2012

RELATIONSHIPS BETWEEN GENDER, SOCIOECONOMIC STATUS, MATH ATTITUDES, AND MATH ACHIEVEMENT: AN INTERNATIONAL INVESTIGATION

Sunny R. Duerr
University of Rhode Island, sunny.duerr@gmail.com

RELATIONSHIPS BETWEEN GENDER, SOCIOECONOMIC STATUS, MATH ATTITUDES, AND MATH ACHIEVEMENT: AN INTERNATIONAL INVESTIGATION

BY

SUNNY R. DUERR

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2012

DOCTOR OF PHILOSOPHY IN PSYCHOLOGY DISSERTATION

OF

SUNNY R. DUERR

APPROVED:

Thesis Committee:

Major Professor Lisa L. Harlow

Minsuk Shim

Jasmine Mena

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2012

ABSTRACT

Investigations into the factors related to math achievement have traditionally been studied within individual countries, despite the existence of large international data sets available for analysis. This dissertation investigated the relationships among gender, socioeconomic status, math attitudes, and math achievement based on information from 50 participating countries in the Trends in International Mathematics and Science Study (TIMSS).

Countries were grouped into clusters using hierarchical cluster analysis. Six cluster solutions were investigated based on average mathematics scores, average science scores, and average math attitude. The clusters were then validated on a separate sample using discriminant function analysis. The validation process utilized several country-level indicator variables, such as the Human Development Index, to ascertain the external validity of the cluster solutions.

Multiple-group latent variable modeling was employed between-clusters and within-clusters to assess the nature and strength of the relationships between gender, socioeconomic status, math attitudes, and math achievement. The findings suggest that math self-confidence has a particularly strong relationship with math achievement, and that value of math has a particularly weak relationship with math achievement. Additionally, gender differences in math achievement appear to have disappeared or now favor female students, but male students report generally higher levels of math self-confidence. Among the implications discussed is the need to promote math self-confidence in education curricula and in teacher education.

ACKNOWLEDGMENTS

Anything worth accomplishing will have its associated challenges, and the completion of a Ph.D. is an exemplar of such a philosophy. Although a person may be able to accomplish great things by sheer force of will, it is more likely that important triumphs are the result of a strong foundation which can be leaned against for support in times of need. In the following paragraphs, I would like to acknowledge those people who have contributed to my success by being part of my foundation.

The selection of a major professor is possibly the most important decision a graduate student can make; a good fit can lead to growth and a rewarding educational experience, while a bad fit can result in stagnation and in some cases animosity. I am exceedingly lucky to have been provided with the opportunity to work with Dr. Lisa L. Harlow, who has proven to be the consummate mentor. In addition to providing insight and much-appreciated expertise as needed, Dr. Harlow recognized certain predilections I possess, and she found or created opportunities for me to develop those predilections into strengths.

In addition to the guidance and support I have received from Dr. Harlow, I would like to acknowledge the contributions of Drs. Jasmine Mena and Minsuk Shim, who are the remaining members of my core committee; their feedback and insights have been greatly appreciated. I would also like to thank Dr. Joe Rossi and Dr. Joan Peckham for participating as an additional dissertation committee member and dissertation chair, respectively.

Prior to entering the Behavioral Science program at URI, I obtained a Master's degree in Applied Statistics and Research Methods from the University of Northern

Colorado. Without the background afforded to me by that program, I would not be where I am today, and my experience at URI would have been drastically different. As such, I would be remiss if I did not acknowledge the role of those professors who provided me with the foundation I had coming into the program at URI: Dr. Susan Hutchinson, Dr. Steven Pulos, and Dr. Trent Lalonde. Finally, I would like to acknowledge Dr. William Douglas Woody, for having a greater positive impact on my intellectual and academic development than he will ever know.

At the core of any solid social foundation is a strong network of support in the form of family and friends; I am fortunate in that I have some of the best of each. While some people in my family may not understand or be able to empathize with my academic pursuits, they have never wavered in their support of my endeavors, and they at least attempt to look interested as I babble about latent variables and statistically significant predictors. Without the support of my mother and father in particular, I would not have been able to get through my Bachelor's program, let alone my Master's or this Doctorate.

Finally, I would like to acknowledge the unconditional support I have received from my lovely wife, who has been a motivational inspiration to me since I met her. The passion she feels for her field makes me question whether or not any of the rest of us are passionate enough about ours.

PREFACE

This dissertation has been prepared in manuscript format. There are two manuscripts included, which have been prepared as Chapter 1 and Chapter 2. The manuscripts have been prepared according to the formatting rules for their respective journals in regard to the location of tables and figures. The tables have been numbered by chapter accordingly. Following the second manuscript is a single appendix containing a general discussion.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

TABLE PAGE

LIST OF FIGURES

CHAPTER 1

Identifying and Validating Clusters of Countries in the TIMSS 2007 Data Set

Submitted to the *International Review of Education* on December 30, 2012

Identifying and Validating Clusters of Countries in the TIMSS 2007 Data Set

In 2006 the Bush Administration released the American Competitiveness Initiative (Domestic Policy Council, 2006), in which a call was made for a renewed push to promote education in the subjects of science, technology, engineering, and mathematics (commonly referred to as STEM disciplines). According to data from international studies on academic performance the United States has fallen behind in math and science (Koretz, 2009; Schmidt & McKnight, 1998), which is concerning for a country that has long self-identified as a world-leader in education (Kuenzi, Matthews, & Mangan, 2006).

Education is one of the key indicators of a society's development and stability. Education can reduce social and economic inequality at the individual level (Lott & Bullock, 2007), and lay a foundation for a country's social and economic development. As the world's workforce is increasingly globalized due to technological advances, education plays a key role in developing and/or maintaining a competitive advantage. In many discussions in the U.S., mathematics and science receive special emphasis for being particularly important to our country's future well-being (Glenn, 2000; Kuenzi et al., 2006; National Research Council, 2007).

One major influential factor in mathematics education is math attitude; in the United States, this is a generally expressed as a positive correlation, with higher math attitude being indicative of higher math achievement (Harlow, Burkholder, & Morrow, 2002; Schreiber, 2002). This relationship has been investigated for decades (Aiken & Dreger, 1961; Anttonen, 1969), and continues to be of interest because attitude is a much more malleable variable than cognitive ability or background variables such as SES (Singh, Granville, & Dika, 2002).

As could be inferred from Bandura's self-efficacy theory (Bandura, 1997), the differences in self-efficacy levels between boys and girls in math related topics may explain much of the sex-based differences seen in math performance. As an example of this, Ethington (1992) demonstrated the importance of the student's attitude toward math as a key component in the sex-based difference in performance and that the value placed on math was more influential for boys, but indirect psychological influences such as math affect were more influential for girls in the 746 eighth-grade participants of the Second International Mathematics Study (SIMS). Additionally, Casey, Nuttall, and Pezaris (2001) demonstrated that math self-confidence significantly mediated the relationship between gender and performance in their sample of 187 eighth-grade students. Further, Köller, Baumert, and Schnabel (2001) found that, for the 602 participants in their study, students who are disinterested in math lack the motivation to learn the subject, whereas those who are highly interested often challenge themselves by selecting more advanced math courses, which in turn leads to higher learning rates and a deeper understanding of concepts.

Historically, the majority of the research on the relationship between math attitude and math achievement has been conducted from an ethnocentric perspective; researchers may investigate the relationship between these two variables in individual countries (e.g., Köller, Baumert, & Schnabel, 2001; Ma, 1997; Papanastasiou & Zembylas, 2002), but little research has investigated the relationship from a cross-national perspective, even considering the wealth of data available for such analyses.

An example of such data is the Trends in International Mathematics and Science Study (TIMSS).

TIMSS is a recurring assessment of mathematics and science achievement for 4th and 8th grade students in participating countries. Initially conducted in 1995, the study has released four waves of data as of this writing; a fifth wave, TIMSS 2011, will become available for secondary analysis in January 2013. The purpose of the TIMSS is to provide an international view of mathematics and science achievement, which can then be used by educators and policy makers as a foundation for policy relevant decisions. Prior to the public release of each wave of TIMSS data, a thorough report of the study's summary statistics is published, in which the performance of participating countries is discussed in broad terms of mean comparisons and benchmarking ratios (see Martin, Mullis, & Foy, 2008, for the TIMSS 2007 summary report). These summary reports are often used by mass media and policy makers to compare the performance of one country with that of other countries, or to illustrate how a country compares with the international mean.

As Koretz (2009) pointed out, such comparisons should be interpreted with caution. International averages are not constants, and tend to vary from assessment cycle to assessment cycle. Koretz argues that we should instead make comparisons based on countries that are most similar to our own (i.e., Australia, Canada, and the U.S.), and with countries that consistently outperform our own (i.e., Japan and Singapore). There are, however, no discernible guidelines on what countries should be considered similar to each other and what countries should be considered different from each other, or how such delineations should be made. A systematic investigation designed to identify such patterns of similarity or difference would provide additional guidance for such decisions.

Furthermore, certain advanced analysis methods such as multiple-group latent variable modeling (LVM), which can illustrate group differences in complex statistical models, can accommodate only a limited number of groups, and using all of the available countries from a data set like TIMSS as a grouping variable for such analyses would yield too many groups. However, if countries could be grouped according to similarity, or broken into separate clusters where each country was similar to other countries in its cluster and different from countries in other clusters, procedures such as multiple-group LVM could be applied.

The purpose of this study was to identify meaningful clusters of countries in the TIMSS 2007 8th grade data set. Because attitude is a strong predictor of achievement, attitude was used as a clustering variable in addition to achievement. The resulting clusters were validated using several external variables, which are discussed in detail in the methods section below.

METHODS

Participants

The sample for this analysis included the 8th grade students from 48 of the 50 participating countries and territories in TIMSS 2007; sample characteristics in terms of sample size, average math and science achievement scores, and average math attitude scores are presented in Table 1.1. It should be noted that only 48 countries were investigated as Mongolia and Morocco are excluded from the analysis due to sampling violations reported in Mullis, Martin, and Foy (2008).

INSERT TABLE 1.1 APPROXIMATELY HERE

The complex sampling design of studies like TIMSS mandate special consideration in analyses by utilizing what are known as sampling weights included in the data set. These weights account for sampling design, take into account stratification and disproportionate sampling of subgroups, and include adjustments for non-response (Foy & Olson, 2009). The TOTWGT sample weight in TIMSS 2007 is the weighting variable used to calculate student population estimates within countries, and use of this variable will ensure that subgroups are properly and proportionally represented in population estimates; using the TOTWGT variable inflates the sample size in the analysis to reflect the approximate size of the population (i.e., the total weight).

However, when making cross-national comparisons TOTWGT may not be applicable because larger countries will be overrepresented in the analysis. For analyses in which countries should be weighted equally, the SENWGT sample weight is preferred. SENWGT, presumed to be an acronym for senate weight, is a transformation of TOTWGT which produces a weighted sample of 500 for each country (Foy & Olson, 2009); in this way, the SENWGT variable forces each country to have equal representation, hence the name senate weight. Because the current analysis is concerned with making cross-national comparisons, the SENWGT weighting variable was used.

Measures

Achievement variables.

TIMSS achievement variables are standardized with a mean of 500 and a standard deviation of 100. TIMSS provides scores for sub-topics within each subject. For math, sub-topics include algebra, geometry, statistics, and so on. For science, sub-topics include physics, chemistry, biology, and so on. The achievement variables used in this study are the overall math and science achievement scores provided by TIMSS.

Achievement scores measured by TIMSS are complicated. The TIMSS attempts to measure achievement over a broad variety of math and science topics. In order to reduce time demands on each student, a complex matrix-sampling booklet design is implemented (Williams et al., 2009). This design requires that individual students respond to a relatively small number of items from the overall battery of assessment items. Item responses are then aggregated across all students to provide coverage of a wide range of content.

Because each student responds to only a selection of possible items, TIMSS utilizes an item response theory (IRT) scaling approach based on multiple imputation techniques to create a set of plausible values (Foy, Galia, & Li, 2007; Foy & Olson, 2009; Williams et al., 2009). Plausible values are essentially imputed scores based on a student's item responses in conjunction with background variables. Imputed scores based on limited information certainly contain some amount of error, and to account for this error, scores should be imputed multiple times; the result of each of these imputations is considered a plausible value, or a score that a given student could have received, had the student answered all items in the TIMSS assessment. The TIMSS data set contains five plausible values per student for each achievement related variable.

To accommodate the use of plausible values in any analysis, the analysis must be conducted once for each plausible value. The results of the separate analyses are then combined into a single result which includes parameter estimates and standard errors incorporating both sampling and imputation error (Foy et al., 2007). All analyses which utilize TIMSS achievement variables in this study use the five plausible values provided by the TIMSS data set. The achievement and attitude scores used in this study were calculated using the International Association for the Evaluation of Educational Achievement (IEA) International Database Analyzer (IDB Analyzer). The IDB Analyzer helps data analysts conquer the complexities associated with the sampling design used by international databases; the IDB Analyzer provides an interface which produces SPSS syntax that will accurately accommodate the data set's complexities.

Caution must be taken in the interpretation of assessments which use plausible values. Due to the use of imputation, analyses based on plausible values cannot be considered representative of a given individual student's achievement; rather, plausible values represent a range of reasonable values for a given student's achievement.

Attitude variables.

TIMSS provides three attitude variables for math and science: positive affect toward math/science, self-confidence in math/science, and value of math/science. The scores for these variables range between 1 and 3, with 1 indicating high attitude and 3 indicating low attitude. These variables were rescaled by subtracting the participant's attitude score from 4, resulting in an ordinal scale with scores between 1 and 3 where low scores indicate low attitude toward math/science and high scores indicate high attitude toward math/science.

Several countries do not have data for science attitude. These countries are Algeria, Armenia, Bosnia/Herzegovina, Bulgaria, Cyprus, Czech Republic, Georgia, Hungary, Indonesia, Lebanon, Lithuania, Malta, Mongolia, Morocco, Romania, Russian Federation, Slovenia, Sweden, Syria, Ukraine, and Serbia. These 21 countries represent a large proportion of the countries in the cluster analysis. As such, science attitude is not included as a variable in the cluster analysis.

Cluster validation variables.

A key component in any cluster analysis is providing evidence of the validity of the resulting clusters. To validate the resulting clusters in this study, several country-level variables, external to the TIMSS data and unrelated to math and science achievement or math attitude, were used. These variables include measures of democracy, human development, education, economic freedom, freedom of the press, and gender equality. To operationalize these variables, the following indices were used: the Democracy Index (Economist Intelligence Unit, 2011); the Human Development Index (Human Development Report, 2011); the Education Index (Human Development Report, 2011); the Economic Freedom score (The Heritage Foundation, 2011); the Freedom of the Press score (Reporters Without Borders, 2012); and the Gender Equality score (Human Development Report, 2011). Table 1.2 provides the index values for the validation variables, and a summary of these measures follows.

INSERT TABLE 2 APPROXIMATELY HERE

Human Development Index, Education Index, and Gender Equality Index

The measures for human development, education, and gender equality all come from the Human Development Report (2011), published by the United Nations Development Programme. According to the United Nations Development Programme, the Human Development Index (HDI) is "a composite index measuring average achievement in three basic dimensions of human development – a long and healthy life, knowledge, and a decent standard of living" (p. 130). The HDI has a rating scale between 0 and 1, with numbers closer to 1 indicating higher levels of human development. For 2011, scores ranged from 0.541 (Ghana) to 0.943 (Norway).

Also found in the Human Development Report, the Education Index is a composite of adult literacy rates and enrollment ratios at the primary, secondary, and tertiary levels of education. The Education Index is also on a rating scale between 0 and 1, with numbers closer to 1 indicating a higher level of education for the country. For 2011, scores ranged between 0.627 (Ghana) and 0.993 (Australia), with one case of missing data (Georgia).

The Gender Equality Index comes from the Gender Inequality Index within the Human Development Report. The Gender Inequality Index is a composite measure which represents a country's gender-based inequality in reproductive health, empowerment, and the labor market. The Gender Inequality Index is on a rating scale between 0 and 1, with numbers closer to 1 representing high levels of inequality. In order to have the scale coincide with the other scales in the analysis (i.e., high scores are more positive), this scale was rescaled by subtracting the provided score from 1, yielding scores such that numbers closer to 1 represent a high level of equality. Thus, we have renamed the scale to the Gender Equality Index for the purpose of this study. In 2011, scores on the Gender Equality scale ranged from 0.354 (Saudi Arabia) and 0.951 (Sweden). Six territories lack a score on the Gender Equality Index: Taiwan, Hong Kong, Serbia, Bosnia and Herzegovina, Palestine, and Egypt.

Democracy Index

The Democracy Index "provides a snapshot of the state of democracy worldwide for 165 independent states and two territories…The overall Democracy index is based on five categories: electoral process and pluralism; civil liberties; the functioning of government; political participation; and political culture" (Economist Intelligence Unit, 2011, p. 1). The Democracy Index original scores are on a scale from 0 to 10, with higher scores indicating higher levels of democracy. These scores were divided by 10 in order to force the scale to be comparable with the other scales in the analysis (i.e., ranging from 0 to 1, with numbers closer to 1 indicating higher levels of democracy). Final scores on the Democracy Index ranged between 0.177 (Saudi Arabia) and 0.980 (Norway).

Economic Freedom Index

The Index of Economic Freedom is a joint venture between The Heritage Foundation and the Wall Street Journal. The index is a composite of ten components of economic freedom: property rights, freedom from corruption, fiscal freedom, government spending, business freedom, labor freedom, monetary freedom, trade freedom, investment freedom, and financial freedom. Each of these ten components is rated from 1 to 100, and the overall economic freedom score for a country is the average of these ten components. For the current study, the reported value for the

Economic Freedom Index was divided by 100 to yield scores ranging from 0 to 1, with numbers closer to 1 indicating higher levels of economic freedom. Final scores on the Economic Freedom Index ranged between 0.421 (Iran) and 0.897 (Hong Kong), with one case of missing data (Palestine).

Freedom of the Press Index

The Press Freedom Index is a report measuring the treatment of journalists and media in countries (Reporters without Borders, 2012). The report is based upon a 40-item questionnaire which assesses the state of press freedom in each country. Scores on the Press Freedom Index range between -10.00 and 142.00, with smaller numbers indicating greater press-related freedom. For the purposes of the current study, the scores for this index were first subtracted from 150, yielding scores ranging from 8 to 160. These values were then divided by 160, yielding scores on a 0 to 1 scale, with numbers closer to 1 indicating higher levels of press-related freedom. Final scores for the Freedom of the Press Index ranged between 0.075 (Syria) and 1.00 (Norway).

Procedures

Hierarchical cluster analysis was conducted to identify groups of countries similar to each other yet different from other groups of countries. Punj and Stewart (1983) reviewed several methods of cluster analysis, concluding that Ward's (1963) clustering algorithm consistently performed best among hierarchical clustering techniques. Ward's algorithm forms mutually exclusive groups or clusters, starting with n clusters (i.e., one for each participant in the sample) and iteratively reducing the number of clusters by 1. At each stage a given unit is determined to either fit into an existing cluster or to form a new cluster with another unit, ultimately resulting in a single cluster.

Cluster analysis is inherently an exploratory procedure. It is for this reason that a key component of cluster analysis is the evaluation of the reliability (the degree to which the cluster solutions are consistent) and validity (the degree to which the cluster solutions are meaningful) of the resulting clusters. Evidence of reliability can be demonstrated by testing the structure of the cluster solutions on a separate sample, known as cross-validation (Sherman & Sheth, 1977). Validity can be demonstrated by assessing the identified clusters on variables other than those used for the cluster analysis (Punj & Stewart, 1983). In order to create a cross-validation sample, the TIMSS sample was split into two approximately equal halves using the random selection feature in SPSS 18.0. The initial cluster analysis was performed on an initial (model building) sample, and the subsequent validation analyses were performed on a second (cross-validation) sample. Descriptive statistics for the model building sample can be seen in Table 1.3, and descriptive statistics for the cross-validation sample can be seen in Table 1.4.

INSERT TABLE 1.3 APPROXIMATELY HERE

INSERT TABLE 1.4 APPROXIMATELY HERE

Because the goal for this analysis was to identify clusters of countries that could be useful for multiple-group latent variable modeling approaches, the number of countries per cluster needed to be limited to no more than 10. With 48 countries under consideration, a 4-cluster solution would automatically contain at least one cluster that would be too large for multiple-group analysis, while a 5-cluster solution would not.

For this reason, I decided to begin with the 4-cluster solution and proceed until the clustering solution produced multiple clusters that were significantly smaller than the others (i.e., one or two countries).

The initial cluster analysis for this study was conducted using the default cluster analysis functions available in the R program's *hclust* command. The cluster analysis was performed using Ward's clustering algorithm with squared Euclidean distance for the distancing measure. The cluster analysis routine was performed 6 times, with the number of clusters set to a specific value between 4 and 9 for each analysis (i.e., once to obtain a 4-cluster solution, once for a 5-cluster solution, and so on). This provided six initial cluster solutions to explore during the initial model-testing phase.

In order to investigate the reliability and validity of the cluster solutions that resulted from the initial cluster analysis, the following processes were performed once for each cluster solution. First, countries were assigned a group identifier based on cluster membership from the initial cluster analysis. This identifier was used as the grouping variable for a discriminant function analysis (DFA) and a series of ANOVAs (using the R default MANOVA function and the Anova function from the CAR package, respectively).

DFA is a procedure in which several continuous independent variables are used to predict membership in a categorical grouping variable. For the DFA, cluster membership was included as the dependent variable and the validation variables discussed previously were included as independent variables. A common procedure associated with discriminant function analysis is the assessment of the predictive accuracy of a classification system (Harlow, 2005; Tabachnick & Fidell, 2007).

Applying that purpose to this study, the cluster solutions identified in the initial cluster analysis could be considered reasonably accurate if countries are correctly classified at a rate that is greater than what could be expected by chance (Harlow, 2005, pp. 141-142).

Because cluster membership for the validation analyses was based on the results of the initial cluster analysis, a highly accurate comparison of predicted and actual cluster membership using the cross-validation sample and the validation variables outlined previously would provide elegant evidence for the reliability and validity of the clusters. This analysis was carried out for each of the initial cluster solutions using the lda function in the MASS package in the R software.

RESULTS

As previously described, the TIMSS 2007 data set was split into two roughly equally sized subsamples, a model building sample and a cross-validation sample. The hierarchical cluster analysis using Ward's (1963) clustering algorithm was conducted six times on the model building sample, with the number of clusters to investigate set between 4 and 9; this yielded six different cluster solutions. A dendogram, a commonly used diagram for displaying clustering patterns from a cluster analysis, can be seen in Figure 1.1; the six initial cluster solutions are shown in Table 1.5.

INSERT FIGURE 1.1 APPROXIMATELY HERE

INSERT TABLE 1.5 APPROXIMATELY HERE

To begin validation of the cluster solutions, a DFA was conducted on the cross-validation sample using cluster membership as the categorical dependent variable, and the TIMSS achievement variables and the previously discussed external validation variables as the continuous independent variables. Again, as outlined by Harlow (2005), the cluster solutions identified in the initial cluster analysis could be considered reasonably accurate if countries are correctly classified by the DFA at a rate that is greater than what could be expected by chance.

DFA is mathematically similar to MANOVA, and utilizes the same methods for examining model fit. For overall model fit, several F-tests can be examined, including those associated with Roy's Largest Root, Wilkes' Lambda, Pillai's Trace, and Hotelling-Lawley Trace; because these different F-tests all yield similar interpretations, only Wilkes' Lambda is discussed. As a follow up to significant F-tests for DFA, independent one-way ANOVAs were conducted for each independent variable. The results of both the DFA and follow up ANOVA tests can be seen in Table 1.6.

<u>INSERT TABLE 1.6 APPROXIMATELY HERE</u>

The results of the omnibus tests for the DFA suggest that the clusters are significantly different from each other based on the linear combination of independent variables. This provides some preliminary evidence for the validity of the cluster solutions. However, greater evidence would be provided if the cluster solutions from the model building sample were found to be accurate in predicting group membership in the cross-validation sample. The results of this analysis for each cluster solution follows.

The 9-Cluster Solution (Table 1.7)

<u>INSERT TABLE 1.7 APPROXIMATELY HERE</u>

The predictive accuracy for the 9-cluster solution (shown in Table 1.8) was quite high; 37 out of the 48 countries (77%) were correctly classified. Based on prior probabilities provided by the initial cluster analysis, 6.375 countries (13.28%) could be expected to have been correctly classified by chance. However, the 5th cluster in this solution included only a single country (Lebanon), which in subsequent cluster solutions was grouped within another cluster. Additionally, the last cluster in this solution contains only 2 countries, Ghana and Qatar. Thus, these two clusters did not appear very stable or robust.

INSERT TABLE 1.8 APPROXIMATELY HERE

The 8-Cluster Solution (Table 1.9)

INSERT TABLE 1.9 APPROXIMATELY HERE

The predictive accuracy for the 8-cluster solution (shown in Table 1.10) was also high; 36 out of the 48 countries (75%) were correctly classified. Based on prior probabilities provided by the initial cluster analysis, 6.625 countries (13.80%) could be expected to have been correctly classified by chance. The major change between this cluster solution and the 9-cluster solution is that Lebanon, formerly the only country in Cluster 5, is now included in cluster 4. The last cluster in this solution once again contains only Ghana and Qatar.

INSERT TABLE 1.10 APPROXIMATELY HERE

The 7-Cluster Solution (Table 1.11)

INSERT TABLE 1.11 APPROXIMATELY HERE

The predictive accuracy for the 7-cluster solution (shown in Table 1.12) was again high, with 41 out of the 48 countries (85%) being correctly classified. Based on prior probabilities provided by the initial cluster analysis, 8.667 countries (18.06%) could be expected to have been correctly classified by chance. The major change for this cluster solution from the 8-cluster solution is that two large clusters, previously Cluster 3 and Cluster 4, have been merged. Qatar and Ghana were once again the only countries in the last cluster in this solution.

INSERT TABLE 1.12 APPROXIMATELY HERE

The 6-Cluster Solution (Table 1.13)

INSERT TABLE 1.13 APPROXIMATELY HERE

The predictive accuracy for the 6-cluster solution (shown in Table 1.14) was 83%, or 40 of 48 countries correctly classified. Based on prior probabilities provided by the initial cluster analysis, 9.917 countries (20.66%) could be expected to have been correctly classified by chance. The major change in this cluster solution from the 7-cluster solution is that two clusters, the 5^{th} and 6^{th} clusters in the 7-cluster solution, are merged to form one larger cluster. The last cluster in this solution contains, once again, only Qatar and Ghana.

INSERT TABLE 1.14 APPROXIMATELY HERE

The 5-Cluster Solution (Table 1.15)

INSERT TABLE 1.15 APPROXIMATELY HERE

The predictive accuracy for the 5-cluster solution (shown in Table 1.16) was the highest of all cluster solutions with 90% correct classification (43 out of 48 countries). Based on prior probabilities provided by the initial cluster analysis, 10.833 countries (22.57%) could be expected to have been correctly classified by chance. The major change in this cluster solution from the 6-cluster solution is that the cluster containing Qatar and Ghana has been merged into a larger cluster.

INSERT TABLE 1.16 APPROXIMATELY HERE

The 4-Cluster Solution (Table 1.17)

INSERT TABLE 1.17 APPROXIMATELY HERE

The 4-cluster solution (shown in Table 1.18) yielded 37 out of 48 countries (77%) being correctly classified. Based on prior probabilities provided by the initial cluster analysis, 16.292 countries (33.94%) could be expected to have been correctly classified by chance. The major change in this cluster solution over the 5-cluster solution is the combination of two larger clusters to form one very large cluster containing 21 countries.

INSERT TABLE 1.18 APPROXIMATELY HERE

General Cluster Solution Discussion

Cluster analysis cannot be considered complete until the resulting clusters have been assigned meaning (Punj & Stewart, 1983; Ward, 1963). As Tukey (1977) implies, a picture is worth a thousand words, and the following figures begin to illustrate some of the patterns in the clusters of countries. Figure 1.2 shows the countries arranged by cluster according to average math achievement score, and Figure 1.3 shows the countries arranged by cluster according to math attitude score.

INSERT FIGURE 1.2 APPROXIMATELY HERE

INSERT FIGURE 1.3 APPROXIMATELY HERE

To begin the process of assigning meaning to the cluster solutions, we first consider 3 of the clusters that were most stable across all 6 of the cluster solutions.

The first of these three clusters consists of Japan, Taiwan, South Korea, Hong Kong, and Singapore. These five Asian countries share similar religious views (predominantly Eastern philosophies of Buddhism and Daoism, though South Korea does have a large Christian population), have generally high ratings of economic freedom, and generally high ratings for human development. Additionally, these five countries are consistently at the top of international assessments for mathematics and science achievement, yet these countries also report some of the lowest ratings for student math attitude.

The second of the stable clusters includes 9 countries: Australia, Czech Republic, England, Hungary, Lithuania, Russian Federation, Slovenia, Sweden, and the United States. Although these countries do not have geographic proximity, they do share certain cultural characteristics. For example, all of these countries are predominantly Christian, and several of them share cultural history (i.e., the U.S. and Australia were both once English colonies, while many of the remaining countries were once part of the Soviet Union). These countries are all political democracies, have high ratings for human development and education, and the citizens for most of these countries enjoy the highest ratings for gender equality, democracy, economic freedom, and freedom of the press.

The third consistently stable cluster contains Thailand, Jordan, Tunisia, Turkey, Bahrain, Iran, and Syria. Although this cluster eventually ended up being merged with another cluster in the 4-cluster solution, these countries represented their own independent cluster up until that point. These 7 countries are predominantly Islamic in religion, and most of them share geographic proximity with the exception being

Thailand. This cluster of countries generally has very low ratings for democracy, human development, gender equality, and freedom of the press.

Now we will turn to the remaining clusters, which were not as consistent across cluster solutions. This is to be expected, since hierarchical cluster analysis is a process of combining similar groups with each other to the point of maximum inclusion. What this means for the current analysis is that we ended up with two mega-clusters. These mega clusters are clusters which, by the 5-cluster solution, had absorbed several other clusters over the course of the analysis.

The first mega-cluster is comprised of 14 countries: Armenia, Italy, Malta, Norway, Scotland, Serbia, Ukraine, Bosnia and Herzegovina, Bulgaria, Cyprus, Romania, Israel, Lebanon, and Malaysia. As a mega-cluster, these countries are all similar in terms of human development, education, economic freedom, and freedom of the press. The majority of these countries are predominantly Christian, with the exceptions being Israel, Lebanon, and Malaysia. In the early cluster solutions, these 14 countries were represented by 2 distinct clusters, with Armenia, Italy, Malta, Norway, Scotland, Serbia, and Ukraine in the first cluster, and the remaining countries in the second cluster. The countries in the first group of the mega-cluster (Norway, Italy, Scotland, Malta, Serbia, Armenia, and Ukraine) have geographic proximity with each other; this group also has slightly higher ratings of democracy, political rights, gender equity, and civil rights compared with the mega-cluster's second group (Bosnia and Herzegovina, Bulgaria, Cyprus, Romania, Israel, Lebanon, and Malaysia).

The second mega-cluster is comprised of 13 countries: Colombia, Georgia, Palestine, Algeria, Egypt, Indonesia, Kuwait, Oman, Botswana, El Salvador, Saudi

Arabia, Ghana, and Qatar. Most of the countries in this mega-cluster are Muslim, although Colombia, Georgia, Botswana, and El Salvador are predominantly Christian. As a mega-cluster, this group of countries has some of the lowest ratings for all of the validation variables, and they are all among the lowest scores for mathematics and science achievement. Interestingly, these countries have among the highest math attitude scores. In the early cluster solutions, this mega-cluster was represented by three smaller clusters. Botswana, El Salvador, Kuwait, Palestine, and Saudi Arabia were one cluster; Colombia, Georgia, Algeria, Egypt, Indonesia, and Oman were the second cluster; and Qatar and Ghana represented a third cluster.

In addition to the findings of the cluster analysis and the validation analysis, an interesting pattern in the relationship between math attitude and math achievement emerged. As was previously stated, this relationship is generally expressed as a positive relationship at the individual level with higher math attitude generally indicative of higher math achievement. However, when observed at the country-level, the relationship is less clear; in fact, the relationship becomes counterintuitive. As Figure 1.4 illustrates, at the country level, the relationship between math attitude and math achievement appears to be a fairly strong, negative correlation.

INSERT FIGURE 1.4 APPROXIMATELY HERE

DISCUSSION

The purpose of this study was to investigate the TIMSS 2007 8th grade data set with the intent to identify meaningful clusters of countries. Six different cluster solutions (i.e., solutions containing between 4 and 9 clusters) were identified via cluster analysis, cross-validated using discriminant function analysis, and then externally validated on several additional variables of sociocultural and political interest. The findings of these analyses suggest guidelines for cluster membership, allowing future researchers to make international comparisons among countries that are considered similar to their own based on cluster membership.

Although the classification analyses lent support for each of the six initially identified cluster solutions in terms of the percentage of correctly classified countries, there are some considerations that should be made. First, the cluster containing only Lebanon in the 9-cluster solution immediately disappeared and was absorbed by another cluster. Depending on the researcher's purpose, it may be advisable to simply begin with the 8-cluster solution rather than including a cluster consisting of only a single country. Additionally, the 4-cluster solution may be the point at which the clusters begin to be less meaningful, as the resulting 21-country cluster in this solution may be too large for any meaningful between-cluster comparisons. Based on this information, we would recommend the use of the 8-cluster, 7-cluster, 6-cluster, or 5-cluster solutions, each of which had high reliability in our analyses.

Which of these cluster solutions would be best for any given analysis is dependent on the nature of the research question, but it may also be influenced by the limits of technology. Although they had high predictive reliability for our analysis, the 7-cluster solution through the 4-cluster solution all contain at least one cluster that is too large for current technology to handle when conducting multiple-group LVM analyses, which may make the 8-cluster solution the ideal for researchers seeking to employ those methods. This may not seem intuitive, since the 8-cluster solution had slightly lower predictive accuracy (75% accuracy) than all of the cluster solutions, but considering the ratio of expected accurate predictions by chance alone (13.8%), the 8-cluster solution can still be considered quite accurate. However, if cluster size is not necessarily a concern for the researcher the 5-cluster solution, having the highest predictive accuracy among the cluster solutions at 90% accuracy, may be a desirable choice for future investigations.

As was previously discussed, cluster analysis is an exploratory procedure. The purpose of the analysis is to identify any patterns of relationships among group members based on similarities within members of the same group and differences between members of other groups. The resulting clusters are largely based on the variables included in the cluster analysis and the variables used to validate the clusters. As such, it is to be expected that other random samples taken from the TIMSS data set could yield slightly different cluster solutions. This could lead to several interesting additional investigations. Of particular interest would be a cluster analysis using the soon-to-be-available TIMSS 2011 data to see if the cluster solutions investigated here replicate in that data set. Similar analyses could be done on the previous versions of TIMSS as well, although earlier administrations of TIMSS had fewer participating countries which would influence the resulting cluster solutions.

Finally, it is worth discussing the negative relationship between math achievement and math attitude at the country level. First, it is important to remember that this does not mean that math attitude has a negative impact on math achievement at the individual level. However, it could indicate significant cultural differences in the emphasis placed on either math achievement or math attitude, or both. These cultural differences may be manifesting in the TIMSS 2007 data as Extreme Response Style

(ERS), a type of confound driven by group differences in the actual attitude and its respective response patterns (Eid, Langeheine, & Deiner, 2003; Morren, Gelissen, & Vermunt, 2012; Poortinga & van de Vijver, 1987). Another possible explanation could be that this is an excellent example of Yule-Simpson’s Paradox, wherein a correlation evidenced in a number of groups disappears or reverses direction when the groups are combined. The international level of the comparison may be masking important subgroups within the sample, creating a situation in which the gestalt of the correlation may be something different than the sum of its parts.

Table 1.1: Descriptive statistics for TIMSS Math Achievement, Science Achievement and Math Attitude by country.

Country		Math Achievement		Science Achievement			Math Attitude	
	N	Mean	SD	Mean	SD	N	Mean	SD
Algeria	5,447	386.75	59.25	408.06	62.60	5,018	2.67	0.38
Armenia	4,689	498.68	84.74	487.96	101.14	3,786	2.32	0.56
Australia	4,069	496.23	79.43	514.79	80.32	3,939	2.30	0.55
Bahrain	4,230	398.07	83.60	467.45	86.03	4,123	2.53	0.49
Bosnia and Herzegovina	4,220	455.86	77.80	465.75	79.44	3,949	2.27	0.58
Botswana	4,208	363.54	76.58	354.53	99.42	3,933	2.58	0.41
Bulgaria	3,079	463.63	101.60	470.28	102.62	3,796	2.29	0.59
Chinese Taipei	4,046	598.30	105.51	561.00	89.27	4,018	2.01	0.66
Colombia	4,873	379.64	78.94	417.18	76.65	4,689	2.60	0.42
Cyprus	4,399	465.48	89.32	451.62	85.32	4,327	2.38	0.57
Czech Republic	4,845	503.81	73.69	538.88	71.39	4,807	2.22	0.56
Egypt	6,582	390.56	100.25	408.24	99.38	6,012	2.69	0.39
El Salvador	4,063	340.44	72.82	387.27	69.77	3,894	2.56	0.39
England	4,025	513.40	83.58	541.50	85.40	3,938	2.38	0.53
Georgia	4,178	409.62	96.46	420.90	83.33	3,559	2.47	0.51
Ghana	5,294	309.37	91.60	303.27	108.36	5,001	2.62	0.39
Hong Kong, SAR	3,470	572.49	93.73	530.21	80.97	3,437	2.23	0.59
Hungary	4,111	516.90	84.68	539.03	76.58	4,066	2.23	0.56
Indonesia	4,203	397.11	87.34	426.99	74.18	4,140	2.58	0.37
Iran, Islamic Republic of	3,981	403.38	86.09	458.93	81.34	3,345	2.54	0.49
Israel	3,294	463.25	98.87	467.92	100.91	3,126	2.47	0.53
Italy	4,408	479.63	76.23	495.15	77.52	4,287	2.23	0.63
Japan	4,312	569.81	85.42	553.82	77.11	4,275	1.96	0.56
Jordan	5,251	426.89	102.21	481.72	97.72	4,971	2.66	0.43
Korea, Republic of	4,240	597.27	92.07	553.14	75.86	4,230	2.08	0.61
Kuwait	4,091	353.67	78.64	417.96	89.24	3,821	2.52	0.52
Lebanon	3,786	449.06	74.64	413.61	96.81	3,538	2.52	0.50
Lithuania	3,991	505.82	79.74	518.56	78.21	3,942	2.34	0.52
Malaysia	4,466	473.89	79.25	470.80	88.20	4,448	2.47	0.45
Malta	4,670	487.75	91.77	457.17	113.86	4,630	2.29	0.57
Mongolia	4,499	432.17	81.49	449.31	73.56	4,116	2.61	0.44
Morocco	3,060	380.78	80.33	401.83	78.55	2,768	2.65	0.39
Norway	4,627	469.22	65.66	486.76	73.27	4,479	2.34	0.54
Oman	4,752	372.43	94.94	422.50	95.74	4,560	2.67	0.37
Palestinian National Auth	4,378	367.15	102.44	404.13	110.93	4,153	2.50	0.48
Qatar	7,184	306.79	93.36	318.85	125.87	6,843	2.51	0.53
Romania	4,198	461.32	99.75	461.90	87.89	4,054	2.29	0.57
Russian Federation	4,472	511.73	83.08	529.57	77.65	4,347	2.40	0.53
Saudi Arabia	4,243	329.34	76.43	403.25	77.98	3,997	2.48	0.50

Table 1.1 (Continued)

Country	N	Math Achievement Mean	Math Achievement SD	Science Achievement Mean	Science Achievement SD	N	Math Attitude Mean	Math Attitude SD
Scotland	4,070	487.41	79.73	495.73	81.12	3,991	2.37	0.51
Serbia	4,045	485.77	89.45	470.31	84.72	3,894	2.23	0.58
Singapore	4,599	592.79	92.96	567.25	103.89	4,581	2.43	0.55
Slovenia	4,043	501.48	71.62	537.54	72.02	3,970	2.18	0.53
Sweden	5,215	491.30	70.05	510.69	78.03	4,889	2.33	0.55
Syria, Arab Republic of	4,650	394.84	82.40	451.98	74.71	4,173	2.59	0.46
Thailand	5,412	441.39	91.62	470.61	82.73	5,369	2.47	0.41
Tunisia	4,080	420.41	66.52	444.90	60.48	3,948	2.57	0.48
Turkey	4,498	431.81	108.74	454.16	91.89	4,365	2.53	0.47
Ukraine	4,424	462.16	89.23	485.06	83.99	4,182	2.40	0.52
United States	7,377	508.45	76.74	519.99	82.27	7,261	2.40	0.54

Note: N = Sample size; SD = Standard Deviation; Sample sizes for Math Achievement and Science Achievement are the same with the exception of Bulgaria, which had a Science Achievement sample size of 4,019.

Table 1.2: Mean ratings for validation variables by country.

Country	DEM	HDEV	EDUC	ECON	PRESS	GENEQ
Algeria	0.344	0.698	0.886	0.524	0.588	0.588
Armenia	0.409	0.716	0.909	0.697	0.769	0.657
Australia	0.922	0.929	0.993	0.825	0.913	0.864
Bahrain	0.292	0.806	0.893	0.777	0.156	0.712
Bosnia and Herzegovina	0.524	0.733	0.988	0.575	0.816	NA
Botswana	0.763	0.633	0.788	0.688	0.863	0.493
Bulgaria	0.678	0.771	0.930	0.649	0.756	0.755
Chinese Taipei	0.746	0.687	0.927	0.708	0.856	NA
Colombia	0.663	0.710	0.881	0.680	0.522	0.518
Cyprus	0.729	0.840	0.910	0.733	0.956	0.859
Czech Republic	0.819	0.865	0.938	0.704	0.969	0.864
Egypt	0.395	0.644	0.697	0.591	0.328	NA
El Salvador	0.647	0.674	0.798	0.688	0.879	0.513
England	0.816	0.863	0.957	0.745	0.925	0.791
Georgia	0.474	0.733	NA	0.704	0.700	0.582
Ghana	0.602	0.541	0.627	0.594	0.869	0.402
Hong Kong, SAR	0.592	0.898	0.879	0.897	0.831	NA
Hungary	0.704	0.816	0.960	0.666	0.875	0.763
Indonesia	0.653	0.617	0.840	0.560	0.513	0.495
Iran, Islamic Republic of	0.198	0.707	0.793	0.421	0.084	0.515
Israel	0.753	0.888	0.945	0.685	0.742	0.855
Italy	0.774	0.874	0.965	0.603	0.815	0.876
Japan	0.808	0.901	0.965	0.728	0.944	0.877
Jordan	0.389	0.698	0.870	0.689	0.583	0.544
Korea, Republic of	0.806	0.897	0.949	0.698	0.858	0.889
Kuwait	0.374	0.760	0.872	0.649	0.763	0.771
Lebanon	0.532	0.739	0.857	0.601	0.741	0.560
Lithuania	0.724	0.810	0.968	0.713	0.913	0.808
Malaysia	0.619	0.761	0.851	0.663	0.588	0.714
Malta	0.828	0.832	0.887	0.657	0.816	0.728
Norway	0.980	0.943	0.989	0.703	1.000	0.925
Oman	0.326	0.705	0.790	0.698	0.594	0.691
Palestine	0.497	0.641	0.886	NA	0.463	NA
Qatar	0.318	0.831	0.888	0.705	0.650	0.451
Romania	0.654	0.781	0.915	0.647	0.850	0.667
Russian Federation	0.392	0.755	0.933	0.505	0.525	0.662
Saudi Arabia	0.177	0.770	0.828	0.662	0.417	0.354

Table 1.2 (Continued)

Country	DEM	HDEV	EDUC	ECON	PRESS	GENEQ
Scotland	0.816	0.863	0.957	0.745	0.925	0.791
Serbia	0.633	0.766	0.891	0.580	0.756	NA
Singapore	0.589	0.866	0.913	0.872	0.556	0.914
Slovenia	0.776	0.884	0.969	0.646	0.880	0.825
Sweden	0.950	0.904	0.974	0.719	0.972	0.951
Syria, Arab Republic of	0.199	0.632	0.773	0.513	0.075	0.526
Thailand	0.655	0.682	0.888	0.647	0.553	0.618
Tunisia	0.553	0.698	0.772	0.585	0.561	0.707
Turkey	0.573	0.699	0.828	0.642	0.500	0.560
Ukraine	0.594	0.729	0.939	0.458	0.600	0.665
United States	0.811	0.910	0.968	0.778	0.850	0.701

Note: DEM = Democratic Index; HDEV = Human Development Index; EDUC = Education Index; ECON = Economic Freedom Index; PRES = Freedom of the Press Index, GENEQ = Gender Equality Index

Table 1.3: Descriptive statistics for model building sample.

		Math		Science			Attitude	
Country	N	Mean	SD	Mean	SD	N	Mean	SD
Algeria	2,781	386.46	58.51	408.86	62.42	2559	2.67	0.38
Armenia	2,374	498.99	84.84	487.44	102.06	1911	2.33	0.56
Australia	2,048	496.89	77.76	513.39	78.94	1980	2.31	0.55
Bahrain	2,140	398.75	85.23	467.58	87.14	2092	2.52	0.50
Bosnia and Herzegovina	2,157	453.20	77.45	464.11	79.23	2003	2.25	0.58
Botswana	2,060	362.71	75.90	353.62	98.31	1924	2.58	0.42
Bulgaria	1,984	461.74	101.27	470.11	103.52	1864	2.29	0.58
Chinese Taipei	2,045	598.57	106.19	560.70	89.70	2029	2.00	0.66
Colombia	2,413	379.81	78.84	418.11	77.65	2320	2.60	0.42
Cyprus	2,227	467.02	89.92	452.62	86.46	2192	2.40	0.56
Czech Republic	2,446	503.29	72.30	537.61	70.07	2420	2.23	0.55
Egypt	3,175	390.86	100.34	408.34	99.39	2924	2.69	0.39
El Salvador	2,089	341.36	72.38	388.70	68.70	1992	2.57	0.39
England	1,995	516.36	83.01	544.64	85.18	1958	2.37	0.53
Georgia	2,041	407.36	96.36	419.94	83.62	1729	2.48	0.50
Ghana	2,615	309.28	90.07	302.32	107.47	2482	2.61	0.39
Hong Kong, SAR	1,713	572.99	92.62	531.04	81.16	1695	2.24	0.58
Hungary	1,973	518.76	84.68	541.44	76.16	1953	2.23	0.55
Indonesia	2,128	397.11	87.77	427.67	73.90	2094	2.56	0.37
Iran, Islamic Republic of	2,052	402.45	87.75	458.16	83.18	1733	2.54	0.50
Israel	1,655	462.84	98.11	467.53	99.50	1566	2.48	0.52
Italy	2,152	478.20	75.97	493.43	77.02	2100	2.22	0.62
Japan	2,189	569.10	85.89	553.48	76.48	2172	1.95	0.56
Jordan	2,585	424.34	101.58	479.18	96.67	2450	2.66	0.43
Korea, Republic of	2,080	596.04	93.31	552.53	77.02	2075	2.08	0.62
Kuwait	2,050	352.98	78.39	417.15	88.89	1919	2.52	0.52
Lebanon	1,913	447.82	75.26	411.40	99.55	1787	2.52	0.50
Lithuania	1,979	508.88	78.27	521.13	77.56	1955	2.35	0.52
Malaysia	2,230	471.63	78.41	467.73	87.26	2218	2.47	0.45
Malta	2,337	488.92	92.11	459.00	114.22	2317	2.30	0.57
Mongolia	2,342	433.71	81.20	450.13	73.57	2134	2.60	0.44
Morocco	1,525	379.54	80.18	400.31	78.50	1386	2.65	0.39
Norway	2,361	469.80	66.21	486.96	73.23	2298	2.35	0.53
Oman	2,339	372.57	94.94	422.67	94.97	2254	2.67	0.37
Palestine	2,197	369.42	102.51	405.60	110.73	2084	2.51	0.49
Qatar	3,619	308.37	93.23	320.09	125.29	3444	2.50	0.52
Romania	2,152	456.34	102.16	458.06	90.17	2071	2.28	0.57
Russian Federation	2,287	512.58	82.61	530.59	76.63	2217	2.41	0.53
Saudi Arabia	2,121	329.03	75.43	403.37	77.23	1994	2.48	0.50

Table 1.3 (Continued)

		Math		Science			Attitude	
Country	N	Mean	SD	Mean	SD	N	Mean	SD
Scotland	2,016	484.80	79.37	493.51	81.96	1978	2.36	0.52
Serbia	2,026	485.46	88.47	471.65	84.27	1950	2.24	0.58
Singapore	2,304	591.07	92.42	564.84	103.63	2296	2.43	0.55
Slovenia	2,011	502.15	71.25	537.43	72.07	1977	2.20	0.54
Sweden	2,635	493.64	70.40	512.54	78.77	2459	2.34	0.54
Syria, Arab Republic of	2,304	396.07	82.95	453.70	74.72	2076	2.58	0.46
Thailand	2,732	441.03	92.78	469.67	83.65	2715	2.47	0.41
Tunisia	2,067	419.84	66.34	444.92	59.84	2001	2.57	0.48
Turkey	2,263	430.64	107.85	453.50	91.25	2203	2.52	0.48
Ukraine	2,282	460.10	90.75	482.99	85.71	2166	2.39	0.52
United States	3,784	508.17	76.63	520.19	82.13	3724	2.38	0.55

Notes: N = Sample size; Mean Math = average score for math achievement; SD Math = standard deviation for math achievement; Mean Sci = average score for science achievement; SD Sci = standard deviation for science achievement; Mean Att = average score for math attitude; SD Att = standard deviation for math attitude.

Table 1.4: Descriptive statistics for cross-validation sample.

		Math		Science			Attitude	
Country	N	Mean	SD	Mean	SD	N	Mean	SD
Algeria	2666	387.05	60.00	407.23	62.78	2459	2.67	0.39
Armenia	2315	498.37	84.62	488.47	100.21	1875	2.32	0.57
Australia	2021	495.57	81.05	516.19	81.67	1959	2.28	0.55
Bahrain	2090	397.38	81.89	467.31	84.87	2031	2.54	0.48
Bosnia and Herzegovina	2063	458.67	78.06	467.46	79.63	1946	2.29	0.57
Botswana	2148	364.34	77.21	355.42	100.48	2009	2.58	0.41
Bulgaria	2035	465.45	101.88	470.45	101.70	1932	2.28	0.59
Chinese Taipei	2001	598.03	104.80	561.31	88.83	1989	2.01	0.66
Colombia	2460	379.47	79.02	416.30	75.67	2369	2.60	0.42
Cyprus	2172	463.89	88.66	450.60	84.12	2135	2.37	0.57
Czech Republic	2399	504.33	75.06	540.16	72.69	2387	2.22	0.57
Egypt	3407	390.27	100.15	408.15	99.36	3088	2.69	0.39
El Salvador	1974	339.48	73.26	385.79	70.84	1902	2.55	0.40
England	2030	510.52	84.03	538.44	85.50	1980	2.38	0.53
Georgia	2137	411.79	96.50	421.83	83.02	1830	2.46	0.52
Ghana	2679	309.46	93.09	304.22	109.22	2519	2.63	0.39
Hong Kong, SAR	1757	572.00	94.80	529.40	80.77	1742	2.22	0.59
Hungary	2138	515.18	84.64	536.83	76.90	2113	2.22	0.56
Indonesia	2075	397.10	86.90	426.30	74.46	2046	2.59	0.36
Iran, Islamic Republic of	1929	404.36	84.30	459.74	79.34	1612	2.54	0.49
Israel	1639	463.66	99.62	468.32	102.29	1560	2.46	0.53
Italy	2256	480.96	76.45	496.76	77.94	2187	2.24	0.64
Japan	2123	570.54	84.92	554.16	77.75	2103	1.98	0.57
Jordan	2666	429.43	102.76	484.25	98.68	2521	2.66	0.43
Korea, Republic of	2160	598.45	90.83	553.73	74.72	2155	2.07	0.61
Kuwait	2041	354.36	78.87	418.75	89.57	1902	2.52	0.52
Lebanon	1873	450.31	73.97	415.84	93.91	1751	2.52	0.51
Lithuania	2012	502.79	81.06	516.02	78.75	1987	2.33	0.53
Malaysia	2236	476.18	80.02	473.92	89.03	2230	2.47	0.46
Malta	2333	486.59	91.41	455.33	113.46	2313	2.28	0.57
Mongolia	2157	430.49	81.77	448.41	73.54	1982	2.62	0.43
Morocco	1535	382.03	80.44	403.36	78.55	1382	2.66	0.39
Norway	2266	468.61	65.09	486.55	73.31	2181	2.34	0.54
Oman	2413	372.30	94.94	422.34	96.48	2306	2.67	0.37
Palestine	2181	364.90	102.31	402.66	111.10	2069	2.48	0.48
Qatar	3565	305.19	93.46	317.60	126.44	3399	2.51	0.53
Romania	2046	466.60	96.83	465.98	85.21	1983	2.31	0.57
Russian Federation	2185	510.83	83.56	528.48	78.71	2130	2.40	0.52
Saudi Arabia	2122	329.65	77.42	403.12	78.72	2003	2.47	0.49

Table 1.4 (Continued)

		Math		Science			Attitude	
Country	N	Mean	SD	Mean	SD	N	Mean	SD
Scotland	2054	489.97	79.99	497.92	80.21	2013	2.39	0.50
Serbia	2019	486.09	90.47	468.89	85.16	1944	2.23	0.58
Singapore	2295	594.51	93.46	569.69	104.10	2285	2.44	0.55
Slovenia	2032	500.81	71.97	537.66	71.96	1993	2.16	0.52
Sweden	2580	488.89	69.61	508.79	77.21	2430	2.32	0.55
Syria, Arab Republic of	2346	393.61	81.83	450.26	74.66	2097	2.59	0.45
Thailand	2680	441.75	90.41	471.58	81.78	2654	2.47	0.41
Tunisia	2013	421.01	66.69	444.88	61.12	1947	2.58	0.47
Turkey	2235	433.01	109.63	454.84	92.54	2162	2.55	0.46
Ukraine	2142	464.38	87.50	487.30	82.03	2016	2.41	0.52
United States	3593	508.76	76.84	519.78	82.42	3537	2.41	0.53

Notes: N = Sample size; Mean Math = average score for math achievement; SD Math = standard deviation for math achievement; Mean Sci = average score for science achievement; SD Sci = standard deviation for science achievement; Mean Att = average score for math attitude; SD Att = standard deviation for math attitude.

Table 1.5: Six initial cluster solutions for 48 countries in TIMSS 2007.

	Cluster Solution							Cluster Solution					
Country	9	8	7	6	5	4	Country	9	8	7	6	5	4
Japan	1	1	1	1	1	1	Cyprus	4	4	3	3	3	3
Chinese Taipei	1	1	1	1	1	1	Israel	4	4	3	3	3	3
Korea, Republic of	1	1	1	1	1	1	Malaysia	4	4	3	3	3	3
Hong Kong, SAR	1	1	1	1	1	1	Lebanon	5	4	3	3	3	3
Singapore	1	1	1	1	1	1	Thailand	6	5	4	4	4	3
Czech Republic	2	2	2	2	2	2	Jordan	6	5	4	4	4	3
Slovenia	2	2	2	2	2	2	Tunisia	6	5	4	4	4	3
Hungary	2	2	2	2	2	2	Turkey	6	5	4	4	4	3
England	2	2	2	2	2	2	Bahrain	6	5	4	4	4	3
Australia	2	2	2	2	2	2	Iran, Islamic Republic of	6	5	4	4	4	3
Russian Federation	2	2	2	2	2	2	Syria, Arab Republic of	6	5	4	4	4	3
Lithuania	2	2	2	2	2	2	Palestine	7	6	5	5	5	4
United States	2	2	2	2	2	2	Botswana	7	6	5	5	5	4
Sweden	2	2	2	2	2	2	Kuwait	7	6	5	5	5	4
Norway	3	3	3	3	3	3	El Salvador	7	6	5	5	5	4
Italy	3	3	3	3	3	3	Saudi Arabia	7	6	5	5	5	4
Scotland	3	3	3	3	3	3	Oman	8	7	6	5	5	4
Malta	3	3	3	3	3	3	Columbia	8	7	6	5	5	4
Serbia	3	3	3	3	3	3	Egypt	8	7	6	5	5	4
Armenia	3	3	3	3	3	3	Algeria	8	7	6	5	5	4
Ukraine	3	3	3	3	3	3	Georgia	8	7	6	5	5	4
Bulgaria	4	4	3	3	3	3	Indonesia	8	7	6	5	5	4
Bosnia and Herzegovinia	4	4	3	3	3	3	Qatar	9	8	7	6	5	4
Romania	4	4	3	3	3	3	Ghana	9	8	7	6	5	4

Table 1.6: Omnibus fit, DFA classification accuracy, and R^2 results for differences between clusters.

	Cluster Solution								
	9			8			7		
DFA *F* tests *(DF)*									
Pillai's Trace *F* test *(DF)*	2.19	*(72,304)*	***	2.43	*(63,217)*	***	3.06	*(54,186)*	***
Hotelling-Lawley *F* test *(DF)*	30.46	*(72,234)*	***	27.67	*(63,163)*	***	25.01	*(54,146)*	***
Roy's Greatest Root *F* test *(DF)*	296.56	*(9,38)*	***	239.90	*(9,31)*	***	174.47	*(9,31)*	***
Wilks' Lambda *F* test *(DF)*	5.78	*(72,196)*	***	6.08	*(63,146)*	***	7.04	*(54,137)*	***
ANOVA *F* tests *(DF)*									
DEMIND	4.059	*(8,39)*	**	4.623	*(7,40)*	***	5.291	*(6,41)*	***
HDEVIND	5.787	*(8,39)*	***	6.616	*(7,40)*	***	7.689	*(6,41)*	***
EDUCIND	7.756	*(8,38)*	***	8.606	*(7,39)*	***	10.085	*(6,40)*	***
ECONFRDM	1.938	*(8,38)*		2.197	*(7,39)*		2.602	*(6,40)*	*
PRESFRDM	6.675	*(8,39)*	***	7.803	*(7,40)*	***	9.272	*(6,41)*	***
GENDEQ	11.436	*(8,33)*	***	11.243	*(7,34)*	***	13.221	*(6,35)*	***
MACH	159.850	*(8,39)*	***	182.270	*(7,40)*	***	187.52	*(6,41)*	***
SACH	121.710	*(8,39)*	***	106.260	*(7,40)*	***	102.03	*(6,41)*	***
MATT	14.109	*(8,39)*	***	15.316	*(7,40)*	***	17.238	*(6,41)*	***
Additional DFA Information									
Overall DFA Accuracy		79 %			75 %			85 %	
Eigenvalue *(R^2)*	108.13	*(94.86)*		69.65	*(93.05)*		50.65	*(91.23)*	

	Cluster Solution								
	6			5			4		
DFA *F* test *(DF)*									
Pillai's Trace *F* test *(DF)*	3.25	*(45,155)*	***	2.99	*(36,124)*	***	4.11	*(27,93)*	***
Hotelling-Lawley *F* test *(DF)*	18.13	*(45,127)*	***	16.63	*(36,106)*	***	6.67	*(27,83)*	***
Roy's Greatest Root *F* test *(DF)*	99.16	*(9,31)*	***	71.97	*(9,31)*	***	17.75	*(9,31)*	***
Wilks' Lambda *F* test *(DF)*	6.68	*(45,123)*	***	6.35	*(36,106)*	***	5.19	*(27,85)*	***
ANOVA *F* tests *(DF)*									
DEMIND	6.678	*(5,42)*	***	3.721	*(4,43)*	*	4.646	*(3,44)*	**
HDEVIND	9.631	*(5,42)*	***	7.869	*(4,43)*	***	8.981	*(3,44)*	***
EDUCIND	12.756	*(5,41)*	***	8.781	*(4,42)*	***	13.416	*(3,43)*	***
ECONFRDM	3.008	*(5,41)*	*	3.588	*(4,42)*	*	4.897	*(3,43)*	**
PRESFRDM	10.148	*(5,42)*	***	2.565	*(4,43)*		3.184	*(3,44)*	*
GENDEQ	15.842	*(5,36)*	***	11.664	*(4,37)*	***	11.764	*(3,38)*	***
MACH	116.310	*(5,42)*	***	74.427	*(4,43)*	***	53.434	*(3,44)*	***
SACH	134.170	*(5,42)*	***	158.190	*(4,43)*	***	55.294	*(3,44)*	***
MATT	22.966	*(5,42)*	***	14.290	*(4,43)*	***	16.372	*(3,44)*	***
Additional DFA Information									
Overall DFA Accuracy		83 %			90 %			77 %	
Eigenvalue *(R^2)*	28.79	*(89.65)*		20.89	*(92.49)*		5.15	*(79.17)*	

Note: DEMIND = Democratic Index; HDEVIND = Human Development Index; EDUCIND = Education Index; ECONFRDM = Economic Freedom Index; PRESFRDM = Freedom of the Press Index, GENDEQ = Gender Equality Index; MACH = Math Achievement Score; SACH = Science Achievement Score; MATT = Math Attitude Score; ***, **, and * indicate significance at the p < .001, .01, and .05 levels, respectively.

Table 1.7: Countries in the 9-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine	7
4	Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia	6
5	Lebanon	1
6	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
7	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia	5
8	Oman, Columbia, Egypt, Algeria, Georgia, Indonesia	6
9	Qatar, Ghana	2

Table 1.8: DFA Classification results for the 9-cluster solution.

	Actual Cluster								
Predicted Cluster	1	2	3	4	5	6	7	8	9
1	5	0	0	0	0	0	0	0	0
2	0	9	1	0	0	0	0	0	0
3	0	0	4	0	0	0	0	0	0
4	0	0	2	3	0	0	0	0	0
5	0	0	0	3	1	0	0	0	0
6	0	0	0	0	0	7	0	0	0
7	0	0	0	0	0	1	6	1	0
8	0	0	0	0	0	0	1	2	0
9	0	0	0	0	0	0	0	2	0
Correctly Classified	5	9	4	3	1	7	6	2	0
Incorrectly Classified	0	1	0	2	3	0	2	1	2
Prior Probabilities	10.42%	18.75%	14.58%	12.50%	2.08%	14.58%	10.42%	12.50%	4.17%
Accuracy of Cluster	100%	100%	57%	67%	100%	100%	75%	67%	0%

Note: Prior probabilities are based on the model-building cluster solution.

Table 1.9: Countries in the 8-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine	7
4	Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon	7
5	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
6	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia	5
7	Oman, Columbia, Egypt, Algeria, Georgia, Indonesia	6
8	Qatar, Ghana	2

Table 1.10: DFA Classification results for the 8-cluster solution.

	Actual Cluster							
Predicted Cluster	1	2	3	4	5	6	7	8
1	5	0	0	0	0	0	0	0
2	0	9	1	0	0	0	0	0
3	0	0	3	3	0	0	0	0
4	0	0	3	4	0	0	0	0
5	0	0	0	0	7	1	0	0
6	0	0	0	0	0	6	1	0
7	0	0	0	0	0	1	2	2
8	0	0	0	0	0	0	0	0
Correctly Classified	5	9	3	4	7	6	2	0
Incorrectly Classified	0	1	3	3	1	1	3	0
Prior Probabilities	10.42%	18.75%	14.58%	14.58%	14.58%	10.42%	12.50%	4.17%
Accuracy of Cluster	100%	100%	43%	57%	100%	75%	67%	0%

Note: Prior probabilities are based on the model-building cluster solution.

Table 1.11: Countries in the 7-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine, Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon	14
4	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
5	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia	5
6	Oman, Columbia, Egypt, Algeria, Georgia, Indonesia	6
7	Qatar, Ghana	2

Table 1.12: DFA Classification results for the 7-cluster solution.

	Actual Cluster						
Predicted Cluster	1	2	3	4	5	6	7
1	5	0	0	0	0	0	0
2	0	9	1	0	0	0	0
3	0	0	13	0	0	0	0
4	0	0	0	7	1	0	0
5	0	0	0	0	6	1	0
6	0	0	0	0	1	1	2
7	0	0	0	0	0	1	0
Correctly Classified	5	9	13	7	6	1	0
Incorrectly Classified	0	1	0	1	1	3	1
Prior Probabilities	10.42%	18.75%	29.17%	14.58%	10.42%	12.50%	4.17%
Accuracy of Cluster	100%	100%	93%	100%	75%	33%	0%

Note: Prior probabilities are based on the model-building cluster solution.

Table 1.13: Countries in the 6-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine, Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon	14
4	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
5	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia, Oman, Columbia, Egypt, Algeria, Georgia, Indonesia	11
6	Qatar, Ghana	2

Table 1.14: DFA Classification results for the 6-cluster solution.

	Actual Cluster					
Predicted Cluster	1	2	3	4	5	6
1	5	0	0	0	0	0
2	0	8	2	0	0	0
3	0	1	11	0	1	0
4	0	0	0	7	2	0
5	0	0	0	0	8	1
6	0	0	0	0	1	1
Correctly Classified	5	8	11	7	8	1
Incorrectly Classified	0	2	2	2	1	1
Prior Probabilities	10.42%	18.75%	29.17%	14.58%	22.92%	4.17%
Accuracy of Cluster	100%	89%	85%	100%	67%	50%

Note: Prior probabilities are based on the model-building cluster solution.

Table 1.15: Countries in the 5-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine, Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon	14
4	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
5	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia, Oman, Columbia, Egypt, Algeria, Georgia, Indonesia, Qatar, Ghana	13

Table 1.16: DFA Classification results for the 5-cluster solution.

	Actual Cluster				
Predicted Cluster	1	2	3	4	5
1	5	0	0	0	0
2	0	9	0	0	0
3	0	0	18	2	0
4	0	0	0	9	0
5	0	0	0	1	2
Correctly Classified	5	9	18	9	2
Incorrectly Classified	0	0	2	0	1
Prior Probabilities	10.42%	18.75%	29.17%	14.58%	27.08%
Accuracy of Cluster	100%	100%	90%	75%	100%

Note: Prior probabilities are based on the model-building cluster solution.

Table 1.17: Countries in the 4-cluster solution.

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine, Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon, Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	21
4	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia, Oman, Columbia, Egypt, Algeria, Georgia, Indonesia, Qatar, Ghana	13

Table 1.18: DFA Classification results for the 4-cluster solution.

	Actual Cluster			
Predicted Cluster	1	2	3	4
1	5	0	0	0
2	0	8	3	0
3	0	1	17	3
4	0	0	4	7
Correctly Classified	5	8	17	7
Incorrectly Classified	0	3	4	4
Prior Probabilities	10.42%	18.75%	50.00%	20.83%
Accuracy of Cluster	100%	89%	71%	70%

Note: Prior probabilities are based on the model-building cluster solution.

Figure 1.1: Initial cluster dendogram for TIMSS 2007 data.

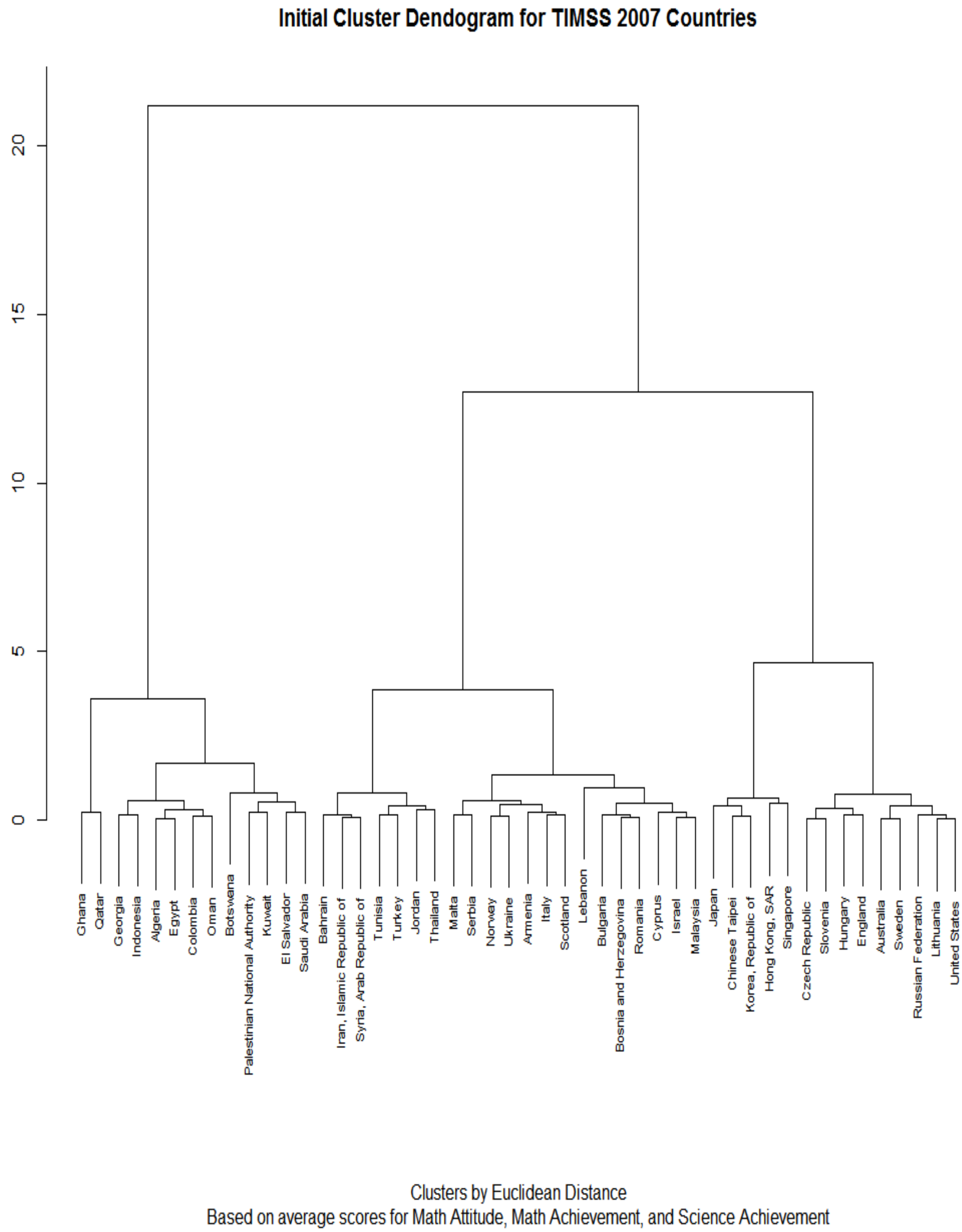


Figure 1.2: Dotplot of average math achievement by country.

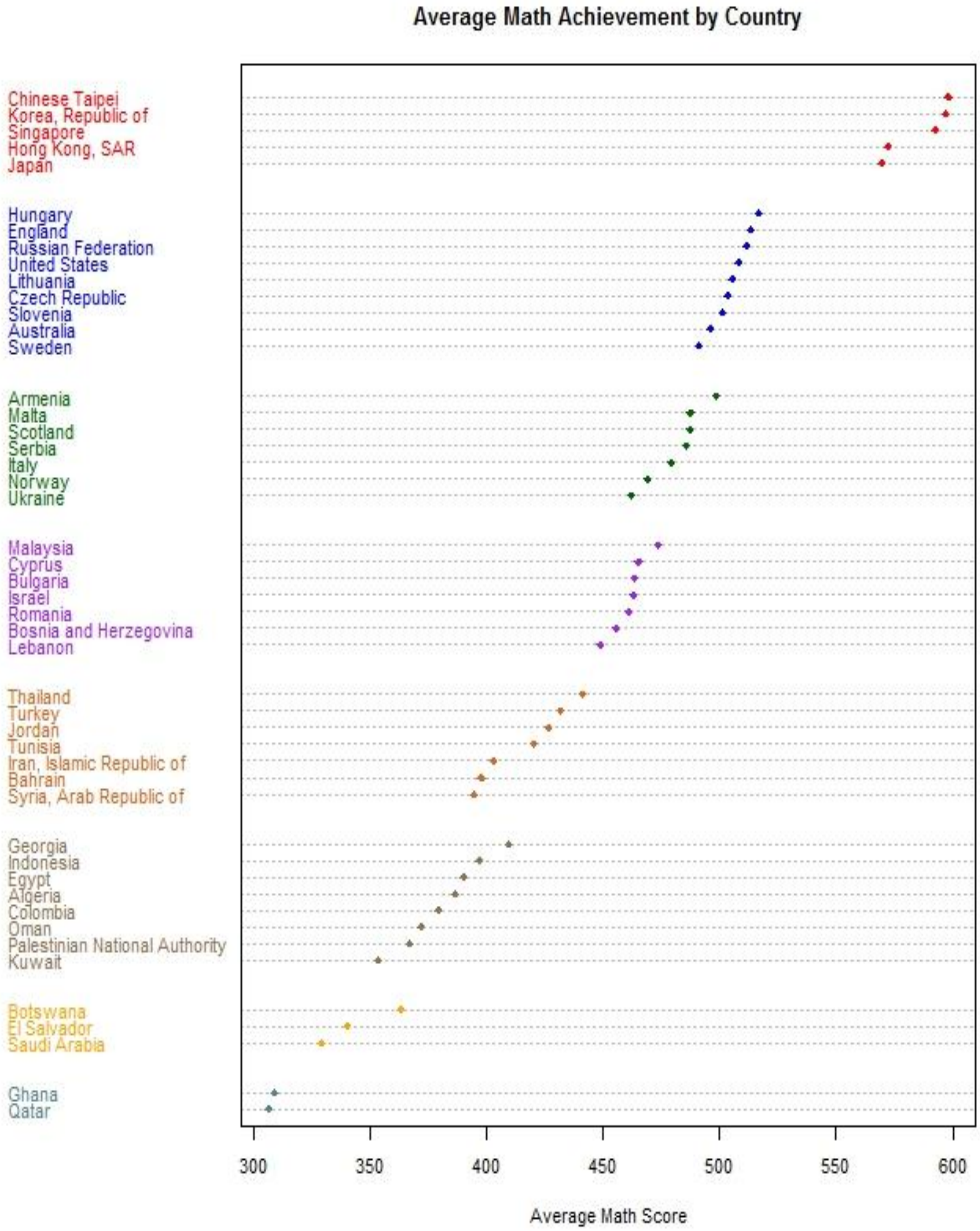


Figure 1.3: Dotplot of average math attitude by country.

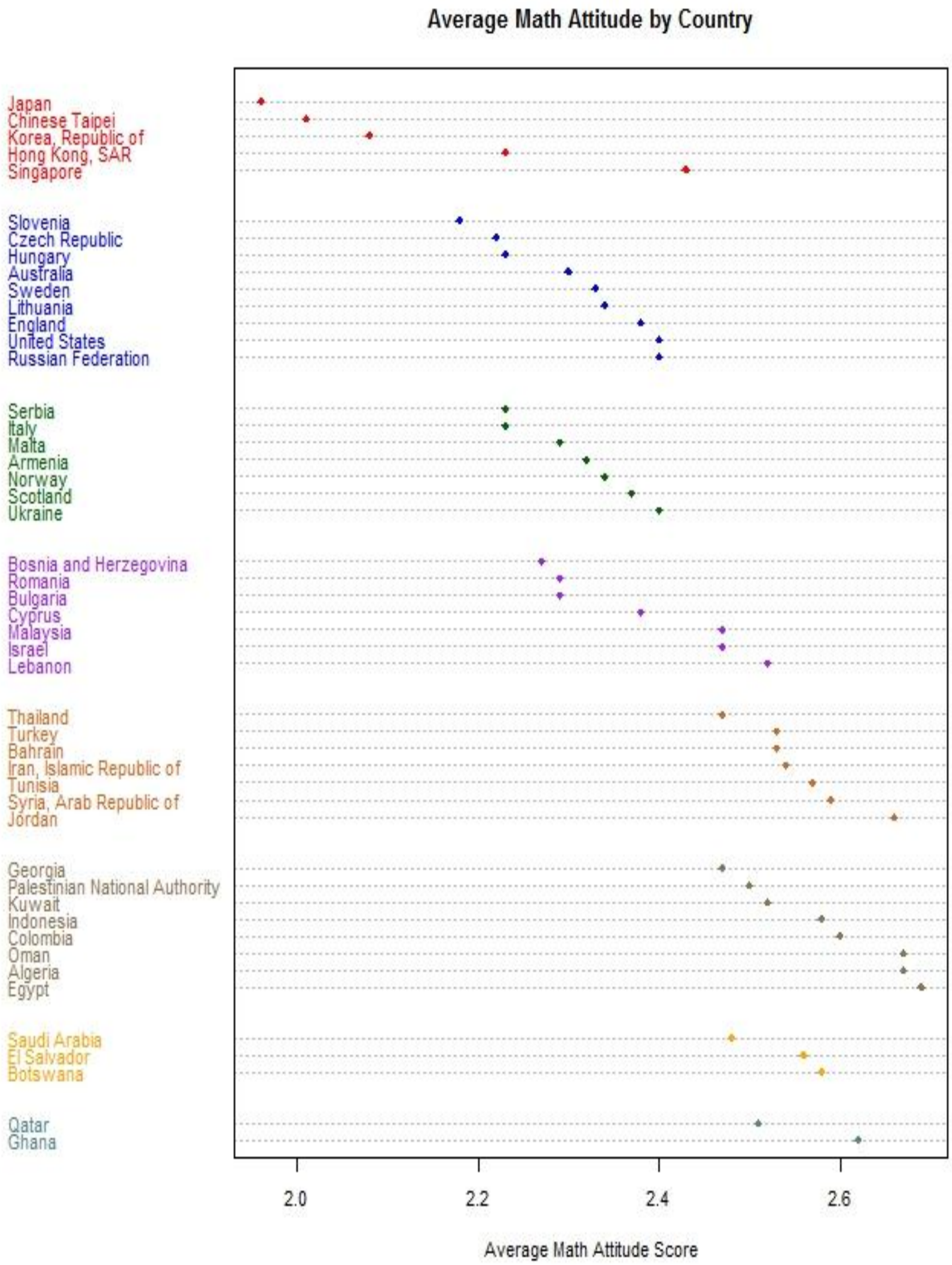
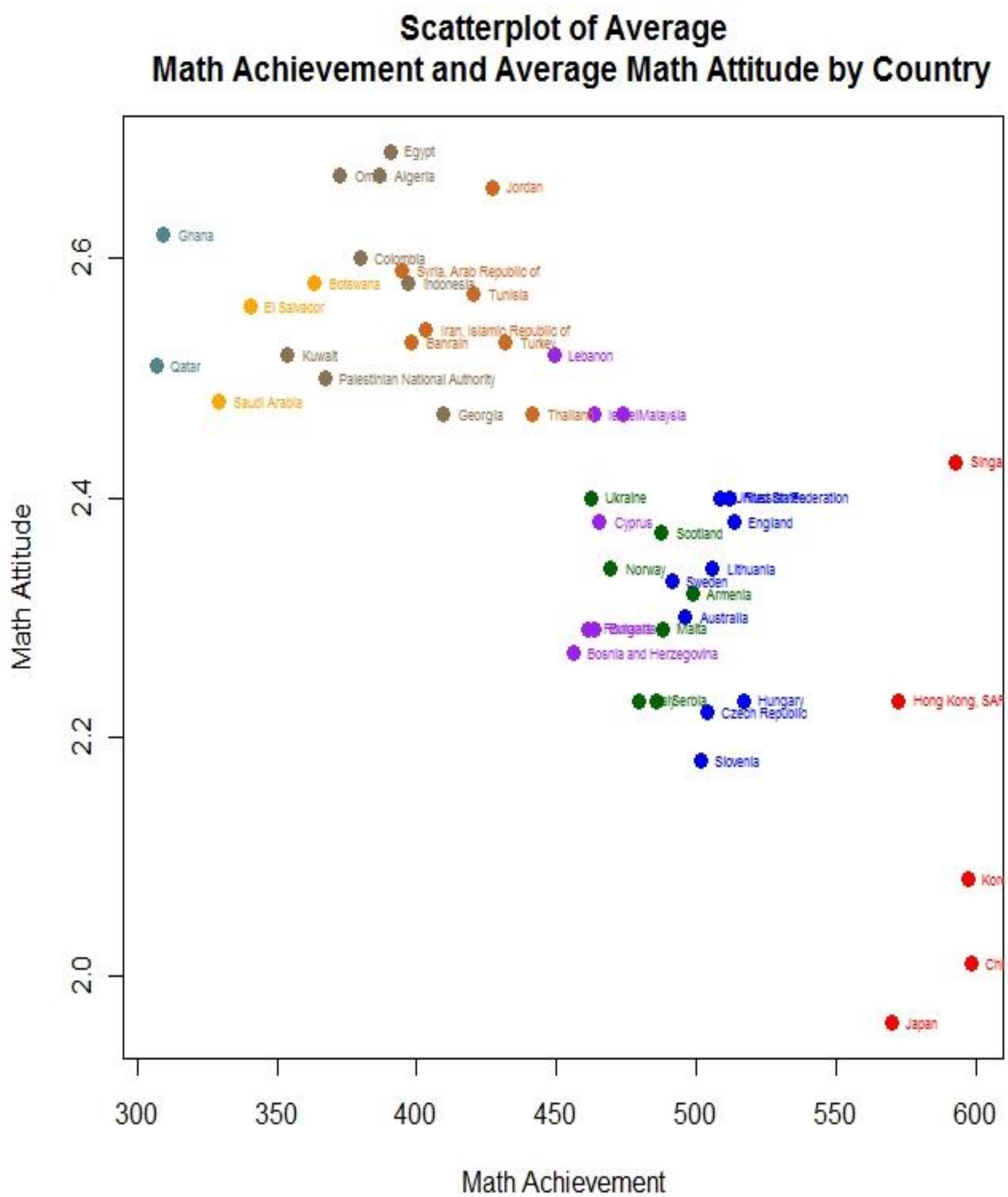


Figure 1.4: Scatterplot of math achievement and math attitude by country.



References

Aiken, L. R., & Dreger, R. M. (1961). The effect of attitudes on performance in mathematics. *Journal of Educational Psychology, 52*(1), 19-24.

Anttonen, R. G. (1969). A Longitudinal Study in Mathematics Attitude. *The Journal of Educational Research, 62*(10), 467-471.

Domestic Policy Council, Office of Science and Technology Policy (2006). *American competetiveness initiative: Leading the world in innovation*. Retrieved from http://www.innovationtaskforce.org/docs/ACI%20booklet.pdf on November 5, 2012.

Economist Intelligence Unit (2011). Democracy Index 2011: Democracy Under Stress: Economist Intelligence Unit.

Eid, M., Langeheine, R., & Diener, E. (2003). Comparing typological structures across cultures by multigroup latent class analysis. A primer. *Journal of Cross-Cultural Psychology, 34,* 195-210.

Foy, P., Galia, J., & Li, I. (2007). Scaling the data from the TIMSS 2007 mathematics and science assessments. In J. F. Olson, Martin, M. O., & Mullis, I.V.S. (Ed.), *TIMSS 2007 Technical Report*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. Boston: TIMMS & PIRLS International Study Center, Lynch School of Education, Boston College.

Glenn, J. (2000). *Before it's too late: A report to the nation*. Washington, DC: Department of Education. Retrieved from http://www.nationalmathandscience.org/sites/default/files/resource/before%20its%20too%20late.pdf on September 20, 2012.

Harlow, L. L. (2005). *The essence of multivariate thinking: Basic themes and methods*. Mahwah, NJ: Erlbaum.

Harlow, L. L., Burkholder, G. J., & Morrow, J. A. (2002). Evaluating attitudes, skill, and performance in a learning-enhanced quantitative methods course: A structural modeling approach. *Structural Equation Modeling Journal*, *9*, 413-430.

The Heritage Foundation (2011). *The index of economic freedom*. Washington, DC: The Heritage Foundation.

Human Development Report (2011). *Sustainability and equity: A better future for all*. New York, NY: United Nations Development Programme.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education*, *32*(5), 448-470.

Koretz, D. (2009). How do American students measure up? Making sense of international comparisons. *The Future of Children*, *19*(1), 37-51.

Kuenzi, J. J., Matthews, C. M., & Mangan, B. F. (2006). *Science, technology, engineering and mathematics (STEM) education issues and legislative options.* Report of the Congressional Research Service, Library of Congress No. RL33434.

Lott, B. E., & Bullock, H. E. (2007). *Psychology and economic injustice: Personal, professional, and political intersections*. Washington, DC: American Psychological Association.

Ma, X. (1997). Reciprocal relationships between attitude toward mathematics and achievement in mathematics. *The Journal of Educational Research, 90*(4), 221-229.

Morren, M., Gelissen, J., & Vermunt, J. (2012). The impact of controlling for extreme responding on measurement equivalence in cross-cultural research. *Methodology, 8*(4), 159-170.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eight grades*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Research Council (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* Washington, DC: National Academies Press.

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.

Poortinga, Y.H., & van de Vijver, F.J.R. (1987). Explaining cross-cultural differences: Bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18,* 259-282.

Punj, G., & Stewart, D. W. (1983). Cluster analysis in marketing research: Review and suggestions for application. *Journal of Marketing Research, 20*(2), 134-148.

Reporters without Borders (2012). *World Press Freedom Index 2011-2012*. Paris: Reporters Without Borders.

Schmidt, W. H., & McKnight, C. C. (1998). What can we really learn from TIMSS? *Science*, *282*(5395), 1830-1831.

Schreiber, J. B. (2002). Institutional and student factors and their influence on advanced mathematics achievement. *The Journal of Educational Research, 95*(5), 274-286.

Sherman, L., & Sheth, J. N. (1977). Cluster analysis and its applications in marketing research. In J. N. Sheth (Ed.), *Multivariate Methods for Market and Survey Research*. Chicago: American Marketing Association.

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.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics, 5th Ed.* Boston, MA: Allyn & Bacon.

Tukey, J.W. (1977). *Exploratory data analysis.* Boston, MA: Addison-Wesley.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236-244.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., . . . Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide*.

Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

CHAPTER 2

An International Investigation of the Relationships between Gender, Socioeconomic Status, Math Attitude, and Math Achievement

To be submitted to the *International Review of Education* on December 21, 2012

An International Investigation of the Relationships between Gender, Socioeconomic Status, Math Attitude, and Math Achievement

Despite over three decades of research the debate surrounding sex-based differences in mathematics performance continues. Although much of the current research on the topic suggests that gender differences in mathematics achievement have disappeared in the United States (Else-Quest, Hyde, & Linn, 2010; Hyde, 2005; Lachance & Mazzocco, 2006; Lindberg, Hyde, Petersen, & Linn, 2010; Liu & Wilson, 2009), there is some evidence that gender differences in math achievement persist outside of the United States (Kim & Law, 2012; Köller, Baumert, & Schnabel, 2001; Ngware, Ciera, Abuya, Oketch, & Mutisya, 2012).

In the United States there is a national push to promote education in the subjects of science, technology, engineering, and mathematics, commonly referred to as STEM disciplines (see the American Competitiveness Initiative published by the Domestic Policy Council, 2006). STEM education is widely seen as being a crucial foundation for the future of the United States (Kuenzi, Matthews, & Mangan, 2006), and there is a belief that “the future well-being of our nation and people depends not just on how well we educate our children generally, but on how well we educate them in mathematics and science specifically” (Glenn, 2000, p. 4).

At least part of the push for increased focus on STEM education has been caused by the findings of multinational assessments like the TIMSS (Trends in International Mathematics and Science Study) and PISA (Program for International Student Assessment), where the United States has been performing sluggishly in such assessments for decades (Koretz, 2009; Schmidt & McKnight, 1998). Most of the findings from these studies are based on mean differences or regression analyses, and more complex evaluations have rarely been reported, possibly due to the complexity of the data. In the increasingly global communities and economies that are being developed, rigorous multivariate research needs to be conducted to look at these relationships on an international scale in order to understand such relationships from a global and more overarching perspective.

Additionally, the underrepresentation of women in STEM disciplines has garnered considerable attention (Beede, Julian, Langdon, McKittrick, Khan, & Doms, 2011). Köller, Baumert, and Schnabel (2001) found that attitude toward mathematics had an effect on course selection, and that female students were less enthusiastic about and less likely to enroll in advanced mathematics courses than their male counterparts. Because STEM disciplines require an understanding of upper-level mathematics, female students who have shied away from increasingly challenging mathematics courses are less likely to have had the prerequisites for upper-level math.

The Relationship between Socioeconomic Status and Math Achievement

Williams, Williams, Kastberg, and Jocelyn (2005) state that the relationship between academic achievement and socioeconomic status is so vital that few studies would enter publication without considering SES as a construct. Sirin's (2005) meta-analysis of articles published between 1990 and 2000 provides strong support for the strength of this relationship; the results of this meta-analytic study support the belief of many researchers that SES is one of the most important factors to consider in all areas of education achievement, and it may be the most widely used variable in education research.

In terms of mathematics achievement specifically (as opposed to academic achievement in general), White's (1982) meta-analysis of 143 studies suggested the existence of a small, positive correlation ($r = .20$) between SES and math achievement when SES was measured at the student level. Similar conclusions were reached by Welch, Anderson and Harris (1982) and Yando, Seitz, and Zigler (1979).

The Relationship between Math Attitude and Math Achievement

The relationship between attitude toward math and math achievement is generally expressed as a positive correlation – the higher a student's attitude toward math is, the better the student will do in math classes and on math assessments (Harlow, Burkholder & Morrow, 2002; Schreiber, 2002). This relationship has been investigated for decades (see Aiken & Dreger, 1961; and Anttonen, 1969 for some early work), and continues to be important to consider because cognitive ability and background variables (such as SES) are difficult to change, whereas affective variables can be more easily targeted for intervention (Singh, Granville, & Dika, 2002).

The Relationships between Sex, Math Achievement and Math Attitude

Some early investigations into sex-based differences in math performance painted a stark picture; "huge sex differences" were reported (Benbow & Stanley, 1980), with females on the losing end of the comparison. However, as Rossi (1983) pointed out, effect sizes that account for less than 5% of the explained variance should not be characterized in terms such as "huge", "large", "substantial", or "sizeable". Indeed, even before 1980 researchers were characterizing sex differences in math performance as small (i.e., Fennema & Sherman, 1977), a position supported by many researchers then and now (Felson & Trudeau, 1991; Fennema, Peterson, Carpenter, & Lubinski, 1990; Joffe & Foxman, 1984; Lieu & Wilson, 2009; Lindberg, Hyde, Petersen, & Linn, 2010; Linn and Hyde, 1989; Rossi, 1983).

Sex-based differences in math achievement are likely not attributable to differences in ability. Rather, as could be inferred from Bandura's self-efficacy theory (Bandura, 1997), the differences in self-efficacy levels between boys and girls in math related topics may explain much of the sex-based differences seen in math achievement, and there are examples in the literature that support this interpretation. Ethington (1992) demonstrated the importance of the student's attitude toward math as a key component in the sex-based difference in performance and that the value placed on math was more influential for boys, but indirect psychological influences such as math affect were more influential for girls. Additionally, Casey, Nuttall, and Pezaris (2001) demonstrated that math self-confidence significantly mediated the relationship between gender and performance. Further, Köller, Baumert, and Schnabel (2001) found that students who are disinterested in math lack the motivation to learn the subject, whereas those who are highly interested often challenge themselves by selecting more advanced math courses, which in turn leads to higher learning rates and a deeper understanding of concepts.

A recently conducted study (Duerr & Harlow, 2011) investigated the differences between sex, SES, math attitude, and math achievement using the 8th grade United States participant data from TIMSS 2007. The findings from that study were: 1) sex-based mean differences in the math achievement latent construct were not statistically significant; 2) sex-based path coefficient differences between the SES and math achievement constructs, and between the math attitude and math achievement constructs were statistically significant but small in magnitude; 3) factor loadings for the math attitude construct were different depending on participant sex, with math affect and value of math loading more highly for female participants and math self-confidence loading more highly for male participants.

The Current Study

That sex-based differences in math achievement have receded in the United States is widely accepted by researchers. However, this is not necessarily the case for all countries. Furthermore, what remains unclear is the relationship between math attitudes and achievement, specifically whether these relationships differ across diverse cultures. Because the world is becoming more of a global community, it is time to start looking at topics like this from a multinational perspective.

The purpose of this study was to extend the work from Duerr and Harlow (2011) to a multinational scale. Using the TIMSS 2007 data and through the use of multiple-group latent variable modeling, relationships between gender, SES, math attitude, and math achievement were investigated between and within clusters of 44 participating countries from TIMSS 2007. It was hypothesized that the math attitude variables would have strong, positive relationships with math achievement across countries, but that the strength of the relationships would change depending on cluster membership and depending on the participant's country of origin.

Additionally, it was hypothesized that sex-based differences in math achievement would be small in magnitude, but that gender differences would be seen among the math attitude variables. Specifically, based on the findings of Ethington (1992) it was hypothesized that self-confidence and math affect would have a stronger relationship with math achievement for female participants than for male participants, while the value of math would have a stronger relationship with math achievement for male participants than for female participants. It was further hypothesized that the magnitude of these differences would change depending on the participant's country of origin.

METHODS

Participants

The sample for this study is comprised of students from countries that participated in TIMSS 2007 at the 8th grade level; the sample sizes for each country are broken down by gender in Table 2.1 below. Although the TIMSS 2007 data set contains information on 225,277 students from 50 countries and territories at the 8th grade level, not all countries could be included in the present analyses, for three reasons: 1) TIMSS administrators report that Morocco and Mongolia violated sampling procedures (Foy & Olson, 2009; Mullis, Martin, & Foy, 2008); 2) England and Scotland did not collect data on parental education level and cannot be included; and 3) Qatar and Ghana, which comprise their own cluster, did not have values for one of the math achievement variables (math reasoning) in the TIMSS 2007 data set and cannot be included; (how countries were organized into clusters is briefly discussed below, and is described in detail in Duerr & Harlow, under review). The final overall sample size for countries included in the present study was 197,155 8th grade students from 44 countries and territories.

INSERT TABLE 2.1 APPROXIMATELY HERE

Measures

Math Achievement.

Math achievement is represented as a latent variable comprised of three continuous indicator variables from the TIMSS 2007 data. The three indicator variables represent three different cognitive domains for math achievement: Math Knowing, Math Applying, and Math Reasoning. Descriptive statistics for these three indicator variables for the countries included in this study can be seen in Table 2.2. TIMSS math achievement scores are built to have a mean of 500 with a standard deviation of 100. Because latent variable modeling works best when all indicator variables are on a similar scale, even when the indicators are categorical (L. K. Muthén & Muthén, 2009), the raw scores for the math achievement indicator variables were divided by 100 so these values would be comparable to the scales used for attitudes, discussed below.

<u>INSERT TABLE 2.2 APPROXIMATELY HERE</u>

Math Attitude.

Math attitude is represented in this study as three separate latent variables: Math Self-Confidence, Positive Math Affect, and Value Placed on Math. These latent variables are each comprised of student responses to several attitude-oriented questions. The latent variables and their indicator questions are shown in Table 2.3. The responses for the attitude variables are on a 4-point Likert-like scale, with low values representing high endorsement and vice-versa. As an example, one question for Positive Math Affect is “I like math”, with responses ranging from 1 (agree a lot) to 4 (disagree a lot). To simplify the interpretation of results in the latent variable model, the scale scores were reversed by subtracting the score from 5. This yielded scores ranging from 1 to 4 with lower scores indicating low endorsement for the item.

INSERT TABLE 2.3 APPROXIMATELY HERE

Socioeconomic Status.

Socioeconomic status is represented in this study as a single latent variable comprised of three ordinal indicators. The three questions asked of students to assess socioeconomic status (SES) had to do with parental education level and the number of books in the student's home. For all three of these questions, low responses are indicative of low SES, and high responses are indicative of higher SES.

In the TIMSS data sets, the number of books in the home is on a 5-point response scale, with the lowest response being "None or very few (0 to 10 books)" and the highest response being "Three or more bookcases (over 200 books)." Parent education level is assessed on the International Standard Classification of Education (ISCED) scale, which ranges from 0 (no education) to 6 (Doctoral or professional degree). One difficulty when using the ISCED scale to measure SES in the TIMSS data is a lack of response to the questions for a number of participants, or the selection of the 8th option, which is "I don't know". This option was treated as nonresponse for the purpose of this analysis, and is discussed in more detail below.

Gender

Gender is a dichotomous variable, with female participants coded as 1 and male participants coded as 2.

Procedure

This study investigates differences between and among clusters of countries through a multiple-group structural equation modeling (SEM) framework. As such, there are three components to the analysis that need to be discussed: 1) the latent variable model being investigated, 2) how the countries were separated into clusters, and 3) how the clusters were used to conduct the multiple-group analyses.

The Latent Variable Model.

The path diagram shown in Figure 2.1 illustrates the latent variable model used to investigate the relationships between Gender, Math Attitude, and Math Achievement, with SES included as a covariate. As the figure shows, Math Achievement is a latent variable comprised of the three previously discussed continuous indicator variables from the TIMSS data set. Math Attitude is represented by three separate latent variables comprised of three or four ordinal indicator variables. SES is a single latent variable comprised of three ordinal indicator variables, and Gender is a dichotomous variable.

Identifying Clusters of Countries.

Multiple-group structural equation modeling is unable to incorporate a large number of groups, and could certainly not handle the number of countries included in this study. In order to make the multiple-group procedure possible, the countries in the TIMSS data set needed to be separated into smaller groups or clusters. To accomplish this, a hierarchical cluster analysis was conducted in a previous study (Duerr & Harlow, under review), which allowed for the identification of several clusters of countries. Clustering was based on country-level mean TIMSS math and science achievement scores and math attitude values.

Of the cluster solutions identified in Duerr and Harlow (under review), a variant of the 8-cluster solution was chosen for the present study primarily because this solution's clusters all contain fewer than 10 groups, which is the point after which multiple-group analysis becomes infeasible (L. K. Muthén, 2011); and the clusters appeared conceptually or geographically cohesive. The cluster membership for this cluster solution can be seen in Table 2.4. It should be noted that although this has been referred to as the 8-cluster solution in Duerr and Harlow (under review), the cluster comprised of Qatar and Ghana has been excluded due to the previously discussed missing data, yielding 7 clusters for the present analyses.

INSERT TABLE 2.4 APPROXIMATELY HERE

Between- and Within-Cluster Analyses.

To investigate the performance of the latent variable model across countries, a series of multiple-group comparisons was conducted. The first analysis utilized the data for all 44 countries in the study, with cluster membership acting as the grouping variable for a multiple-group analysis. This analysis provides an overall assessment of the model's performance at the international level, with clusters of countries being compared with other clusters of countries that are similar within their respective clusters.

Following this initial multiple-group analysis a series of additional multiple-group analyses was conducted for the three clusters deemed to be the most salient for U.S. policymakers, providing an assessment for the model's performance within clusters. For these analyses, countries are the grouping variable within a cluster, such that countries which have been identified as being similar to each other are compared with each other.

Data Analysis

Multiple-Group Analyses.

This study employs multiple-group structural equation modeling methodology and follows procedures outlined in Kline (2011) and L. K. Muthén and Muthén (2009). Multiple-group analysis involves testing a hypothesized model over several steps, with each step placing an additional constraint on parameters within a specified model. As outlined by L. K. Muthén and Muthén (2009), the model is first tested with all parameters freely estimated. Constraints are then applied to factor loadings and regression weights, then to factor variances, then to covariances, and ending with the model in which all parameters are constrained. Through each step of this process model fit is assessed via a chosen fit index where significant changes in the fit index indicate that the model with more constraints does not fit as well as the model with fewer constraints, which indicates a lack of invariance among the groups.

Which fit indicators should be used when conducting any structural modeling analysis has been the subject of much research (Bentler, 1990; Hu & Bentler, 1999; Steiger, 2000; Steiger & Lind, 1980), but there is some consensus in regard to the use of chi-square as a measure of model fit; due to chi-square's susceptibility to sample size and SEM's reliance on large samples, the use of chi-square as a test of model fit is generally accepted as being a poor choice. Several alternative fit indices have been proposed over the years, with some of the more popular choices being Root Mean Squared Error of Approximation (RMSEA, Steiger & Lind, 1980) and the

Comparative Fit Index (CFI, Bentler, 1990). When using RMSEA as a measure of model fit, values greater than 0.10 are considered to have poor fit, and RMSEA values of 0.08 and 0.05 represent acceptable and good fit, respectively (Browne & Cudeck, 1993). CFI values less than 0.90 are generally considered an indication of poor fit, whereas values of 0.95 and higher represent good fit (Hu & Bentler, 1999). It should be noted that the use of cutoff values in the determination of model fit is the subject of debate, as can be seen in Chen, Curran, Bollen, Kirby, and Paxton (2008).

Cheung and Rensvold (2002) investigated the performance of several Goodness of Fit Indices (GFIs) within the context of multiple-group comparisons. Cheung and Rensvold demonstrated that, of the available GFIs, ΔCFI is one of the superior choices for assessing model fit in multiple-group analyses because 1) the correlation between CFI and ΔCFI is small, indicating independence between the overall fit and incremental fit; 2) ΔCFI is not significantly affected by model complexity; and 3) CFI is the most frequently used goodness-of-fit (GFI) indicator, and reporting other GFIs would be redundant. Cheung and Rensvold (2002) do note that ΔGamma-hat and ΔMcDonald's Noncentrality Index are not redundant with ΔCFI and could be reported; however, Mplus does not offer these indices in conjunction with the imputation procedures required for the present study and thus just ΔCFI is used.

Accommodating TIMSS Complexity.

Like the overall math achievement score, the math achievement indicator variables for TIMSS are plausible values. A detailed review of the complexities associated with the TIMSS data collection process, including the use of plausible values for large-scale assessment, is provided in Appendix 1, and a comprehensive explanation can be found in Foy, Galia, and Li (2007). To briefly summarize the concept of plausible values, a single participant is subjected to only a portion of all possible test items from the TIMSS test bank. The student's responses, along with his or her background characteristics, are used to create a series of possible overall scores for the student via imputation; these possible scores are called plausible values. The use of plausible values introduces additional error into any statistical model, which must be accommodated in the analysis. To accommodate this source of error each analysis is conducted once for each plausible value, and the results are averaged. The TIMSS data set includes five calculated plausible values for the overall math achievement score and for the three cognitive domain indicator variables.

An additional complexity associated with the large-scale assessments like TIMSS comes in the form of the sampling design, which employs stratified cluster sampling. A thorough discussion of the sampling procedure for TIMSS assessments can be found in Foy and Olson (2009), and a detailed summary can be found in Appendix 1. A brief overview of the procedure is that schools are randomly selected according to certain regional characteristics, and then two classrooms are randomly selected from each chosen school. The students within these classrooms are then assessed, forming the sample. This sampling design introduces another source of error that must be acknowledged and accommodated in any subsequent analysis.

Accommodating the error associated with the sampling design is accomplished through the use of jackknife repeated replication (JRR). Using JRR, several replicates of the original sample are created and compared with the original sample; the variation between the estimate for the original sample and the estimate for the replicate becomes the jackknife estimate of sampling error for the statistic. Using these jackknife estimates allow for the creation of replicate weights found in the TIMSS data set. These replicate weights must be used to estimate a given parameter estimate, with one estimate for each of the 75 replicate weights and one estimate for the original sample for a total of 76 estimates of the parameter (a detailed explanation of the analysis considerations, including a discussion on the source of the 76 estimates of the parameters, can be found in Appendix 1).

A final methodological accommodation associated with using the large-scale data is the use of sample weights. The use of weighting variables forces a given country's sample to be representative of the country's population in terms of the background variables used for participant selection. Sampling weights for the TIMSS are values assigned to each participant based upon that participant's actual probability of having been selected for participation; this probability is based on the sample selection characteristics associated with the participant's school and classroom, which are known values due to the methodology used for sample selection (Foy & Olson, 2009).

Missing Data.

One issue encountered when using large-scale data sets is the inevitable existence of missing data, and the TIMSS data set is no exception to this. Missing data can take many forms, and the source of missingness will determine how it should be addressed. In some cases, the responses may be missing because it was not asked of the participants; this is the case for parental education information for England and Scotland in TIMSS 2007. In this situation it is difficult to justify the use of any imputation procedure to reconcile the missing data, and the observations for those countries must be excluded from the analysis.

Another source of missing data for the TIMSS is that caused by the balanced incomplete block assessment, which asks each participant to complete only a selection of the items from the test bank. This is systematically missing data, which is accommodated through the jackknife repeated replication procedure discussed previously.

Finally, there is participant nonresponse, where the participant simply does not answer the question. This is not problematic for TIMSS achievement score data, because the calculation of plausible values accommodates this type of missingness. However, this can be problematic for background variables, such as the questions regarding math attitude and socioeconomic status. Additionally, the “I don’t know” response option for parental education level offers no more information than does leaving the question unanswered.

Although many different approaches for handling missing data have been advocated over the years, including mean substitution, nearest neighbor substitution, and listwise deletion, it is generally agreed that these other methods are inadequate and lead to inaccurate parameter estimates. Fortunately, the increase in processing power available in computers has led to new methods in the accommodation of missing data in an analysis; for structural modeling analyses involving categorical indicators, weighted least squares is the preferred method (B. Muthén, 1984; B. Muthén & Satorra, 1995; L. K. Muthén & Muthén, 2009), and is the method employed in the current study.

RESULTS

Overview

The bulk of the present study is interpreted in terms of regression coefficients between latent variables. Regression coefficients provide a researcher with an effect size for a given relationship between two variables and, when interpreted in conjunction with their accompanying *t*-statistic, two specific questions can be answered: is the effect meaningful in terms of magnitude, and is the effect meaningful in terms of statistical significance? I have attempted to interpret the findings in this study from within this paradigm, but in order to do so I should clarify what my decision-making process was.

In terms of magnitude, it is often difficult to determine what qualifies as an important finding. Taking the inferred advice of Rossi (1983), effect sizes that explain less than 5% of the variance in math achievement research will be reported but dismissed as being inconsequential to the discussion. In terms of statistical significance, it should be acknowledged that the present study relies on very large sample sizes; even at the country level the samples consist of thousands of participants of both genders, in which case the sample sizes may promote statistical significance of even very small effects. For this reason, statistical significance will be reported at the $p < 0.05$, $p < 0.01$, and $p < 0.001$ levels, but *p*-values greater than 0.01 will not be discussed in great detail.

A final note on the interpretation of regression coefficients has to do with which coefficients to report. Several sources, including Harlow (2005); Kline (2011); and Pedhazur (1997); state that unstandardized coefficients are preferable to standardized coefficients when making comparisons across samples. However, Montgomery, Peck, and Vining (2006); and Tabachnick and Fidell (2007) state that standardized regression coefficients can be used, cautiously, to judge the magnitude of a given variable's effect relative to other variables in the model; the primary hazard of doing so manifests when subsequent samples have a different range of responses, which will affect the associated coefficients.

Based on this information, the present study reports both standardized and unstandardized regression coefficients. Unstandardized regression coefficients are discussed when effects are compared across groups (i.e., when assessing differences in the magnitude of the effect of self-confidence on math achievement between Australia and the United States), while standardized regression coefficients are interpreted when the magnitude of a variable's effect is being compared with the effects of other variables within the same group (i.e., when assessing differences in the magnitude of the effects for self-confidence and value of math in the United States).

Between-Clusters Analysis

The between-clusters analysis began by testing the model with all parameters freely estimated. The model was assessed for goodness of fit at this step by examining RMSEA and CFI values; the RMSEA of 0.056 and CFI of 0.94 for this step are both within the generally acceptable parameters for good model fit.

INSERT TABLE 2.5 APPROXIMATELY HERE

Constraints were then placed on factor loadings and regression coefficients, with the change in the CFI indices (ΔCFI) reviewed to determine whether or the models were invariant. Following recommendations in Cheung and Rensvold (2002), a ΔCFI value greater than 0.01 was considered to be evidence of a lack of invariance. The ΔCFI for this first step (0.014) is greater than the 0.01 threshold advocated by Cheung and Rensvold, suggesting that there is a lack of invariance between the clusters of countries in this comparison; this leads to the conclusion that the model does not perform equally across the clusters. Because the only constraints on the model for this step were the constraints placed on the factor loadings and regression weights, it can be concluded that the differences in model performance at the cluster level can be attributed to at least one or more of these loadings and weights; i.e., the regressions and loadings show differences between the clusters of countries. The standardized factor loadings for the between-clusters analysis can be seen in Table 2.6, and the unstandardized regression weights can be seen in Table 2.7.

INSERT TABLE 2.6 APPROXIMATELY HERE

INSERT TABLE 2.7 APPROXIMATELY HERE

The primary focus for the purpose of this study is on the regression weights between the latent constructs, specifically between the math attitude constructs and math achievement. Based on the standardized regression coefficients the conclusion can be drawn that, of the variables included in the model, math self-confidence had the largest relationship with math achievement for all clusters; the magnitude of this relationship is considerably larger than that for nearly any other variable in the model, including gender and SES.

Also of note is the nature of the relationship between Value of Math and Math Achievement. The relationship between value of math and math achievement increases as average math achievement scores within clusters decrease; that is to say, the magnitude of this relationship is miniscule (less than 1% of the variance accounted for according to the standardized regression weight) for the clusters that score above the international average, but it may be large enough to be considered an important contributor (accounting for 4% to 10% of the variance) in clusters that perform more poorly.

The relationship between Positive Math Affect and Math Achievement is confusing. Positive Math Affect would appear to have a negative relationship with Math Achievement, suggesting that students with higher math affect perform more poorly. This is counterintuitive to say the least; previous research has repeatedly shown that this relationship is positive in direction, and that students who enjoy learning about mathematics outperform those who do not (see Ryan & Deci, 2000; Lepper, Corpus, & Iyengar, 2005). There are some statistical anomalies that could be present that would influence the directionality of this variable in the model. For instance, given the correlations between the latent variables (shown in Table 2.7), the argument could be made that multicollinearity among the latent constructs is the principle culprit; the correlation between Positive Math Affect and Math Self-Confidence is quite large at 0.878. This could cause the magnitude and directionality of relationships with these variables to be unstable.

INSERT TABLE 2.7 APPROXIMATELY HERE

Gender's relationship with math achievement, while statistically significant across clusters, was consistently small in magnitude and explained less than 1% of the variance in most clusters. This is a finding consistent with much of the contemporary research on gender differences in mathematics achievement, including a comprehensive meta-analysis on the topic by Lindberg, Hyde, Petersen, and Linn (2010). What have not been studied in great detail are potential international differences regarding gender and math attitude. In terms of the present study's between-clusters analysis, the relationship between gender and math attitude is minute in most cases, although there are a few noted exceptions. For example, the relationship is rather larger for Cluster 4's countries, accounting for 7.5% of the variance for value of math, to 24% of the variance for math self-confidence. Gender also seems to have more links with math attitude in Cluster 1 countries (Asia), accounting for 3% (value of math) to 12% (math self-confidence) of the variance.

Within-Clusters Analyses

The within-clusters analyses consist of three separate multiple-group latent variable analyses. The model was tested separately for each of three clusters: Cluster 1 (Asian countries), Cluster 2 (Cold War nations), and Cluster 3 (Northern and Mediterranean European countries). Like the between-clusters analysis, these analyses began by testing the model with all parameters freely estimated for each cluster and no parameters constrained to be equal across clusters. The model was assessed for goodness of fit at this step by examining RMSEA and CFI values. The model performed differently for each of the three clusters in this analysis.

For the cluster of Asian countries, ΔCFI was less than 0.01 for not only the first comparison (all parameters freely estimated versus factor loadings and regression weights constrained), but also for the fully constrained model. This suggests that this cluster is quite homogeneous in terms of the variables in this model. The second cluster, however, does not display this same degree of homogeneity; although the model fits acceptably well when all parameters are freely estimated for the Cold War nations (CFI = 0.910), the same cannot be said when the factor loadings and regression weights are constrained to group equality (CFI = 0.889). Because models with CFI values lower than 0.900 are considered to have unacceptable fit, and because the ΔCFI value is slightly larger than acceptable (ΔCFI = 0.021), the constrained model does not appear fully appropriate. Finally, the model with all parameters freely estimated does not appear to fit well for the Mediterranean and Northern European countries, with a CFI of 0.889.

Once again based on the unstandardized regression coefficients (see Tables 2.8 through 2.13 for the standardized and unstandardized coefficients for these three clusters of countries), the conclusion can be drawn that, of the variables included in the model, math self-confidence had the largest relationship with math achievement for nearly all countries within these clusters; the magnitude of this relationship is considerably larger than that for nearly any other variable in the model, including gender and SES. Also of note is that Value of Math is significantly related to math achievement at the $\alpha = 0.01$ level for only two countries (Czech Republic and South Korea) across all three analyses. Finally, positive math affect again shows a negative relationship with math achievement, but this appears to be the result of multicollinearity in the model; the correlation between the math self-confidence and positive math affect latent constructs is very high (see Tables 2.14 through 2.16).

<u>INSERT TABLE 2.8 APPROXIMATELY HERE</u>

<u>INSERT TABLE 2.9 APPROXIMATELY HERE</u>

<u>INSERT TABLE 2.10 APPROXIMATELY HERE</u>

INSERT TABLE 2.11 APPROXIMATELY HERE

INSERT TABLE 2.12 APPROXIMATELY HERE

INSERT TABLE 2.13 APPROXIMATELY HERE

INSERT TABLE 2.14 APPROXIMATELY HERE

INSERT TABLE 2.15 APPROXIMATELY HERE

INSERT TABLE 2.16 APPROXIMATELY HERE

The cluster of Asian countries yielded some additional interesting information: none of the attitude constructs had large relationships with math achievement in Hong Kong, including Math Self-Confidence, and the statistically significant effects that were observed were only significant at the $\alpha = 0.05$ level. This pattern was repeated to a certain degree in Australia, Sweden, the United States, Armenia, Malta, and Norway; in each of these countries the link between Math Self-Confidence and Math Achievement was much smaller (though still statistically significant at the $\alpha < 0.001$ level) than in other countries in their respective clusters.

The role of gender in the model differed by country within each cluster. For many countries, gender was not statistically significantly related to Math Achievement, and for most of those where gender was statistically significantly related to Math Achievement, the magnitude of the effect was trivial and accounted for less than 1% of the variance. There were some countries wherein this was not the case, however. In Australia, for example, the relationship between gender and Math Achievement was both statistically significant and large in magnitude ($b = -0.841$, $p < 0.001$) in favor of female participants; a similar relationship can be seen in Armenia ($b = -0.643$, $p < 0.001$).

Another important finding for this analysis can be seen in the relationship between gender and the math attitude variables. Although the strength and directionality of the relationship does differ depending on country, the relationship is either not statistically significant or the magnitude of this relationship is very small, accounting for less than 1% of the variance, in all countries except Chinese Taipei, Hong Kong, and Australia. In each of these three countries gender has a statistically significant and relatively strong relationship with all of the math attitude variables, and the relationship favors male participants.

DISCUSSION

The purpose of this study has been to investigate the relationships between gender, socioeconomic status, math attitudes, and math achievement from an international perspective by applying multiple-group latent variable modeling techniques to clusters of countries within the TIMSS 2007 data set. The following sections contain some general conclusions that can tentatively be made based on the results of these analyses.

The Relationships between Math Attitude and Math Achievement

First, the role of self-confidence is clearly shown to be an important one when it comes to math achievement; the self-confidence latent variable was the most influential variable in the model for all of the clusters in the between-clusters analysis, and for almost all of the countries in the within-clusters analyses. Simply stated, students who performed highly relative to their peers on the TIMSS 2007 math assessment were also consistently more likely to endorse feeling confident in their ability to perform well on math-oriented tasks. This was the case between clusters as well as between countries.

The importance of self-confidence for math achievement has been well established (Ethington, 1992; Ethington & Wolfle, 1984; Ganley & Vasilyeva, 2011; Lloyd, Walsh, & Yailagh, 2005; Nosek & Smyth, 2011). The findings of the present study, in addition to what has been found in previous literature for nearly 30 years, suggest that educators should be placing a considerable amount of emphasis on building math self-confidence as math is being taught in the classroom. The question is whether or not this information has been integrated into the primary and secondary education systems through pre-service teacher education programs and through published curricula in math education; this is a topic that should be investigated by future research. Research on math achievement has tended to focus on sex-based differences, and through the course of those efforts several content analyses were conducted and used to make the case that math curricula were potentially unfairly biased against girls (Sadker & Sadker, 1994; Spender & Sarah, 1980). It would be beneficial to the field of math education to make similar inquiries into the content of current curricula to determine whether or not proper attention is being paid to developing self-confidence in learning math for both boys and girls.

A second conclusion that can be drawn from the present analyses has to do with the relationship between the value students place on math and math achievement. In this study, student value of math did not have a statistically significant relationship with math achievement, and even when it was statistically significant the relationship was very small. There are several questions that this raises. For instance, how much of the student response for value of math is a reflection of the way they value math, as opposed to a reflection of the values placed on math by significant adults in their lives (i.e., relatives and/or teachers)?

Finally, the relationship between positive math affect and math achievement was difficult to interpret in this study, largely due to the high correlation between positive math affect and math self-confidence; the correlation was large enough that the two latent variables may well be measuring the same construct in this data set. Although the measurement items used to represent math attitude all loaded onto separate factors during a preliminary factor analysis, their use in the current model had unforeseen effects on the covariance matrix.

The Impact of Gender on Math Attitude and Math Achievement

Gender's link with math achievement has been investigated for decades, with several researchers (e.g., Lieu & Wilson, 2009; Lindberg, Hyde, Petersen, & Linn, 2010) having reached the conclusion that any meaningful gender differences have disappeared, at least in the United States. The findings from this study support such conclusions and extend them to include most of the countries that were investigated; either gender's relationship with math achievement is not statistically significant, or it is of negligible practical importance. Of particular interest is the finding that, in TIMSS 2007 data and for those countries with large and statistically significant effects, the gender gap is in favor of female participants. Whether or not this role-reversal in the math achievement gender gap will result in a similar reversal in the emphasis of research and political discussion remains to be seen, though some preliminary evidence would support such a claim; researchers in a handful of countries have already begun investigating the “boy turn” (Weaver-Hightower, 2003). The “boy turn” is a phrase used to reference a shift in attention in gender-effect research to boys (i.e., “now it’s the boys’ turn to be looked at”, or “now it’s time to turn our attention to the boys”).

Gender’s relationship with math attitude is similar to its relationship with math achievement: the relationship is most commonly small in magnitude and is often not statistically significant. What is particularly interesting is the apparent reversal of direction in the relationship: in countries where gender is a significant and potentially meaningful predictor of math attitude, male participants display more positive attitudes toward math than female participants. However, the number of countries for which this is the case (i.e., the relationship is both statistically significant and has a large magnitude of effect) is limited to three: Chinese Taipei, Hong Kong, and Australia. Why gender has a stronger relationship with math attitude for only these territories is unknown, though the status of both Hong Kong and Australia as former British colonies, and of Chinese Taipei as a dissenting territory within the People’s Republic of China could hold some information as to the reason for the relationship; perhaps the strength of this relationship is in some way a residual effect in these communities as they search for their sense of national identity.

Limitations of the Current Study and Suggestions for Future Research

There are naturally limitations for this study, many of which are tied to the nature of the data. Because this is a secondary data analysis of public-use data, the variables included in the model are necessarily limited to the variables available in the data. The present analysis has suffered from this in at least one major way, and that is the apparent overlap (represented by very high correlations) between the indicator variables used to create the math self-confidence and positive math affect variables. As pointed out earlier, these two constructs are supposed to be different from each other, but they may very well be close to measuring the same thing.

There are currently no mathematical or methodological solutions to the presence of multicollinearity in structural models, and conventional wisdom would suggest that the variables should either be merged into a single variable or one of the two should be dropped from the analysis. This may be the impetus for a follow-up study, wherein only the effects of math self-confidence are analyzed in the model, since the other two attitude variables either suffered from multicollinearity (math affect) or were not consistently significantly influential in the model (value of math). Alternatively, the model could be tested using each of the three attitude variables separately, though this would have the negative side effect of not being able to compare the strength of the resulting regression coefficients with each other because the models would not be accounting for the shared variance among the attitude variables in their relationships with math achievement.

Another limitation for these analyses is the use of categorical indicators for the majority of the independent latent constructs. Although Mplus can accommodate the use of categorical variables through the application of the weighted least squares estimation method, the resulting interpretations are not entirely clear. For example, it was observed that math self-confidence is a significant predictor for math achievement, but the regression coefficients associated with that relationship cannot be interpreted as though they are continuous. Indeed, the only thing that can really be said about the relationship is that higher self-confidence is associated with higher achievement, and the magnitude of the relationship becomes difficult to interpret in a quantitative sense.

Each of the previous limitations shares a common theme: the measurement of the attitude variables has impacted the quality of the analysis and the ability to interpret the results. It would be in the best interests of education researchers, therefore, to develop methods of attitude measurement that do not suffer from these issues. There is some evidence that the Implicit Association Test, or IAT (Greenwald, McGhee, & Schwartz, 1998) can be used to fill this need. The IAT has successfully been used by Nosek and Smyth (2011) as an accurate measure of the cognitive processes associated with math attitudes. Use of the IAT would circumvent many of the biases associated with self-report assessments; in the case of math attitudes, it may help to remove the salience of parental and/or teacher expectation from the participant's responses, yielding a more accurate reporting of positive math affect and value of math in particular.

Yet another limitation with this study is that despite its methodological complexity the TIMSS data set is still cross sectional in nature. Although the TIMSS has four current assessment points available (with a fifth available in January 2013), the use of these data sets in conjunction with each other would still only qualify as longitudinal at the country level because the students assessed differ from assessment cycle to assessment cycle. Therefore, causal inferences cannot be included in the final conclusions in this or any current TIMSS related study. Instead, the findings in this study should be used to inspire future research, ideally longitudinal in nature, to fully investigate the effects of math attitude on math achievement.

The final limitation I would like to acknowledge is that this analysis is contingent upon the model being tested. Regardless of the relatively good fit between the model and the TIMSS 2007 data, in the end it is still simply a model of the relationships among the latent constructs, and it is not necessarily the best model. Other models, those that include other variables from the TIMSS data set, or perhaps exclude variables included here, may perform equally well or possibly more admirably. It is also a reasonable expectation that the presented model is inadequate due to variables of interest that are not available within the TIMSS data, such as the value parents place on math and/or education. Researchers who utilize statistics, and particularly those who employ modeling methods, should be reminded of the fundamental lesson of Magritte's *La Trahison des Images*: the model is simply a conceptual representation, rather than the physical manifestation of any kind of Truth.

TABLES

Table 2.1: Sample size by gender for each country in the TIMSS 2007 8th grade data set.

Country	Female	Male	Total	Country	Female	Male	Total
Algeria	2680	2767	5447	Lebanon	2051	1735	3786
Australia	1843	2226	4069	Lithuania	2016	1975	3991
Bahrain	1974	2256	4230	Malaysia	2362	2104	4466
Armenia	2305	2384	4689	Malta	2374	2296	4670
Bosnia/Herzegovina	2068	2152	4220	Mongolia	2302	2191	4493
Botswana	2193	2015	4208	Morocco	1607	1450	3057
Bulgaria	2045	1974	4019	Oman	2245	2507	4752
Chinese Taipei	1943	2103	4046	Norway	2290	2337	4627
Colombia	2484	2389	4873	Qatar	3639	3545	7184
Cyprus	2196	2203	4399	Romania	2094	2104	4198
Czech Republic	2335	2510	4845	Russia	2326	2146	4472
El Salvador	2137	1926	4063	Saudi Arabia	2226	2017	4243
Georgia	2119	2059	4178	Singapore	2246	2353	4599
Palestine	2373	2005	4378	Slovenia	2022	2021	4043
Ghana	2424	2870	5294	Sweden	2494	2721	5215
Hong Kong	1748	1722	3470	Syria	2339	2311	4650
Hungary	2051	2060	4111	Thailand	2955	2457	5412
Indonesia	2178	2025	4203	Tunisia	2121	1959	4080
Iran	1786	2195	3981	Turkey	2093	2405	4498
Israel	1737	1556	3293	Ukraine	2294	2130	4424
Italy	2114	2294	4408	Egypt	3258	3324	6582
Japan	2142	2170	4312	United States	3721	3656	7377
Jordan	2800	2451	5251	Serbia	1999	2046	4045
South Korea	2016	2224	4240	England	2086	1939	4025
Kuwait	2273	1818	4091	Scotland	2057	2013	4070

Note: N = 225,277; n_{female} = 113,181; n_{male} = 112,096.

Table 2.2: Descriptive statistics for math achievement indicator variables by country.

	Math Knowing		Math Reasoning		Math Applying	
Country	Mean	SD	Mean	SD	Mean	SD
Algeria	370.84	66.51	501.73	79.21	411.85	61.43
Armenia	506.74	76.63	452.21	85.98	492.72	90.86
Australia	487.48	70.70	413.32	87.90	499.94	78.55
Bahrain	394.66	88.19	489.44	101.06	402.65	78.09
Bosnia and Herzegovina	477.98	75.88	455.00	108.87	440.31	81.58
Botswana	376.47	76.08	591.42	108.88	351.10	82.06
Bulgaria	476.97	98.19	415.72	81.49	457.87	103.82
Chinese Taipei	593.72	104.25	460.85	97.84	592.17	101.80
Colombia	364.18	79.31	499.81	77.29	383.94	80.75
Cyprus	468.46	77.95	389.30	109.12	465.07	92.79
Czech Republic	502.41	68.27	381.31	101.34	504.30	75.59
Egypt	392.10	100.85	521.37	68.09	393.28	101.74
El Salvador	335.59	78.85	556.98	99.81	346.66	72.37
Georgia	426.85	98.23	512.64	88.45	401.06	102.32
Hong Kong, SAR	573.64	87.20	462.45	99.52	568.63	92.27
Hungary	518.29	80.31	483.46	80.11	513.36	84.03
Indonesia	396.62	96.33	567.80	92.73	398.33	88.12
Iran, Islamic Republic of	403.31	83.56	440.36	98.36	401.57	88.82
Israel	473.28	89.12	579.02	92.70	455.90	102.74
Italy	476.04	71.45	429.41	91.27	482.99	74.28
Japan	560.00	77.22	485.76	84.65	565.04	82.93
Jordan	431.75	101.72	467.82	70.38	422.24	101.56
Korea, Republic of	596.33	90.93	474.75	88.16	595.25	93.48
Kuwait	347.03	85.73	448.65	79.81	361.05	80.18
Lebanon	464.07	74.59	383.31	90.61	448.02	74.53
Lithuania	507.60	80.39	397.11	94.74	511.39	77.61
Malaysia	476.67	75.58	475.38	74.86	478.06	79.64
Malta	490.44	86.10	448.59	114.04	492.21	92.60
Norway	458.17	51.44	490.12	84.14	477.21	68.12
Oman	372.08	101.22	495.81	79.17	367.53	96.15
Palestinian National Authorit	365.22	107.99	405.06	89.52	370.81	97.76
Romania	470.06	100.19	456.24	87.17	462.05	96.76
Russian Federation	521.15	82.05	425.48	66.30	509.61	81.49
Saudi Arabia	307.73	90.52	440.72	107.09	335.25	81.75
Serbia	500.10	84.37	510.22	73.15	478.46	93.08
Singapore	581.46	81.19	444.74	96.35	593.03	90.29
Slovenia	499.74	68.50	396.50	93.36	502.98	70.10
Sweden	478.24	54.38	504.73	72.14	497.10	71.59
Syria, Arab Republic of	393.14	88.91	473.78	94.68	400.91	82.53
Thailand	436.00	86.83	517.61	83.23	446.40	88.08
Tunisia	420.62	66.44	495.31	81.15	423.36	70.82
Turkey	439.24	108.73	496.03	79.09	424.92	107.10
Ukraine	471.27	89.89	464.64	98.99	463.75	88.02
United States	513.98	68.43	523.90	72.62	502.65	79.42

Table 2.3: Latent constructs and their indicator variables from TIMSS 2007

Latent Construct	Indicator Variable	Description of Indicator
SES	MOMED	Mother's education level
	DADED	Father's education level
	NOBOOKS	Number of books in the home
Math Achievement	MKNOW	Knowledge of mathematics
	MAPPLY	Application of mathematics
	MREASON	Mathematics reasoning
Math Self-Confidence	MDOWELL	Q: I usually do well in mathematics
	MDIFF	Q: Mathematics is more difficult for me than for many of my classmates (reverse scored)
	MNOTSTR	Q: Mathematics is not one of my strengths (reverse scored)
	MLRNQCK	Q: I learn things quickly in mathematics
Math Affect	MENJOY	Q: I enjoy learning mathematics
	MBORING	Q: Mathematics is boring (reverse scored)
	MLIKE	Q: I like mathematics
Value of Math	MHELPDLY	Q: I think learning mathematics will help me in my daily life.
	MLRNOTHR	Q: I need mathematics to learn other school subjects.
	MNEEDUNI	Q: I need mathematics to get into the University of my choice.
	MNEEDJOB	Q: I need to do well in mathematics to get the job I want.

Notes: Parent education variables are ordered categorical ranging from 1 (no education) to 7 (graduate or professional degree. Number of books is ordered categorical ranging from 1 (none or very few) to 5 (more than 200). Math achievement variables are continuous. Math attitude variables (self confidence, affect, value) are all Likert-like scales ranging from 1 (agree a lot) to 4 (disagree a lot).

Table 2.4: Cluster membership of countries based on Duerr & Harlow (under review).

Cluster	Countries in Cluster	No. of Countries
1	Japan, Chinese Taipei, Republic of Korea, Hong Kong, Singapore	5
2	Czech Republic, Slovenia, Hungary, England, Australia, Russian Federation, Lithuania, United States, Sweden	9
3	Norway, Italy, Scotland, Malta, Serbia, Armenia, Ukraine	7
4	Bulgaria, Bosnia and Herzegovinia, Romania, Cyprus, Israel, Malaysia, Lebanon	7
5	Thailand, Jordan, Tunisia, Turkey, Bahrain, Islamic Republic of Iran, Arab Republic of Syria	7
6	Palestinian National Authority, Botswana, Kuwait, El Salvador, Saudi Arabia	5
7	Oman, Columbia, Egypt, Algeria, Georgia, Indonesia	6
8	Qatar, Ghana	2

Table 2.5: Unstandardized factor loadings and regression weights for between-clusters analysis.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Construct	Indicator				λ			
F1	X1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X2	0.803	0.873	0.811	0.849	1.022	1.008	0.645
	X3	0.954	0.923	0.845	0.821	1.095	1.146	0.851
	X4	0.938	0.951	0.990	1.063	1.171	1.464	1.082
F2	X5	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X6	0.821	0.821	0.789	0.699	0.734	0.669	0.884
	X7	0.814	0.821	0.882	0.778	0.893	0.779	0.736
F3	X8	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X9	0.959	0.930	0.949	0.989	0.986	0.888	1.047
	X10	1.095	1.136	1.202	1.375	1.363	1.148	1.397
	X11	1.162	1.109	1.238	1.312	1.242	1.048	1.418
F4	X12	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X13	1.034	1.042	1.080	1.087	0.964	0.915	0.957
	X14	1.013	0.974	0.986	0.996	0.861	0.945	0.856
F5	X15	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X16	3.481	1.567	1.769	1.537	1.504	1.721	2.274
	X17	3.433	1.632	1.783	1.546	1.546	1.841	2.179
To	From				*b*			
Ach	SCM	***0.838***	***0.884***	***0.599***	***0.774***	***0.816***	***0.745***	***1.067***
	PATM	***-0.262***	***-0.422***	***-0.095***	***-0.279***	***-0.445***	***-0.348***	***-0.806***
	VALM	***0.074***	***0.092***	***-0.080***	***-0.062***	***0.273***	***0.243***	***0.388***
	Gender	***-0.309***	***-0.137***	***-0.076***	***-0.059***	***-0.041***	***-0.124***	***-0.117***
SCM	SES	***0.333***	***0.454***	***0.303***	***0.313***	***0.348***	***0.281***	***0.116***
	Gender	***0.632***	***0.142***	***0.115***	***0.043***	0.011	***0.047***	***0.101***
PATM	SES	***0.198***	***0.192***	***0.223***	0.001	***0.121***	**0.029**	***-0.120***
	Gender	***0.559***	***-0.052***	-0.023	***-0.077***	-0.016	***0.089***	***0.058***
VALM	SES	***0.170***	***0.128***	***0.132***	0.015	***0.148***	***0.088***	-0.011
	Gender	***0.272***	***0.040***	***0.105***	*-0.032*	***-0.086***	***-0.063***	***-0.056***

Note: ***Itallic and Bold*** = *p <0 .001; Itallic* = *p* < 0.01; **Bold** = *p* <0 .05; All other values are NS (*p* > 0.05).

Table 2.6: Standardized factor loadings and regression weights for between-clusters analysis.

		Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Construct	Indicator				λ			
F1	X1	0.863	0.865	0.813	0.785	0.780	0.687	0.697
	X2	0.820	0.781	0.709	0.653	0.549	0.377	0.437
	X3	0.890	0.819	0.746	0.676	0.668	0.519	0.528
	X4	0.878	0.823	0.834	0.807	0.748	0.648	0.678
F2	X5	0.948	0.885	0.860	0.883	0.886	0.832	0.849
	X6	0.853	0.754	0.798	0.797	0.662	0.606	0.630
	X7	0.980	0.971	0.934	0.958	0.887	0.855	0.843
F3	X8	0.771	0.737	0.757	0.706	0.662	0.661	0.687
	X9	0.795	0.715	0.613	0.618	0.612	0.561	0.587
	X10	0.905	0.814	0.765	0.802	0.754	0.714	0.688
	X11	0.897	0.837	0.776	0.813	0.765	0.720	0.725
F4	X12	0.971	0.953	0.938	0.936	0.958	0.890	0.931
	X13	0.966	0.947	0.953	0.949	0.952	0.927	0.919
	X14	0.873	0.833	0.815	0.824	0.837	0.787	0.770
F5	X15	0.858	0.870	0.891	0.894	0.871	0.757	0.820
	X16	0.890	0.845	0.869	0.844	0.777	0.648	0.754
	X17	0.519	0.578	0.477	0.552	0.573	0.602	0.551
To	From				β			
Ach	SCM	0.918	1.073	0.671	0.862	0.762	0.678	0.917
	PATM	-0.312	-0.522	-0.112	-0.374	-0.472	-0.384	-0.842
	VALM	0.070	0.094	-0.084	-0.062	0.216	0.213	0.328
	Gender	-0.187	-0.096	-0.052	-0.501	-0.024	-0.082	-0.072
SCM	SES	0.316	0.456	0.332	0.345	0.388	0.309	0.137
	Gender	0.349	0.082	0.071	0.327	0.007	0.034	0.073
PATM	SES	0.173	0.189	0.231	0.001	0.118	0.026	-0.116
	Gender	0.284	-0.029	-0.013	-0.490	-0.009	0.053	0.034
VALM	SES	0.188	0.151	0.155	0.019	0.194	0.101	-0.013
	Gender	0.175	0.027	0.069	-0.275	-0.065	-0.048	-0.041

Table 2.7: Correlation/covariance matrix for latent constructs in the between clusters analysis.

	Math Self-Confidence	Positive Math Affect	Value of Math	Math Achievement	Socioeconomic Status
Math Self-Confidence	**0.819**	*0.782*	*0.374*	*0.460*	*0.245*
Positive Math Affect	0.878	**0.969**	*0.509*	*0.397*	*0.146*
Value of Math	0.532	0.665	**0.606**	*0.205*	*0.125*
Math Achievement	0.616	0.488	0.318	**0.683**	*0.177*
Socioeconomic Status	0.316	0.173	0.188	0.249	**0.735**

Note: Variances are **bold**; Covariances are *itallicized* in the upper diagonal; Correlations are in the lower diagonal

Table 2.8: Unstandardized factor loadings and regression weights for Cluster 1.

		Chinese Taipei	Hong Kong	Japan	Korea	Singapore
Construct	Indicator			λ		
F1	X1	1.000	1.000	1.000	1.000	1.000
	X2	0.768	0.962	0.929	1.002	0.788
	X3	0.934	0.975	1.131	0.960	0.885
	X4	0.935	1.009	0.993	0.912	0.950
F2	X5	1.000	1.000	1.000	1.000	1.000
	X6	0.852	0.880	0.863	0.913	0.849
	X7	1.016	1.016	1.054	1.041	1.019
F3	X8	1.000	1.000	1.000	1.000	1.000
	X9	0.896	0.967	0.977	1.344	0.907
	X10	1.138	1.070	1.206	1.526	1.051
	X11	1.056	1.140	1.121	1.449	1.062
F4	X12	1.000	1.000	1.000	1.000	1.000
	X13	0.991	1.080	1.018	1.026	1.037
	X14	0.956	1.054	1.035	0.929	1.057
F5	X15	1.000	1.000	1.000	1.000	1.000
	X16	1.010	0.949	1.089	1.100	1.142
	X17	0.655	0.586	0.479	0.815	0.596
To	From			*b*		
Ach	SCM	***1.242***	**0.256**	***1.370***	***0.966***	***0.848***
	PATM	**-0.622**	0.151	***-0.747***	***-0.345***	**-0.466**
	VALM	***0.140***	**0.086**	**0.126**	***0.226***	0.040
	Gender	-0.443	**-0.081**	***-0.209***	***-0.169***	***-0.268***
SCM	SES	***0.424***	***0.198***	***0.473***	***0.613***	***0.416***
	Gender	***0.821***	***0.319***	***0.203***	***0.170***	***0.180***
PATM	SES	***0.231***	***0.194***	***0.290***	***0.424***	**0.127**
	Gender	***0.681***	***0.181***	***0.109***	**0.060**	0.055
VALM	SES	***0.226***	***0.213***	***0.233***	***0.246***	***0.129***
	Gender	***0.523***	*0.112*	***0.080***	***0.135***	0.045

Note: ***Itallic and Bold*** = *p <0 .001; Itallic* = *p* < 0.01; **Bold** = *p* <0 .05; All other values are NS (*p* > 0.05).

Table 2.9: Standardized factor loadings and regression weights for Cluster 1.

		Chinese Taipei	Hong Kong	Japan	Korea	Singapore
Construct	Indicator			λ		
F1	X1	0.905	0.879	0.820	0.931	0.920
	X2	0.717	0.805	0.697	0.855	0.693
	X3	0.563	0.858	0.926	0.894	0.815
	X4	0.855	0.887	0.814	0.849	0.874
F2	X5	0.962	0.951	0.927	0.938	0.946
	X6	0.832	0.838	0.800	0.857	0.803
	X7	0.976	0.966	0.977	0.977	0.964
F3	X8	0.777	0.842	0.721	0.605	0.786
	X9	0.701	0.815	0.704	0.812	0.713
	X10	0.876	0.901	0.869	0.921	0.827
	X11	0.818	0.960	0.808	0.875	0.835
F4	X12	0.982	0.967	0.977	0.971	0.991
	X13	0.977	0.974	0.951	0.976	0.946
	X14	0.892	0.862	0.873	0.871	0.858
F5	X15	0.840	0.807	0.819	0.846	0.780
	X16	0.849	0.765	0.892	0.931	0.891
	X17	0.550	0.472	0.392	0.690	0.465
To	From			β		
Ach	SCM	1.254	0.301	1.486	1.025	1.099
	PATM	-0.651	0.195	-0.914	-0.367	-0.616
	VALM	0.116	0.096	0.120	0.155	0.043
	Gender	-0.231	-0.054	-0.138	-0.095	-0.189
SCM	SES	0.363	0.179	0.470	0.555	0.351
	Gender	0.416	0.179	0.123	0.091	0.098
PATM	SES	0.191	0.164	0.256	0.382	0.104
	Gender	0.332	0.095	0.059	0.032	0.029
VALM	SES	0.236	0.204	0.264	0.343	0.128
	Gender	0.325	0.067	0.056	0.111	0.028

Table 2.10: Unstandardized factor loadings and regression weights for Cluster 2.

		Australia	Czech Republic	Hungary	Lithuania	Russia	Slovenia	Sweden	United States
Construct	Indicator				λ				
F1	X1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X2	0.809	0.948	1.012	0.902	1.054	0.902	0.906	0.841
	X3	0.888	1.048	0.927	1.217	1.067	0.849	0.921	0.889
	X4	0.905	0.996	0.961	0.962	1.001	1.022	0.988	0.903
F2	X5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X6	0.790	0.914	0.809	0.726	0.979	0.977	0.970	0.795
	X7	1.031	1.111	1.048	1.029	1.086	1.108	1.032	1.052
F3	X8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X9	0.837	0.884	0.931	1.192	1.037	0.753	0.871	0.884
	X10	1.056	1.189	0.859	1.256	1.080	0.987	1.152	1.221
	X11	1.117	1.096	1.079	1.291	1.180	0.862	1.108	1.148
F4	X12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X13	0.931	1.027	1.090	0.968	1.118	0.997	0.889	1.229
	X14	0.780	0.918	0.960	1.041	1.101	0.954	0.831	0.959
F5	X15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	X16	1.033	1.162	0.946	0.958	0.949	1.006	0.903	1.013
	X17	0.608	0.694	0.816	0.650	0.559	0.535	0.677	0.477
To	From				*b*				
Ach	SCM	***0.458***	***0.879***	***2.609***	***0.979***	***0.759***	***0.828***	***0.337***	***0.467***
	PATM	**-0.108**	***-0.397***	***-2.382***	***-0.300***	***-0.361***	***-0.383***	-0.004	***-0.225***
	VALM	0.036	*-0.110*	**0.739**	**-0.155**	-0.037	-0.005	-0.045	-0.006
	Gender	***-0.841***	***-0.157***	*-0.482*	***-0.138***	*-0.076*	-0.034	-0.048	-0.003
SCM	SES	***0.504***	***0.441***	***0.332***	***0.383***	***0.358***	***0.363***	***0.347***	***0.279***
	Gender	***0.707***	***0.179***	***0.128***	**0.063**	***-0.022***	*0.104*	***0.262***	***0.195***
PATM	SES	***0.388***	***0.149***	***0.142***	***0.135***	***0.096***	***0.132***	***0.318***	***0.092***
	Gender	***0.624***	*-0.096*	-0.062	-0.057	***-0.158***	0.004	**0.086**	0.031
VALM	SES	***0.316***	0.016	-0.010	*0.073*	0.024	*0.065*	***0.255***	***0.119***
	Gender	***0.744***	***0.118***	0.053	*-0.090*	*0.081*	**0.081**	*0.085*	-0.046

Note: ***Itallic and Bold*** = p <0 .001; *Itallic* = p < 0.01; **Bold** = p <0 .05; All other values are NS (p > 0.05).

Table 2.11: Standardized factor loadings and regression weights for Cluster 2.

Construct	Indicator	Australia	Czech Republic	Hungary	Lithuania	Russia	Slovenia	Sweden	United States
					λ				
F1	X1	0.919	0.886	0.887	0.765	0.825	0.851	0.889	0.906
	X2	0.757	0.807	0.777	0.670	0.811	0.778	0.834	0.733
	X3	0.825	0.928	0.823	0.930	0.881	0.723	0.820	0.807
	X4	0.840	0.882	0.852	0.736	0.826	0.870	0.879	0.819
F2	X5	0.935	0.867	0.903	0.901	0.849	0.859	0.933	0.909
	X6	0.751	0.792	0.731	0.654	0.831	0.839	0.905	0.722
	X7	0.961	0.963	0.946	0.927	0.921	0.952	0.962	0.957
F3	X8	0.863	0.668	0.699	0.627	0.660	0.847	0.822	0.686
	X9	0.736	0.591	0.651	0.747	0.685	0.638	0.716	0.606
	X10	0.905	0.793	0.600	0.788	0.713	0.835	0.946	0.837
	X11	0.950	0.731	0.754	0.810	0.779	0.730	0.910	0.788
F4	X12	1.002	0.968	0.952	0.936	0.937	0.942	1.017	0.962
	X13	0.924	0.927	0.956	0.945	0.949	0.951	0.845	0.980
	X14	0.760	0.797	0.852	0.871	0.831	0.817	0.620	0.801
F5	X15	0.763	0.783	0.917	0.902	0.838	0.881	0.813	0.802
	X16	0.788	0.910	0.867	0.864	0.795	0.886	0.733	0.813
	X17	0.463	0.543	0.748	0.586	0.468	0.471	0.550	0.383
To	From				β				
Ach	SCM	0.694	1.197	3.082	1.058	0.960	1.089	0.626	0.807
	PATM	-0.163	-0.527	-2.861	-0.381	-0.471	-0.508	-0.007	-0.388
	VALM	0.052	-0.112	0.688	-0.138	-0.037	-0.007	-0.078	-0.007
	Gender	-0.655	-0.120	-0.320	-0.097	-0.058	-0.026	-0.050	-0.003
SCM	SES	0.394	0.388	0.342	0.451	0.363	0.375	0.314	0.245
	Gender	0.363	0.100	0.072	0.041	-0.013	0.061	0.146	0.107
PATM	SES	0.302	0.135	0.144	0.135	0.094	0.135	0.277	0.081
	Gender	0.318	-0.056	-0.035	-0.032	-0.093	0.002	0.046	0.017
VALM	SES	0.262	0.019	-0.013	0.104	0.030	0.068	0.252	0.139
	Gender	0.404	0.088	0.038	-0.072	0.061	0.048	0.052	-0.033

Table 2.12: Unstandardized factor loadings and regression weights for Cluster 3.

		Armenia	Italy	Malta	Norway	Ukraine	Serbia
Construct	Indicator			λ			
F1	X1	1.000	1.000	1.000	1.000	1.000	1.000
	X2	0.687	0.772	0.717	0.826	0.884	1.340
	X3	0.851	0.984	0.777	0.857	0.973	0.908
	X4	1.110	1.020	0.926	0.901	1.051	1.063
F2	X5	1.000	1.000	1.000	1.000	1.000	1.000
	X6	0.804	0.937	0.924	0.956	0.952	0.971
	X7	1.063	1.029	1.069	1.065	1.101	1.100
F3	X8	1.000	1.000	1.000	1.000	1.000	1.000
	X9	0.848	0.939	1.022	0.904	1.028	0.909
	X10	1.049	1.100	1.154	1.143	1.215	1.110
	X11	1.111	1.086	1.184	1.132	1.152	1.107
F4	X12	1.000	1.000	1.000	1.000	1.000	1.000
	X13	1.226	0.970	1.039	0.917	1.057	1.205
	X14	1.017	0.825	0.948	0.780	1.045	1.052
F5	X15	1.000	1.000	1.000	1.000	1.000	1.000
	X16	1.032	0.939	0.735	0.881	0.998	0.924
	X17	0.642	0.302	0.062	0.520	0.704	0.661
To	From			*b*			
Ach	SCM	***0.363***	***0.808***	***0.510***	***0.373***	***1.220***	***1.015***
	PATM	-0.057	*-0.415*	*-0.131*	***-0.157***	***-0.749***	***-0.457***
	VALM	-0.051	-0.057	-0.046	0.025	0.217	-0.057
	Gender	***-0.643***	**-0.081**	*-0.088*	-0.045	***-0.210***	-0.056
SCM	SES	***0.235***	***0.208***	*0.072*	***0.382***	***0.487***	***0.359***
	Gender	0.457	***0.179***	***0.135***	***0.155***	*0.091*	0.035
PATM	SES	**0.145**	**0.072**	-0.002	***0.203***	***0.187***	-0.024
	Gender	0.376	**0.089**	0.026	0.002	*-0.100*	***-0.101***
VALM	SES	***0.111***	**0.068**	0.020	***0.211***	*0.093*	**-0.059**
	Gender	0.490	0.076	***0.122***	**0.095**	***0.125***	*0.080*

Note: ***Itallic and Bold*** = p <0 .001; *Itallic* = p < 0.01; **Bold** = p <0 .05; All other values are NS (p > 0.05).

Table 2.13: Standardized factor loadings and regression weights for Cluster 3.

		Armenia	Italy	Malta	Norway	Ukraine	Serbia
Construct	Indicator			λ			
F1	X1	0.700	0.888	0.891	0.906	0.840	0.829
	X2	0.488	0.754	0.703	0.769	0.770	0.794
	X3	0.601	0.873	0.693	0.776	0.817	0.753
	X4	0.770	0.906	0.825	0.816	0.883	0.881
F2	X5	0.818	0.942	0.905	0.924	0.856	0.875
	X6	0.668	0.883	0.836	0.883	0.814	0.847
	X7	0.864	0.969	0.967	0.985	0.940	0.960
F3	X8	0.684	0.743	0.663	0.829	0.665	0.733
	X9	0.576	0.697	0.668	0.750	0.680	0.663
	X10	0.661	0.818	0.750	0.947	0.804	0.808
	X11	0.712	0.807	0.773	0.938	0.763	0.799
F4	X12	0.927	0.986	0.949	1.015	0.950	0.945
	X13	0.959	0.947	0.968	0.838	0.963	0.963
	X14	0.745	0.751	0.845	0.619	0.872	0.833
F5	X15	0.865	0.805	0.840	0.875	0.872	0.921
	X16	0.893	0.755	0.617	0.770	0.871	0.851
	X17	0.556	0.243	0.052	0.455	0.614	0.609
To	From			β			
Ach	SCM	0.352	1.108	0.645	0.792	1.237	1.142
	PATM	-0.065	-0.603	-0.168	-0.340	-0.774	-0.544
	VALM	-0.040	-0.065	-0.044	0.048	0.171	-0.050
	Gender	-0.428	-0.063	-0.062	-0.053	-0.127	-0.037
SCM	SES	0.282	0.188	0.068	0.366	0.505	0.398
	Gender	0.310	0.100	0.076	0.085	0.054	0.021
PATM	SES	0.146	0.061	-0.002	0.192	0.190	-0.027
	Gender	0.223	0.047	0.015	0.001	-0.059	-0.058
VALM	SES	0.139	0.073	0.027	0.223	0.124	-0.075
	Gender	0.278	0.051	0.092	0.057	0.094	0.054

Table 2.14: Correlation/covariance matrix for latent constructs for between clusters analysis for Cluster 1.

	Math Self-Confidence	Positive Math Affect	Value of Math	Math Achievement	Socioeconomic Status
Math Self-Confidence	**0.963**	*0.917*	*0.490*	*0.611*	*0.299*
Positive Math Affect	0.917	**1.039**	*0.580*	*0.510*	*0.163*
Value of Math	0.621	0.709	**0.646**	*0.280*	*0.159*
Math Achievement	0.642	0.517	0.359	**0.938**	*0.293*
Socioeconomic Status	0.363	0.190	0.236	0.360	**0.706**

Note: Variances are **bold**; Covariances are *itallicized* in the upper diagonal; Correlations are in the lower diagonal

Table 2.15: Correlation/covariance matrix for latent constructs for between clusters analysis for Cluster 2.

	Math Self-Confidence	Positive Math Affect	Value of Math	Math Achievement	Socioeconomic Status
Math Self-Confidence	**0.950**	*0.695*	*0.474*	*0.229*	*0.293*
Positive Math Affect	0.728	**0.959**	*0.554*	*0.104*	*0.226*
Value of Math	0.528	0.615	**0.847**	*0.031*	*0.184*
Math Achievement	0.366	0.166	0.053	**0.412**	*0.120*
Socioeconomic Status	0.395	0.302	0.262	0.246	**0.583**

Note: Variances are **bold**; Covariances are *itallicized* in the upper diagonal; Correlations are in the lower diagonal

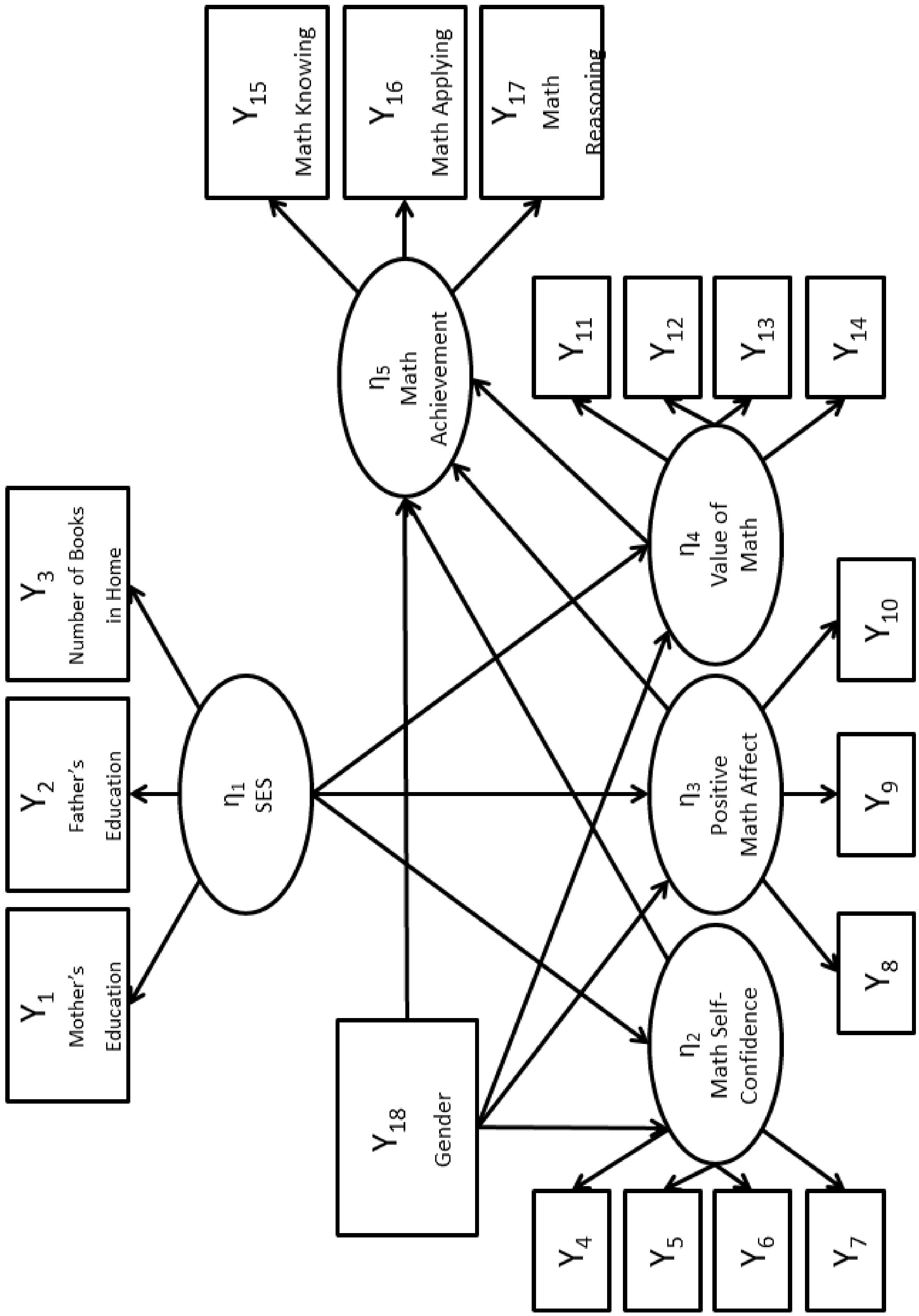
Table 2.16: Correlation/covariance matrix for latent constructs for between clusters analysis for Cluster 3.

	Math Self-Confidence	Positive Math Affect	Value of Math	Math Achievement	Socioeconomic Status
Math Self-Confidence	**0.525**	*0.482*	*0.293*	*0.076*	*0.176*
Positive Math Affect	0.779	**0.731**	*0.411*	*0.051*	*0.108*
Value of Math	0.550	0.654	**0.539**	*-0.019*	*0.083*
Math Achievement	0.139	0.079	-0.034	**0.563**	*0.053*
Socioeconomic Status	0.280	0.146	0.130	0.082	**0.749**

Note: Variances are **bold**; Covariances are *itallicized* in the upper diagonal; Correlations are in the lower diagonal

FIGURES

Figure 2.1: Latent variable model of relationships between math attitude, socioeconomic status, gender, and math achievement.



APPENDIX 1: REVIEW OF TIMSS ANALYSIS CONSIDERATIONS

TIMSS is a complex data set, and some may not be familiar with the many complexities associated with its proper analysis. Balanced Incomplete Block Sampling and Jackknife Replication techniques are not exactly ubiquitous methodologies, and even renowned researchers may be unfamiliar with the details associated with the methods required to analyze large-scale data sets. The purpose of this appendix is to provide a brief yet thorough overview of the complexities associated with the TIMSS data set.

The TIMSS is an international collaboration which attempts to measure mathematics and science achievement at the fourth and eighth grades. TIMSS is a recurring study conducted every four years. At the international level, the study is organized by the International Association for the Evaluation of Educational Achievement (IEA), which also organizes a similar international study of reading literacy, the Progress in International Reading Literacy Study (PIRLS). Although it is the principle organizational entity responsible for TIMSS, the IEA relies heavily on the coordination and support of organizations within each participating country; it is the responsibility of these supporting organizations to select their country's participants, administer the assessments, and consolidate the data. In the United States, this task is handled by the TIMSS & PIRLS International Study Center at the Lynch School of Education, Boston College. Funding for the U.S. participation in TIMSS comes from the budget for the National Center for Education Statistics (NCES).

TIMSS Data Collection Procedures

As an international assessment of mathematics and science achievement with dozens of participating countries and hundreds-of-thousands of student-participants, TIMSS data collection is a very complicated process. The following is a general overview of those procedures; significantly more detailed information on the TIMSS data collection process can be found in Williams et al. (2009).

The populations of interest for the TIMSS are all fourth and eighth grade students in the participating countries. The sampling techniques TIMSS employs aim to produce country-level samples that are representative of the country's population. Because my investigations all focus on the eighth grade data, I will discuss the sampling procedures only for the eighth grade and make the statement that the sampling procedures are generally the same at the fourth grade level. Additionally, for the sake of simplicity, I will discuss the sampling procedure as it relates to the United States, with the acknowledgement that this is the IEA mandated sampling procedure used by all countries.

Sample Selection

TIMSS data collection begins with the two-stage sampling process. First, a sample of schools within the country is selected, and then a sample of classrooms within the school is selected. To create the school-level sample, schools which included an eighth grade were identified and included as possible participants. For each school, a variable was calculated to represent the school's size in relation to the target population; this variable is termed the Measure of Size (MOS). An ordered sampling frame was created based on the school's MOS and four categorical characteristics: school control (public or private); region (Northeast, Southeast, Central, or West); school location relative to population of the surrounding area (large city, mid-size city, urban fringe of a large city, urban fringe of a mid-size city, large town, small town, rural); and race/ethnicity status (above/below 15 percent Black, non-Hispanic, Hispanic, American Indian, or Alaska Native). The ordered sampling frame was sorted by each categorical characteristic, creating an implicit stratification scheme with 128 implicit strata. The sampling frame was then sorted by MOS in descending order.

To select participating schools, schools were randomly selected from each stratum. As schools were identified, a first- and second-alternative was selected as well, based on the primary selection's position in the sampling frame; the school below the selected school was identified as the first-alternative, and the school above the selected school was identified as the second-alternative. Schools were selected within the stratum using proportional probability sampling (PPS) until the target MOS for the stratum was met or exceeded. Alternative schools were invited to participate only if the primary selection declined.

Once schools were identified for participation and agreed to participate, a second sampling frame was created. This sampling frame included all eligible classrooms within the school. All classrooms within the school had an equal probability of being selected for participation, and two classrooms were selected within each school. In return for participating in the study, schools were gifted with an all-in-one printer, and students were gifted with a clock-compass carabiner.

Assessment Administration Procedures

The TIMSS administration procedures are also complex. In order to reduce the time required to complete the TIMSS assessment yet still adequately cover a breadth of subject-matter, balanced incomplete block spiraling (BIB) of assessment items was employed. The BIB assessment design is manifested in the assessment booklets used. Each student used one of 18 different assessment booklets. Although each booklet contained both mathematics and science questions, the exact questions in each booklet differed. In addition to new items for each assessment period, TIMSS collects responses on items from the previous assessment period, creating a set of bridging items. Each student was assigned one assessment booklet, the completion of which required approximately 90 minutes.

Achievement Scores

TIMSS assessments are comprised of a mixture of open-ended responses and multiple-choice options. Items are scored by a team of trained scorers; TIMSS 2007's scoring team consisted of 109 scorers, 12 supervisors, and 2 subject-specific directors. Student responses to individual items were used to construct Item Response Theory (IRT) based score scales, which summarize the achievement results for participants. Because the scale scores are placed on the same scale, the performance of samples of students can be summarized on a single scale (or series of scales), even though different participants responded to different items.

This assessment method has a built-in missing data component. This missing data is accommodated by the use of plausible values. As stated in the TIMSS 2007 Technical Manual (Williams et al., 2009), "each plausible value represents a random selection from the distribution of scale scores of students with similar backgrounds who answered the assessment items in a similar way" (p.79). Thus, a given student's background characteristics and actual response patterns are used to estimate responses for the items that were not asked due to the incomplete block design of assessment; essentially, based on the estimated distribution of the student's achievement, the student's scale scores represent the values he or she could have had, had the student been asked all of the questions.

There are notable benefits associated with using BIB assessment and plausible values. Firstly, the burden of assessment is reduced; students are asked only a sample of questions, rather than being subjected to a battery consisting of all possible questions. This translates into less time on the part of the participant as well as a cost reduction for the overall assessment. Secondly, the use of plausible values to estimate scale scores provides more accurate population-level estimates of average performance and variability than procedures which utilize a single score for each student (Beaton & González, 1995; Olson, Martin, & Mullis, 2008). There is, however, a tradeoff to the use of plausible values: because they are drawn from the estimated distribution of the student's achievement, plausible values are not an accurate or valid measure of a specific individual student's achievement.

TIMSS Analysis Considerations

The complexity of the TIMSS sampling design necessitates added complexity for any analysis using the data set. This complexity comes in two forms: adequately adjusting the sample such that it is representative of the population, and accurately estimating the error introduced by the sampling design.

As was previously discussed, TIMSS assessments utilize a two-stage sampling procedure, with the result being a stratified sampling frame based on several categorical variables and the participating school's enrollment. The construction of the sampling frame allows TIMSS researchers to create weighting variables which, when properly employed, forces the data to be representative of the population for the country. The TIMSS weighting variables are calculated for each individual participant based on their sampling frame characteristics. As detailed in Foy and Olson (2009, pp. 102-105), the probability of any given student within a country being selected for the sample is known because students were selected using the probability sampling method previously described. Therefore, sampling weights can be constructed for each student by taking the inverse of the probability of selection. The use of sampling weights accommodates the complex sampling design by accounting for stratification and any disproportional subgroup sampling; the TIMSS sampling weights also include adjustments for non-response.

The TIMSS data set includes several sample weight variables, each of which has a different purpose. Of primary interest are the TOTWGT (total weight) and SENWGT (senate weight) variables. TOTWGT is the variable name assigned for the sample weight previously described; when used, TOTWGT will ensure that the subgroups within the stratified sample are proportionally represented in population estimates. TOTWGT should be used whenever student-level population estimates within a country are desired.

When properly applied, the TOTWGT variable will inflate the sample size for a given country to approximately the size of the grade-appropriate population for that country (i.e., TOTWGT would inflate the 7,377 U.S. eighth grade participants to approximately the size of the entire eighth grade population). This creates a problem when comparisons between countries are desired; countries with larger populations have more students than countries with smaller populations. This difficulty is accommodated by using the SENWGT weighting variable. SENWGT is a transformation of TOTWGT that forces each country to have a weighted sample size of 500. For analyses where comparisons are being made between countries, the SENWGT variable should be used rather than TOTWGT to allow for an equitable assessment (Foy & Olson, 2009).

The other analysis consideration when using TIMSS data is the proper estimation of the error introduced by the sampling design. There are two forms of error to consider, sampling error caused by the stratified sampling procedure, and imputation error caused by the use of plausible values. To accommodate the sampling error, the jackknife repeated replication technique (JRR) is employed. In JRR, pairs of schools are systematically assigned to sampling zones, creating pseudo-replicates of the original sample; for TIMSS, 75 pseudo-replicates were created. The statistic of interest is calculated once for the overall sample and again for each pseudo-replicate. The variation between the estimate of the original sample and the estimate for the jackknife replicate is the jackknife estimate of the sampling error for the statistic (Foy et al., 2007). The 75 jackknife estimates for sampling error were then used to create 75 replicate weights. Adequately accommodating the error introduced by the stratified sampling procedure, therefore, involves estimating a parameter 76 times, once for the original sample and once for each replicate weight.

This brings up the point of accommodating the plausible values in the analysis. Because plausible values are imputations rather than actual observed scores, there is error associated with the imputation. As stated by Williams et al. (2009), averaging the plausible values and using the resulting mean to calculate a parameter estimate would underestimate the standard error associated with the subsequent analysis. Therefore, the imputation error is accommodated by calculating a given statistic once for each plausible value and then averaging these results over five analyses.

The accommodation of both sampling error and imputation error would ideally result in the calculation of a given parameter estimate 76 times for each of the five plausible values (once for the overall sample and once for each jackknife replicate weight), yielding 380 analyses to be averaged for an accurate parameter estimate. However, Foy et al. (2007) state that a shortcut is available: accommodate the sampling error by estimating the parameter once for each of the 75 replicate weights using only the first plausible value, and then accommodate the imputation error by estimating the parameter once for each plausible value, computing the parameter estimate a total of 80 times.

Analysis Software

As can be seen in the previous sections, analyses of the TIMSS data have some inherently complex considerations associated with them which are not common to most data analysis endeavors. Fortunately, data analysts have some support in overcoming these complexities in the form of the analysis software available. First and foremost, the IEA has created and made available a database analysis package called the International Database Analyzer (IDB Analyzer). The IDB Analyzer is a standalone application that generates SPSS syntax, which can then be used to analyze data from the IEA assessments, including TIMSS. The SPSS syntax generated by the IDB analyzer accounts for the complex sampling design associated with these studies by properly implementing the use of the weighting variables and plausible values.

Although the IDB Analyzer makes some analyses simpler to perform, the range of options available through the IDB Analyzer is limited. An analyst can calculate percentages and means, correlations, and percentiles, and can perform regression analyses, but that is all. Even within these options, there are limitations to what the IDB Analyzer will do. For example, only one variable with plausible values can be used for any given analysis, so analyses comparing more than one plausible values-based variable, such as math achievement with science achievement, are not possible. Additionally, multivariate methods more advanced than multiple linear regression are also not possible through the IDB Analyzer.

Fortunately, the developers of certain advanced analysis software have begun to incorporate methods for analyzing complex data; both LISREL and Mplus have accommodated the use of sampling weights and imputation through plausible values for SEM techniques, and HLM and Mplus can both accommodate these complexities when conducting hierarchical linear modeling (HLM).

APPENDIX 2: MODEL BUILDING PROCEDURE

In this appendix I will briefly discuss the model building procedure used to create the model tested in this analysis. The decisions were based on not only model fit, but also on the goals of the study.

As was stated in the body of this paper, the model used in this study is based on a model that was previously tested with only the U.S. data. However, that model used a single math attitude latent construct, which was comprised of several of the math attitude indicator variables in this study. The findings of that study indicated that the math attitude indicator variables did not contribute equally to the relationship on math achievement, and so the model for the present study was created to investigate those inequalities.

Because the primary goals of the study were to investigate differences among the math affect variables' impact on math achievement, three separate math affect latent variables were used for the model. This is also why a second-order math affect variable was not included in the model; a second-order math affect variable would have removed the paths between the individual math affect variables and math attitude, thereby defeating the goal of the analyses.

The model represented in Figure 1 was originally conceptualized with a regression path between SES and math achievement. However, it was determined that the final model displayed superior fit ($\Delta CFI = 0.039$, $\Delta RMSEA = 0.012$). In this way, SES is included in the model as a covariate; we are not especially interested in the relationships between SES and the other variables in the model, but we want to account for the role that SES plays in the model.

REFERENCES

Beaton, A. E., & González, E. (1995). *The NAEP primer*. Chestnut Hill, MA: Boston College.

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

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In Bollen, K., & Long, J., Editors: *Testing structural equation models.* Newbury Park, CA: Sage.

Chen, F., Curran, P.J., Bollen, K.A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociology Methods Research, 36*(4), 462-494.

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

Duerr, S. R., & Harlow, L. L. (Under Review). Identifying and validating clusters of countries in the TIMSS 2007 data set.

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin, 136*(1), 103-127.

Ethington, C. A. (1992). Gender differences in a psychological model of mathematics achievement. *Journal for Research in Mathematics Education, 23*(2), 166-181.

Ethington, C. A., & Wolfle, L. M. (1984). Sex differences in a causal model of mathematics achievement. *Journal for Research in Mathematics Education, 15*(5), 361-377.

Foy, P., Galia, J., & Li, I. (2007). Scaling the data from the TIMSS 2007 mathematics and science assessments. In J. F. Olson, Martin, M. O., & Mullis, I.V.S. (Ed.), *TIMSS 2007 Technical Report*. Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Foy, P., & Olson, J. F. (2009). *TIMSS 2007 user guide for the international database*. Boston: TIMMS & PIRLS International Study Center, Lynch School of Education, Boston College.

Ganley, C. M., & Vasilyeva, M. (2011). Sex differences in the relation between math performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology, 32*, 235-242.

Glenn, J. (2000). *Before it's too late: A report to the nation*. Washington, DC: Department of Education. Retrieved from http://www.nationalmathandscience.org/sites/default/files/resource/before%20its%20too%20late.pdf on November 6, 2012.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(1464-1480).

Harlow, L. L. (2005). *The essence of multivariate thinking: Basic themes and methods*. Mahwah, NJ: Erlbaum.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

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

IDB Analyzer Version 2 [Computer software and manual]. (2009). Hamburg, Germany: International Association for the Evaluation of Educational Achievement.

Kim, D. H., & Law, H. (2012). Gender gap in maths test scores in South Korea and Hong Kong: Role of family background and single-sex schooling. *International Journal of Educational Development, 32*, 92-103.

Kline, R. (2011). *Principles and practice of structural equation modeling, 3rd Ed.* New York: Guilford Press.

Köller, O., Baumert, J., & Schnabel, K. (2001). Does interest matter? The relationship between academic interest and achievement in mathematics. *Journal for Research in Mathematics Education, 32*(5), 448-470.

Kuenzi, J. J., Matthews, C. M., & Mangan, B. F. (2006). *Science, technology, engineering and mathematics (STEM) education issues and legislative options.* Report of the Congressional Research Service, Library of Congress No. RL33434.

Lachance, J. A., & Mazzocco, M. M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences, 16*, 195-216.

Lepper, M. R., Corpus, J. H., & Iyengar, S. S. (2005). Intrinsic and extrinsic motivational orientations in the classroom: Age differences and academic correlates. *Journal of Educational Psychology, 97*(2), 184-196.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136*(6), 1123-1135.

Liou, P.-Y. (2010). *Cross-National Comparisons of the Association Between Student Motivation for Learning Mathematics and Achievement Linked with School Contexts: Results from TIMSS 2007*. University of Minnesota. UMI Dissertation Publishing.

Liu, O. L., & Wilson, M. (2009). Gender differences in large-scale math assessments: PISA trend 2000 and 2003. *Applied Measurement in Education, 22,* 164-184. doi: 10.1080/08957340902754635.

Lloyd, J. E. V., Walsh, J., & Yailagh, M. S. (2005). Sex differences in performance attributions, self-efficacy, and achievement in mathematics: If I'm so smart, why don't I know it? *Canadian Journal of Education, 28*(3), 384-408.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to linear regression analysis (4th ed.)*. Somerset, NJ: Wiley.

Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eight grades*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115-132.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*, 267-316.

Muthén, L. K. (2011). [How many groups is too many?].

Muthén, L. K., & Muthén, B. O. (2009). *Mplus user's guide, 6th Ed: Statistical analysis with latent variables*. Los Angeles: Muthén & Muthén.

Ngware, M. W., Ciera, J., Abuya, B. A., Oketch, M., & Mutisya, M. (2012). What explains gender gaps in maths achievement in primary schools in Kenya? *London Review of Education, 10*(1), 55-73.

Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal, 48*(5), 1125-1156.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (2008). *TIMSS 2007 Technical Report*. Chestnut Hill, MA: Boston College.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction (3 ed.).* Stamford, CT: Wadsworth.

Rossi, J. S. (1983). Ratios exaggerate gender differences in mathematical ability. *American Psychologist*, 38(3), 348.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54-67.

Sadker, M., & Sadker, D. (1994). *Failing at fairness: How America's schools cheat girls*. New York: Charles Scribner's Sons.

Spender, D., & Sarah, E. (1980). *Learning to lose*. London: Women's Press.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7*(2), 149-162.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of factors*. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics, 5th Ed*. Boston, MA: Allyn & Bacon.

Weaver-Hightower, M. (2003). The "Boy Turn" in research on gender in education. *Review of Educational Research, 73*(4), 471-498.

Williams, T., Ferraro, D., Roey, S., Brenwald, S., Kastberg, D., Jocelyn, L., . . . Stearns, P. (2009). *TIMSS 2007 U.S. Technical Report and User Guide*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

GENERAL APPENDIX

General Discussion

The purpose of this study was to investigate the relationships between gender, socioeconomic status, math attitude, and math achievement from an international perspective. Whereas there has been much research on the individual contributions of gender, SES, and math attitude on math achievement, there is little research that has investigated these relationships within a multivariate framework and from an international perspective. Using publicly available data from a large, international data set (TIMSS 2007), the relationships among these variables was investigated through multiple-group latent variable modeling.

In order to accommodate the large number of groups in the TIMSS 2007 data it was first necessary to partition the number of countries into smaller groups of countries, or clusters. This was accomplished through hierarchical cluster analysis and the resulting clusters were cross-validated on a separate sample quantitatively using discriminant function analysis and qualitatively through the use of several geopolitical indicators. Six different cluster solutions were investigated. Although all six cluster solutions can be considered valid based on the results of the discriminant function analysis, two of the solutions may not have practical value. The nine-cluster solution included a cluster with only a single country (Lebanon), and the four-cluster solution had a cluster that was somewhat large and comprised of 21 countries. For these reasons, these two cluster solutions are not recommended, but the remaining cluster solutions could have varying degrees of use for researchers, based on the questions being investigated.

The cluster solutions from the cluster analysis study were used to inform decisions for the multiple-group latent variable model analyses. Because multiple-group latent variable modeling prefers fewer than 10 groups in order to work, the eight-cluster solution was chosen; this is the only cluster solution in which all of the resulting clusters were comprised of fewer than 10 countries. It is important to note that the cluster solutions appear to have some contextual cohesion, based on several geopolitical indicators, in addition to the mathematical parsimony for the cluster solutions (see Study 1).

The multiple-group latent variable analyses were used to investigate the ways in which sex, math attitudes, and math achievement are related with each other, and how these relationships differ on an international scale. The findings suggest that math self-confidence has a consistently strong relationship with math achievement across countries, and that value of math has a consistently weak relationship with math achievement across countries. Furthermore, the relationship between gender and math achievement, and the relationship between gender and math attitude, is quite small in most cases.

Policy implications from the findings of this study include the need to promote math self-confidence in math curricula. The strength of the relationship between math self-confidence and math achievement is larger than the relationship between math achievement and any other variable in the model, including SES and gender. Since math self-confidence is a more malleable construct than SES, gender, or cognitive ability, and because math self-confidence has such a strong relationship with math achievement, it is imperative that self-confidence is a major consideration for math educators at all levels. Future research on this subject should begin to focus on the development of interventions to increase math self-confidence.

In addition to the need to focus more attention in the math education process onto math self-confidence, the current study suggests a need to develop better measures for math attitude. Suggestions include the use of the Implicit Associations Test (Greenwald et al., 1998) as demonstrated in Nosek and Smyth (2011), although whether or not this would be possible for studies as large as TIMSS remains open to discussion.

A final consideration for this study has to do with a more qualitative interpretation of the findings. That certain countries consistently outperform the United States in math and science achievement is a given; South Korea, Singapore, Hong Kong, Taiwan, and Japan have long been leaders in international math and science assessments. However, there are important cultural differences between these countries and countries like ours (which are defined in the present study as being those countries in the same cluster as ours, such as Australia and Sweden).

A prime example of these cultural differences can be seen in how countries view tutoring, although other specific examples certainly exist. In the United States, students receive tutoring only after they are perceived as being at risk for failure, and expectation is that this supplemental assistance will be funded by the already taxed budget for the school system. Conversely in South Korea there is an expectation that most students receive additional instruction, and this supplemental instruction commonly comes at an additional expense to parents. In essence, students in South Korea are spending nearly twice as much time in school as their American counterparts, and their parents are willing to pay large amounts of money to afford their children this opportunity.

The question becomes whether or not we are willing as a society to adopt the tactics employed by other countries and other cultures to achieve their level of math achievement on international assessments. Furthermore, whether international assessments of math achievement at the fourth and eighth grade levels are an important predictor for a country's well-being has not been established, only assumed. Given the economic situation that has persisted in the United States since 2008, and the general trend toward decreasing funding for education, it seems unlikely that we will be seeing any kind of massive paradigm shift in which the U.S. is willing to spend more money on education, either at the national, state, or individual levels, which is likely what would be necessary to increase math achievement scores by large amounts.

Because it is unlikely that we will be spending more money on education, it is important that we spend the money we do allocate for education efficiently and in a way that it will have the most positive impact. By increasing the attention paid to self-confidence in learning math during the teaching of mathematics, we may be able to decrease the gaps we see between ourselves and countries we desire to emulate in terms of math achievement, without committing tremendous additional resources to the endeavor.