2014). Gender effects on student attitude toward science. *Bulletin of Science, Technology & Society*, *34*(1–2), 7–12.

Stiles, D. A., Adkisson, J. L., Sebben, D., & Tamashiro, R. (2008). Pictures of hearts and daggers: Strong emotions are expressed in young adolescents' drawings of their attitudes towards mathematics. *World Cultures eJournal*, *16*(2). Retrieved from http://escholarship.org/uc/item/1sq263b7 Google Schola

Stromeyer, W. R., Miller, J. W., Sriramachandramurthy, R., & DeMartino, R. (2015). The prowess and pitfalls of Bayesian structural equation modeling: Important considerations for management research. *Journal of Management*, *41*(2), 491–520.

Thurstone, L. L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology*, *26*(3), 249–269.

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

Wang, J., & Staver, J. R. (2001). Examining relationships between factors of science education and student career aspiration. *The Journal of Educational Research*, *94*(5), 312–319.

Wang, T. L., & Berlin, D. (2010). Construction and validation of an instrument to measure Taiwanese elementary students' attitudes toward their science class. *International Journal of Science Education*, *32*(18), 2413–2428.

Wang, X. (2013). Modeling entrance into STEM fields of study among students beginning at community colleges and four-year institutions. *Research in Higher Education*, *54*(6), 664–692.

Ware, N. C., & Lee, V. E. (1988). Sex differences in choice of college science majors. *American Educational Research Journal*, *25*(4), 593–614.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*(1), 15–25.

Weinburgh, M. (1995). Gender differences in student attitudes toward science: A meta- analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, *32*, 387–398.

Xue, Y., & Larson, R. C. (2015). STEM crisis or STEM surplus: Yes and yes. *Monthly Labour Review*. https://doi.org/10.3102/00028312134488622

Zeidan, A. H., & Jayosi, M. R. (2015). Science process skills and attitudes toward science among Palestinian secondary school students. *World Journal of Education*, *5*(1), 13–24.