Boston College

The Graduate School of Education

Department of Educational Research, Measurement, and Evaluation

EXAMINING SCHOOL EFFECTIVENESS AT THE FOURTH GRADE: A HIERARCHICAL ANALYSIS OF THE THIRD INTERNATIONAL MATHEMATICS AND SCIENCE STUDY (TIMSS)

Dissertation

by

STEVEN EDWARD STEMLER

submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

May 2001

Abstract

Examining School Effectiveness at the Fourth Grade: A Hierarchical Analysis of the Third International Mathematics and Science Study (TIMSS)

Dissertation by: Steven E. Stemler

Advisor: Prof. Ina V.S. Mullis

This study explored school effectiveness in mathematics and science at the fourth grade using data from IEA's Third International Mathematics and Science Study (TIMSS). Fourteen of the 26 countries participating in TIMSS at the fourth grade possessed sufficient between-school variability in mathematics achievement to justify the creation of explanatory models of school effectiveness while 13 countries possessed sufficient between-school variability in science achievement. Exploratory models were developed using variables drawn from student, teacher, and school questionnaires. The variables were chosen to represent the domains of student involvement, instructional methods, classroom organization, school climate, and school structure. Six explanatory models for each subject were analyzed using two-level hierarchical linear modeling (HLM) and were compared to models using only school mean SES as an explanatory variable. The amount of variability in student achievement in mathematics attributable to differences between schools ranged from 16% in Cyprus to 56% in Latvia, while the amount of between-school variance in science achievement ranged from 12% in Korea to 59% in Latvia. In general, about one-quarter of the variability in mathematics and science

achievement was found to lie between schools. The research findings revealed that after adjusting for differences in student backgrounds across schools, the most effective schools in mathematics and science had students who reported seeing a positive relationship between hard work, belief in their own abilities, and achievement. In addition, more effective schools had students who reported less frequent use of computers and calculators in the classroom. These relationships were found to be stable across explanatory models, cultural contexts, and subject areas. This study has contributed a unique element to the literature by examining school effectiveness at the fourth grade across two subject areas and across 14 different countries. The results indicate that further exploration of the relationship between school effectiveness and student locus of control warrants serious consideration. Future research on school effectiveness is recommended, perhaps using trend data and looking at different grade levels.

ACKNOWLEDGMENTS

I would like to dedicate this dissertation to my wife and best friend, Karen Stemler. Her sincerity, brilliance, warmth, humor, and love have made all of the difference in my life. Karen has always pushed me to be more than I thought I could be and has always been there whenever I needed her. She has been the source of infinite inspiration, support, and laughter and I am truly grateful to have her in my life.

Next, I would like to thank my dissertation director, Dr. Ina V.S. Mullis. She devoted an incredible amount of time toward this project and gave me insightful feedback on the numerous drafts of this dissertation faster than I had any right to expect. Dr. Mullis' intellect, enthusiasm, and positive attitude are qualities that I hope to emulate throughout my life and throughout my professional career. She has been an outstanding mentor and I am very grateful for everything she has done for me.

I would also like to thank my two readers, Dr. Michael Martin and Dr. Irwin Blumer. Their unique perspectives helped enlighten me to various elements that made this dissertation a much stronger piece of research than I could have ever accomplished on my own. In addition, I would like to gratefully acknowledge Craig Hoyle, Kelvin Gregory, and Eugene Gonzalez.

Finally, I wish to thank my family and friends for their support throughout this process. My parents, Gary and Cathy Stemler, have always stood behind me no matter what decisions I have made and their unconditional love for me has propelled me through life. I am grateful for the close friendships that have developed with Cathy Horn, Laura

O'Dwyer, Damian Bebell, Rob Connor, Ed Fierros, and Martin Pierre. We have shared many good times together and each of these people has broadened my perspective on life.

Table of Contents

CHAPTER 1 - INTRODUCTION **1**

THE PROBLEM 1
ANALYZING SCHOOL EFFECTIVENESS 2
WHAT IS THE IEA? 5
WHAT IS TIMSS? 7
RESEARCH QUESTIONS 11
SIGNIFICANCE OF THE STUDY 13

CHAPTER 2 – LITERATURE REVIEW **16**

OVERVIEW 16
THE CONSTRUCT OF SCHOOL EFFECTIVENESS 16
THE HISTORICAL EVOLUTION OF SCHOOL EFFECTIVENESS RESEARCH 18
Model Mis-specification *20*
The Unit of Analysis Problem *22*
External Validity *23*
Longitudinal v. Cross-sectional Research Designs *24*
UNDERSTANDING THE CONTEXT AND CONTENT OF TIMSS 26
HOW CAN TIMSS CONTRIBUTE TO SCHOOL EFFECTIVENESS RESEARCH? 33
WHY USE HIERARCHICAL LINEAR MODELING (HLM) FOR ANALYSIS? 35
PARTITIONING THE VARIANCE IN ACHIEVEMENT 38
THEORETICAL MODELS FOR EXPLORING SCHOOL EFFECTIVENESS 41
Dependent Variables *42*
Student Background Variables *43*
Student-level Explanatory Variables *46*
Teacher-level/Classroom-level Explanatory Variables *49*
School-level Explanatory Variables *52*
SUMMARY 54

CHAPTER 3 - METHODOLOGY **56**

OVERVIEW 56
CRITERIA FOR VARIABLE INCLUSION 57
Theoretical Importance *57*
Response Rates *59*
Variability *60*
Model Parsimony *61*
Measurement Error *61*
CONSTRUCTING EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS 62
Recoding *67*
Standardization of Explanatory Variables *67*
Control Block *68*
Block 1 - Student Involvement *68*
Block 2 - Instructional Methods *70*
Block 3 – Classroom Organization *72*
Block 4 - School Climate *73*
Block 5 - School Resources *75*
PRELIMINARY SCREENING ANALYSIS 76
Sampling Procedures *77*
Between-School and Within-Country Variability in Achievement *78*
Missing Questionnaire Data *79*

Countries Under Investigation *80*
DATABASE CHARACTERISTICS 80
Plausible Values *81*
Sampling Weights *82*
Intra-class Correlation *84*
HIERARCHICAL LINEAR MODELING 85
Using A Two-Level Hierarchical Linear Model *87*
Level 1: Within-school student-level equation *88*
Level 2: Between-school equations - The Unconditional Model *89*
Level 2: Between-school school equation – The Conditional Model *90*
Fixed v. Random Coefficients *91*
Variable Centering *92*
SUMMARY 92

CHAPTER 4 – RESULTS 94

OVERVIEW 94
DESCRIPTIVE STATISTICS 95
RESEARCH QUESTIONS IN MATHEMATICS 99
Is there sufficient variability in mathematics achievement between schools in each country to develop a model for explaining that variability? *99*
What factors that contribute to schools' effectiveness can be identified and are any of them open to policy manipulation? *106*
Are the variables associated with effective schools at the fourth grade stable across different cultural contexts? *109*
After correcting for the differences in student intake across schools with regard to SES, how much variance in mathematics and science achievement across schools can be explained by each explanatory model? *117*
RESEARCH QUESTIONS IN SCIENCE 118
Is there sufficient variability in science achievement between schools in each country to develop a model for explaining that variability? *118*
What factors that contribute to schools' effectiveness can be identified and are any of them open to policy manipulation? *124*
Are the variables associated with effective schools at the fourth grade stable across different cultural contexts? *127*
After correcting for the differences in student intake across schools with regard to SES, how much variance in science achievement across schools can be explained by each explanatory model? *135*
SUMMARY OF FINDINGS IN MATHEMATICS 136
SUMMARY OF FINDINGS IN SCIENCE 138
COMPARISON OF FINDINGS IN MATHEMATICS AND SCIENCE 141
SUMMARY 141

CHAPTER 5 – CONCLUSIONS 143

OVERVIEW 143
SUMMARY AND INTERPRETATION OF THE FINDINGS 143
PRACTICAL APPLICATION OF RESULTS 150
DIRECTIONS FOR FUTURE RESEARCH 153
LIMITATIONS OF THE STUDY 155
SUMMARY 157

REFERENCES 158

APPENDIX A - DESCRIPTIVE STATISTICS 170

APPENDIX B - COUNTRY LEVEL RESULTS FOR EACH MODEL TESTED IN MATHEMATICS 185

APPENDIX C - COUNTRY LEVEL RESULTS FOR EACH MODEL TESTED IN SCIENCE......200

List of Tables

TABLE 2.1 DISTRIBUTION OF MATHEMATICS ITEMS BY CONTENT REPORTING CATEGORY AND PERFORMANCE EXPECTATION – POPULATION 1* 29
TABLE 2.2 DISTRIBUTION OF SCIENCE ITEMS BY CONTENT REPORTING CATEGORY AND PERFORMANCE EXPECTATION – POPULATION 1* 30
TABLE 3.1 VARIABLES ASSOCIATED WITH EACH EXPLANATORY BLOCK - MATHEMATICS 65
TABLE 3.2 VARIABLES ASSOCIATED WITH EACH EXPLANATORY BLOCK - SCIENCE 66
TABLE 3.3 COUNTRIES MEETING CRITERIA FOR INCLUSION 77
TABLE 4.1 DESCRIPTIVE STATISTICS FOR THE UNITED STATES AT THE UPPER GRADE OF POPULATION 1 98
TABLE 4.2 COMPARISON OF VARIANCE COMPONENTS IN THE FULL AND RESTRICTED UNCONDITIONAL MODELS IN FOURTH GRADE MATHEMATICS 101
TABLE 4.3 COMPARISON OF BETWEEN-SCHOOL VARIANCE IN FOURTH GRADE MATHEMATICS OBSERVED IN THE FULL AND RESTRICTED UNCONDITIONAL MODELS 102
TABLE 4.4 COMPARISON OF PARAMETER VARIANCE AND RELIABILITY ACROSS FULL AND RESTRICTED UNCONDITIONAL MODELS IN FOURTH GRADE MATHEMATICS 103
TABLE 4.5 COMPARISON OF MEANS AND STANDARD DEVIATIONS OF THE FULL AND RESTRICTED SAMPLES IN FOURTH GRADE MATHEMATICS 104
TABLE 4.6 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN THE UNITED STATES 107
TABLE 4.7 NUMBER OF VARIABLES FOUND TO BE SIGNIFICANT AT THE .10 LEVEL ACROSS MODELS IN FOURTH GRADE MATHEMATICS (14 COUNTRIES) 110
TABLE 4.8 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 1 112
TABLE 4.9 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 2 112
TABLE 4.10 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 3 113
TABLE 4.11 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 4 114
TABLE 4.12 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 5 115
TABLE 4.13 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE MATHEMATICS – MODEL 6 116
TABLE 4.14 BETWEEN-SCHOOL VARIANCE IN MATHEMATICS ACHIEVEMENT EXPLAINED BY EACH EXPLANATORY MODEL IN FOURTH GRADE MATHEMATICS 118
TABLE 4.15 COMPARISON OF VARIANCE COMPONENTS IN THE FULL AND RESTRICTED UNCONDITIONAL MODELS IN FOURTH GRADE SCIENCE 120
TABLE 4.16 COMPARISON OF BETWEEN-SCHOOL VARIANCE IN FOURTH GRADE SCIENCE OBSERVED IN THE FULL AND RESTRICTED UNCONDITIONAL MODELS 121
TABLE 4.17 COMPARISON OF PARAMETER VARIANCE AND RELIABILITY ACROSS FULL AND RESTRICTED UNCONDITIONAL MODELS IN FOURTH GRADE SCIENCE 122
TABLE 4.18 COMPARISON OF MEANS AND STANDARD DEVIATIONS IN THE FULL AND RESTRICTED SAMPLES IN SCIENCE 123
TABLE 4.19 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN THE UNITED STATES 125
TABLE 4.20 NUMBER OF VARIABLES FOUND TO BE SIGNIFICANT AT THE .10 LEVEL ACROSS MODELS IN FOURTH GRADE SCIENCE (13 COUNTRIES) 128
TABLE 4.21 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 1 130

TABLE 4.22 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 2 130
TABLE 4.23 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 3 131
TABLE 4.24 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 4 132
TABLE 4.25 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 5 133
TABLE 4.26 INDICATION OF THE DIRECTION OF STATISTICALLY SIGNIFICANT RESULTS FOR FOURTH GRADE SCIENCE – MODEL 6 134
TABLE 4.27 BETWEEN-SCHOOL VARIANCE IN SCIENCE ACHIEVEMENT EXPLAINED BY EACH EXPLANATORY MODEL IN FOURTH GRADE SCIENCE 136
TABLE A.1 DESCRIPTIVE STATISTICS FOR AUSTRALIA AT THE UPPER GRADE OF POPULATION 1 171
TABLE A.2 DESCRIPTIVE STATISTICS FOR CANADA AT THE UPPER GRADE OF POPULATION 1 172
TABLE A.3 DESCRIPTIVE STATISTICS FOR CYPRUS AT THE UPPER GRADE OF POPULATION 1 173
TABLE A.4 DESCRIPTIVE STATISTICS FOR THE CZECH REPUBLIC AT THE UPPER GRADE OF POPULATION 1 174
TABLE A.5 DESCRIPTIVE STATISTICS FOR GREECE AT THE UPPER GRADE OF POPULATION 1 175
TABLE A.6 DESCRIPTIVE STATISTICS FOR HONG KONG AT THE UPPER GRADE OF POPULATION 1 176
TABLE A.7 DESCRIPTIVE STATISTICS FOR IRAN AT THE UPPER GRADE OF POPULATION 1 177
TABLE A.8 DESCRIPTIVE STATISTICS FOR IRELAND AT THE UPPER GRADE OF POPULATION 1 178
TABLE A.9 DESCRIPTIVE STATISTICS FOR KOREA AT THE UPPER GRADE OF POPULATION 1 179
TABLE A.10 DESCRIPTIVE STATISTICS FOR LATVIA AT THE UPPER GRADE OF POPULATION 1 180
TABLE A.11 DESCRIPTIVE STATISTICS FOR NEW ZEALAND AT THE UPPER GRADE OF POPULATION 1 181
TABLE A.12 DESCRIPTIVE STATISTICS FOR PORTUGAL AT THE UPPER GRADE OF POPULATION 1 182
TABLE A.13 DESCRIPTIVE STATISTICS FOR SLOVENIA AT THE UPPER GRADE OF POPULATION 1 183
TABLE A.14 DESCRIPTIVE STATISTICS FOR THE UNITED STATES AT THE UPPER GRADE OF POPULATION 1 184
TABLE B.1 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN AUSTRALIA 186
TABLE B.2 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN GRADE FOUR MATHEMATICS IN CANADA 187
TABLE B.3 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN CYPRUS 188
TABLE B.4 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN THE CZECH REPUBLIC 189
TABLE B.5 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN GREECE 190
TABLE B.6 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN HONG KONG 191
TABLE B.7 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN IRAN 192
TABLE B.8 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN IRELAND 193
TABLE B.9 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN KOREA 194
TABLE B.10 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN LATVIA 195
TABLE B.11 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN NEW ZEALAND 196
TABLE B.12 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN PORTUGAL 197

TABLE B.13 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN SLOVENIA 198
TABLE B.14 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN MATHEMATICS AT THE FOURTH GRADE IN THE UNITED STATES 199
TABLE C.1 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN AUSTRALIA 201
TABLE C.2 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN CANADA 202
TABLE C.3 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN CYPRUS 203
TABLE C.4 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN THE CZECH REPUBLIC 204
TABLE C.5 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN THE GREECE 205
TABLE C.6 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN IRAN 206
TABLE C.7 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN IRELAND 207
TABLE C.8 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN KOREA 208
TABLE C.9 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN LATVIA 209
TABLE C.10 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN NEW ZEALAND 210
TABLE C.11 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN PORTUGAL 211
TABLE C.12 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN SLOVENIA 212
TABLE C.13 RESULTS OF SEVEN EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS IN SCIENCE AT THE FOURTH GRADE IN THE UNITED STATES 213

List of Figures

FIGURE 4.1 GRAPHIC DISPLAY OF THE BETWEEN-SCHOOL VARIANCE IN THE RESTRICTED UNCONDITIONAL MODEL IN FOURTH GRADE MATHEMATICS 105
FIGURE 4.2 GRAPHIC DISPLAY OF THE BETWEEN-SCHOOL VARIANCE IN THE RESTRICTED UNCONDITIONAL MODEL IN FOURTH GRADE SCIENCE 124

CHAPTER 1 - INTRODUCTION

THE PROBLEM

The purpose of this dissertation study is to explore educational practices at the fourth grade that are associated with schools meeting “world-class standards” as defined by their performance on the Third International Mathematics and Science Study (TIMSS). The theoretical framework driving this investigation is drawn from the school effectiveness and school improvement literature. Sammons (1999) points out that, “Over the last twenty years or so increasing academic interest has been devoted to the related research fields of school effectiveness and improvement. Considerable evidence has accumulated at both the primary and secondary levels of the existence of significant differences in schools’ effects on students’ educational outcomes (see reviews by McPherson, 1992; Reynolds, 1992; Scheerens, 1992; Mortimore, 1993).” (pp. 25-26).

Although the exact origins of school effectiveness research as a field may be debated, there is a general consensus in the literature (Arnold & Sedlacek, 1995; Creemers, Reynolds, & Swint, 1994; Sammons, 1999) that the field was developed in reaction to seminal studies by Coleman et al. (1966); the Plowden Committee (1967); and Jencks et al. (1972). The first of these studies dealt with the *Equality of Educational Opportunity* in the United States and is known as the Coleman Report in deference to its primary author. While the technical conclusions of that study stated that once socioeconomic factors were controlled for, school factors did not explain much variance in student achievement, these conclusions were largely taken to mean that schools made no difference in the lives of children. Shortly thereafter, the Plowden committee released a report on *Children and Their Primary Schools* in England that came to the same conclusions as the Coleman report, lending an element of external validity to the findings. In 1972, Jencks et al. conducted another study that came up with conclusions that corroborated the findings of these two reports. It is largely through dissatisfaction with these conclusions that the field of school effectiveness research was born.

In the past two decades, research in the field of school effectiveness has increased substantially. At least in part, this may be attributed to the increasing interest in accountability that has affected students, teachers, and schools alike (Sammons, 1999). In the United States and the United Kingdom in particular, there has been a general movement towards ‘high standards’ and ‘value-added’ education, which has led to many policies specifically aimed at holding schools accountable for educational outcomes.

ANALYZING SCHOOL EFFECTIVENESS

Historically, what exactly constitutes an “effective school” has varied across studies. Some studies have simply examined mean student achievement across schools and identified schools with the highest mean student achievement as “effective” (see Sammons, 1999 for a review of these studies). This approach is only defensible, however, when the effect of schools is independent of pupil background (Aitkin & Longford, 1986). “If the value added depends, say, on students’ socioeconomic status (SES), the effect of a particular school cannot be described without first specifying the SES of the student to whom the effect applies.” (Raudenbush & Bryk, 1989, p. 207).

Furthermore, “since children are not assigned at random to schools, minimizing bias requires that relevant pre-existing differences among students must be controlled lest they confound inferences about school effects.” (Raudenbush & Bryk, 1989, p. 208).

Many studies of school effectiveness have been analyzed using Ordinary Least Squares (OLS) regression analyses in order to control for socioeconomic status (SES) (see for example Coleman et al., 1966; Jencks et al., 1972). Yet, a growing body of research suggests that using Ordinary Least Squares (OLS) regression to analyze school effectiveness data is often inappropriate (Bryk & Raudenbush, 1992; Cronbach, 1976; Cronbach & Webb, 1975; Haney, 1980; Raudenbush & Bryk, 1986). The problem with using an OLS analysis in a school effectiveness study arises from the fact that students are nested within classrooms, which are nested within schools. Research has shown that, “shared experience (e.g., being in the same classroom) causes a dependence of observations.” (Kreft & De Leeuw, 1998, p. 9). In other words, thirty students in the same classroom are more like each other than thirty students selected at random from a school. This dependence of observations is also known as intra-class correlation. If intra-class correlations are present (as when dealing with clustered data), then the OLS assumption of independent errors is violated (Kreft & De Leeuw, 1998). As a result, the standard errors of the coefficients will be underestimated, leading to an increase in the probability of a Type-I error (Pedhazur, 1997).

One approach to this problem is to abandon examination of within-group variance and focus solely on between-group variation. This technique is known as aggregate regression. Unfortunately, because aggregate regression ignores all within-school variation, it tends to throw away a large amount of possibly important variance (Kreft & De Leeuw, 1998). Consequently, this approach is not satisfactory.

Researchers have also turned to the statistical technique known as the Analysis of Covariance (ANCOVA) to help recover some of the information lost in the presence of intra-class correlation. ANCOVA takes into account pre-existing group differences, thereby allowing groups to have unequal intercepts. This technique has advantages in that it allows the researcher to better understand the true nature of the effect controlling for prior differences. ANCOVA is best used, however, in situations that are designed as true experiments (i.e., either subjects are randomly assigned to treatment groups or treatments are randomly assigned to subjects), a situation that is rare in educational research for a variety of reasons (Cook & Campbell, 1979).

In order to combat the limitations of OLS and ANCOVA, a technique known as Hierarchical Linear Modeling (HLM) has been applied to the analysis of school effectiveness data. HLM was designed to allow researchers to, "simultaneously handle measurements made at different levels of a hierarchy." (Kreft & De Leeuw, 1998, p. 1). HLM involves a linear model in which group membership is taken into account and estimates of within group and between group variation can be calculated. This allows the researcher to identify the amount of variance explained within and between groups by variables that are measured at different levels of a hierarchy. As a result, any number of control variables may be entered in at the first level of analysis (student differences), and then any number of explanatory variables can be used in modeling the second level of analysis (school differences).

Due to these methodological advantages, school effectiveness research has exploded as a field of study in the past 15 years. This dissertation will employ HLM analyses to data drawn from the IEA's Third International Mathematics and Science Study (TIMSS) in order to explore school effectiveness in an international context.

WHAT IS THE IEA?

The IEA was established in 1961, "as a cooperative venture between educational *research institutes* in a dozen countries whose directors had met regularly since the late 1950s and had begun to contemplate the task of evaluating national systems of education on a comparative basis." (Husen, 1996, p. 208) At that point in history, "no cross-national surveys in education by means of representative national samples had been attempted." (Husen, 1996, p. 209). With the launching of the Sputnik satellite in 1957, the Soviet Republic established their technological prowess in a way that would forever influence the world's view of science and technology. To industrialized nations of the world, Sputnik was taken as tangible evidence of the outstanding teaching of science in the Russian schools (Husen, 1996). As a result, governments throughout the world began to allocate a tremendous amount of resources to their own schools, particularly in the areas of mathematics and science, in order to keep pace with the Russians. The time soon came when these officials wanted to see not only how much progress their children had made relative to a set standard, but also how they stacked up against their peers in other countries. The launching of Sputnik was just the push needed to encourage the directors of these educational research institutes to contemplate designing a study evaluating national systems of education on a comparative basis (Husen, 1996).

The first subject area that would become the testing ground of international comparison was mathematics.

> The reasons for the decision to choose mathematics as the first subject for a full-scale survey were obvious. In the first place, it was a subject area whose language was universal and whose curriculum content showed a high degree of cross-national overlap. Secondly, it was a subject where the development of standardized tests appeared to be rather straightforward and simple and did not entail problems of the kind later encountered in developing tests in, for instance, civic education. We could hardly foresee the controversies among mathematics educators about inclusion of specific content areas and the relative weight that should be assigned to them (Husen, 1996, p. 211).

In the nearly forty years since the first IEA study of mathematics achievement, the influence of mathematics, science and technology has been increasing at an exponential rate. One major reason for the interest in assessing mathematics, science, and technology is because countries with high rates of unemployment are concerned with preparing youths for jobs that require technical skill in order to increase productivity in light of declining economic and international trade issues (Atkin & Black, 1997). In addition, advances in computer technology sweeping across the globe have served to highlight the importance of mathematics and science literacy. As a result of these and other factors, an increasing number of countries have been eager to participate in studies involving international comparisons of student achievement in order to evaluate their students using a global yardstick.

WHAT IS TIMSS?

The most recent and advanced study of international mathematics and science achievement was the IEA's Third International Math and Science Study (TIMSS). In all, forty-five countries collected data in more than 30 different languages (Beaton et al., 1996). Students representing five different grade levels from three different populations were tested (see Martin & Kelly, 1996 for more information about the grades tested in specific countries). Population 1 consisted of all students enrolled in the two adjacent grades that contained the largest proportion of nine year-olds at the time of testing. Across most countries, this corresponded to grades three and four. Population 2 consisted of all students enrolled in the two adjacent grades that contain the largest proportion of 13 year-olds at the time of testing. Across most countries this corresponded to grades seven and eight. Finally, Population 3 consisted of students enrolled in their final year of secondary education. Although the amount of schooling received by students in this population varied within countries, in a number of countries, including the U.S., this corresponded to grade 12. Participation in the study required all countries to administer the survey to students at Population 2, while participation was voluntary at the other two populations.

As with other international studies before it, the purpose of TIMSS was multifaceted. Among the many aims of TIMSS were to identify countries along a continuum of mathematics and science achievement, to determine the factors associated with effective schools, and to identify the factors positively influencing student achievement. Furthermore, it was hoped that key findings would emerge that would be generalizable to other populations in other contexts.

The TIMSS results were released from late 1996 (grades 7 and 8) to early 1997 (grades 3 and 4 and grade 12). The results for all three populations across both subjects provided a chain of overlapping performances, where most countries had average achievement similar to a cluster of other countries, but from the beginning to the end of the chain there were substantial differences (Mullis et al., 1997).

The TIMSS data has fueled many secondary analyses, including a recent study of *Effective Schools in Science and Mathematics* conducted by Martin, Mullis, Gregory, Hoyle, & Shen (2000). Their study explored a number of explanatory variables associated with effective schools in science and mathematics at the upper grade of Population 2 (i.e., the eighth grade in most countries). The present dissertation study is designed as a complementary and companion volume to that study, with the present investigation being focused on effective schools in mathematics and science at the upper grade of Population 1 (i.e., the fourth grade in most countries). Although this dissertation derives its basic design from the Martin et al. (2000) study, the explanatory variables that will be explored are tailored toward the population under investigation.

In addition to fitting in with prior research, this study is important in light of the fact that TIMSS 2003 is currently being designed. TIMSS 2003 will test students in the fourth and eighth grades in approximately 50 countries and will provide valuable trend data. Consequently, the results of the present investigation may serve to highlight areas warranting further exploration in the questionnaires designed for TIMSS 2003.

The present dissertation study will focus upon the results from the students who were assessed at the upper grade of Population 1, corresponding to the fourth year of formal schooling in most countries. Twenty-six countries completed the steps necessary to be included in the data analyses for Population 1. Most of the preliminary findings were descriptive; however, some relationships among variables were explored. For instance, "Those students who reported either liking mathematics or liking it a lot generally had a higher achievement than students who reported disliking it to some degree." (Mullis et al., 1997, p. 4). In addition, "fourth-grade students who reported having more educational resources in the home had higher mathematics achievement than those who reported little access to such resources." (Mullis et al., 1997, p. 4). While these relationships are revealing, prior research has taught us that any single factor by itself is unlikely to have a dramatic influence upon student achievement (Mullis, Jenkins, & Johnson, 1994; Welch, Walberg, & Fraser, 1986). Student achievement is affected by several factors including home background, school environment, and national context.

Students were assessed in the major subject areas of mathematics and science. Each of these subject areas had a number of content areas that were covered by the assessments. In mathematics, 79 multiple-choice, 15 short-answer, and 8 extended-response items covered six content areas. The content areas and the percentage of test items devoted to each were: (i) whole numbers (25%); (ii) fractions and proportionality (21%); (iii) measurement, estimation, and number sense (20%); (iv) data representation, analysis, and probability (12%); (v) geometry (14%); and (vi) patterns, relations, and functions (10%). The percentage of items by content category and performance expectation is listed in Appendix A of Mullis et al. (1997). In science, 74 multiple-choice, 13 short-answer, and 10 extended-response items covered four content areas. The content areas and the percentage of test items devoted to each were: (i) earth science (18%); (ii) life science (42%); (iii) physical science (31%); (iv) environmental issues and the nature of science (9%). The percentage of items by content category and performance expectation is listed in Appendix A of Martin et al. (1997).

In addition to the information gathered from the achievement tests, TIMSS collected extensive information on a host of background variables in order to gain a better understanding of the relative influence of each of these factors (Mullis et al., 1997). Students participating in the study completed questionnaires related to such topics as their interest in mathematics and science, specific instructional strategies used in the classroom, and the resources in their home. In addition, teachers and administrators completed questionnaires related to such topics as teacher experience, school climate, and structural features of the school (e.g., average class size). In all, data was collected on over 1500 variables across the three populations, making the database an extremely valuable resource for secondary analysis.

Some of the important variables related to school effectiveness reported by Martin et al. (2000) include the amount of homework students do, students attitudes toward the subject area, class size, minor and major behavioral violations by the students, urban location, and the overall mean SES for the school. In particular, students' future aspirations was the most prevalent predictor of school effectiveness in mathematics (significant in 17 countries) while the strongest predictor of school effectiveness in science came with students who reported frequently doing homework in three subject areas. Each of these predictors was important across a wide range of countries and across the subject areas of both mathematics and science. In addition, class size was a significant predictor of school effectiveness in eight countries in mathematics, but was a significant predictor of school effectiveness in only two countries in science (Martin et al., 2000).

RESEARCH QUESTIONS

This dissertation study will investigate the factors associated with effective schools for each country participating in TIMSS at the fourth grade. In TIMSS, at least 150 schools were to be selected per target population and all population samples were to have an effective sample size of 400 students (Foy, Rust, & Schleicher, 1996). This study will begin with a review of the literature in order to determine the factors associated with effective schools in mathematics and science. Theoretical models will then be developed based upon the research and the explanatory power of the models will be tested using HLM. The research questions to be addressed in this dissertation are listed below:

1. Is there sufficient variability between schools in each country to develop a model for explaining that variability?

2. As Gray (1989) has noted, “What factors that contribute to schools’ ‘effectiveness’ can be identified and, crucially, are any of them open to policy manipulation?” (p. 130)
3. Are the variables associated with effective schools at the fourth grade stable across different cultural contexts?
4. After correcting for the differences in student intake across schools with regard to SES, how much variance in mathematics and science achievement across schools can be explained by: student involvement, instructional methods, classroom organization, school climate, and school structural features?
5. Are the variables associated with effective schools at the fourth grade stable across different subject areas (i.e., are the same variables important in both mathematics and science)?

The model employed to test the hypotheses will use variables drawn from different levels of the hierarchy. A two-level hierarchical linear model will be developed using the individual student level variables at level one and school level explanatory variables at level two. HLM is useful for this type of research because it will allow level one variables to be used for control (such as SES), while investigating the variance in achievement across level two variables. Furthermore, HLM allows the variance in achievement to be partitioned across various levels of the hierarchy.

SIGNIFICANCE OF THE STUDY

This dissertation study will be important to a wide variety of audiences including policy makers, educators, administrators, and researchers. In particular, the study will be of interest to researchers in that it will contribute to the knowledge base on school effectiveness by adding the element of international comparisons across two subject areas at the fourth grade. In addition, the study continues the tradition of IEA research into policy relevant areas by complementing the study conducted by Martin et al. (2000) on school effectiveness at the eighth grade. When interpreted jointly, these two studies will provide an unprecedented opportunity for examining school effectiveness internationally across two different grade levels.

Furthermore, the findings will be particularly relevant to policy makers in allowing for greater insight into the strengths of particular educational policies across different cultural contexts.   For example, if class size is found to be a powerful predictor of school effectiveness in mathematics in one country but makes no difference in another country, this will signal a need for researchers to probe deeper into the mechanisms underlying these relationships. If a predictor is found to be important across countries, however, then policies can be confidently formulated on the basis of extremely solid findings. By the same token, if a particular educational policy is a significant predictor of school effectiveness in mathematics, but this same policy is not a significant predictor of school effectiveness in science, then this will point to the need for further exploration of the mechanisms underlying the subject specific nature of the policy.

Finally, both teachers and administrators could potentially learn from the findings of this study. The study will focus on a number of factors related to instructional methods, classroom organization, school climate, and school structure; all topics of great interest and relevance to teachers and administrators alike. There is much to be learned from the findings regardless of their outcome. For instance, if particular instructional methods related to the frequency of homework assignments and the frequency of student testing are found to be related to school effectiveness, this finding may have immediate implications for teachers and administrators. If, however, these variables are not found to be associated with school effectiveness, this conclusion may be equally important.

One of the greatest strengths of the study is the fact that the data being analyzed may truly be characterized as a representative sample of students from each of the participating countries. As a consequence, the results of the study will have a high degree of external validity. Yet, as Sammons (1999) points out,

> The limitations as well as the strengths of school effectiveness approaches need to be recognized and acknowledged, so that unrealistic expectations about radical changes in students' educational performance are not raised, and yet without dampening concern to explore and document how schools can make a difference, and avoiding a return to the sociological determinism which has often played down the school's influence, and led to a culture of low expectations for particular groups of students in some schools (p. x).

We must remain cautiously optimistic about any of the findings from this study. In addition, it must be kept in mind that, "Questions about values in education, the purposes of schooling, the quality of students' educational experiences, and of what constitutes a 'good school' rightly remain the subject of much argument and are unlikely to be resolved." (White & Barber, 1997).

The goal of this study is to investigate a particular theoretical model of school effectiveness at grade four and to identify the explanatory power and generalizability of that model. The model will be tested across two different curricular domains and across a variety of countries participating in TIMSS. In the end, it is important to recognize that the research itself cannot provide a ‘solution’, but rather it may serve to enlighten, and to provide insight into what may and may not be expected at best and at worst by altering particular school policies.

CHAPTER 2 – LITERATURE REVIEW

OVERVIEW

One of the biggest controversies surrounding school effectiveness research is the very definition of what constitutes an “effective” school. This issue has historically been the source of spirited debate and therefore seems an appropriate place to begin a review of the literature. Once the construct of school effectiveness has been operationally defined, a review of the historical evolution of school effectiveness research will be presented. This review will touch upon several major studies of school effectiveness undertaken to date and will delineate the key issues faced by the field throughout the course of its development. The benefits of using the TIMSS database for school effectiveness research will then be discussed. Next, the state-of-the-art technique for analyzing school effectiveness research (i.e., hierarchical linear modeling) will be presented, followed by a brief explanation of variance partitioning. Finally, a review of the most prevalent variables used in prior studies of school effectiveness research will be presented and discussed.

THE CONSTRUCT OF SCHOOL EFFECTIVENESS

What does it mean to say that a school is good? In order to answer such a question, there must be a clear standard or criterion against which the school is to be judged. There are a number of suitable criteria that could be used depending upon one’s beliefs about the purpose of school. To say that a school is good could mean that it produces students who are socially well adjusted, emotionally stable, intellectually adept, physically fit, spiritually fulfilled, or any combination thereof. In light of these diverse criteria, the judgment as to how good a school is arguably lies in the eye of the beholder. It is for this reason that Edmonds (1979a and b) introduced the term “effectiveness” into the literature. “Rather than attempting to define ‘good’, and thus by implication ‘bad’ schools, school effectiveness research focuses deliberately on the narrower concept of effectiveness which concerns the achievement of educational goals measured by student progress.” (Sammons, Thomas, & Mortimore, 1997, p. 8). Although this definition suffers from other problems such as “effective for whom?” and “effective for what?” (see criticisms by Slee & Weiner, 1998; Atkin & Black, 1997; Ranson, 1997), it at least forces researchers to adopt some measurable criteria for evaluating school effectiveness. To that end, West & Hopkins (1996) have proposed that school effectiveness research should narrow its focus down to four domains: (i) student achievement; (ii) student experiences; (iii) teacher and school development; and (iv) community involvement. Yet, when it comes to operationalizing the construct of effectiveness, many studies tend to place the most emphasis on measures of student achievement as the outcome variable due to its capacity for reliable measurement.

For the purposes of this dissertation study, effectiveness will be defined on the basis of schools’ average performance on TIMSS at the fourth grade. In interpreting the findings from this study, it should be kept in mind, however, that although test scores may be what defines a school as effective, they can only be a part of what defines a school as good. Understanding the policies and practices that make for effective schools in various contexts is an important part of this dissertation study, however, users of the results must be cautious that they do not simply attempt to imitate what is working in other contexts without reflecting upon how high test scores fit within their own philosophy of education and their own educational context (see caveats by Atkin & Black, 1997; Reynolds, 2000; Slee & Weiner, 1998).

THE HISTORICAL EVOLUTION OF SCHOOL EFFECTIVENESS RESEARCH

Although administrators around the world have been evaluating the extent to which their schools produce quality students for hundreds of years (Madaus, Scriven, & Stufflebeam, 1983), the development of a formal field of school effectiveness research is fairly recent. The field is most often traced back to the publication of a report by Coleman et al. in 1966 on the *Equality of Educational Opportunities* in the United States (Burstein, 1980; Bryk & Raudenbush, 1992; Sammons, 1999; Teddlie & Reynolds, 2000). The Coleman report, as it is often called, was noteworthy in that it represented the first major piece of educational research to conclude that the impact of schools upon student achievement was minimal after controlling for the socioeconomic status of the students (Coleman et al., 1966).

Shortly after the release of the Coleman report in the United States, the results from a study called *Children and their Primary Schools* (1967) in England, known also as the Plowden Report, were released. The conclusions of the Plowden Report echoed the sentiments of the Coleman report that home background factors explained more of the variance in school achievement than did school factors (Teddlie, Reynolds, & Sammons, 2000).

The uneasiness many educators were feeling about these conclusions was intensified when Jencks et al. (1972) conducted a study in the United States whose results corroborated the findings of the Coleman report. Researchers disagree about whether the studies by Coleman et al., the Plowden Committee, and Jencks et al. served as springboards for school effectiveness research (Fitz-Gibbon & Kochan, 2000) or actually represent the first examples of school effectiveness research (Reynolds, Teddlie, Creemers, Scheerens, & Townsend, 2000). What is clear, however, is that the advancement of school effectiveness research as a field had its origins with these seminal studies and their bleak implications that schools made very little difference in the lives of children.

The studies by Coleman et al. (1966) and Jencks et al.(1972) were characteristic of the early studies of school effectiveness in that they employed an input-output model of effectiveness (Teddlie & Reynolds, 2000). This model assumes that student achievement could be linked directly to the amount of physical resources within a school. Many researchers found this approach overly simplistic, however. As a consequence, a number of school effectiveness researchers of the 1970s and 1980s began to conduct case study research in poor and urban settings placing most of their emphasis on exploring the processes involved in effective schooling (see Bidwell & Kasarda, 1980 for a review of the case study research).

The goal of the case study researchers, with their intense focus on processes, was to dispel the beliefs that schools could have little or no impact on student achievement (Reynolds et al., 2000). This new found focus on processes was likely influenced by the dominant set of program evaluation models at the time known as the input-process-output (IPO) models (Madaus, Haney, & Kreitzer, 1992). In time, an extension of the traditional IPO model, known as the context-input-process-product (CIPP) model (Stufflebeam, 1983), would emerge and remain the state-of-the-art model for studying school effectiveness even until the present day (Reynolds et al., 2000). The advantage to the CIPP model is that it takes the educational context into account when investigating school effectiveness. As a result of this increased sensitivity to the context of school effects, subsequent studies of school effectiveness tended to be much more explicit about the external validity of their findings.

Although the field of school effectiveness research has only been an active academic field for a little more than 30 years, researchers have had to confront and overcome several key issues as the field has developed. Four of the major issues that have been addressed are: (i) model mis-specification, (ii) the unit of analysis problem, (iii) external validity, and (iv) longitudinal v. cross-sectional research designs. Each of these issues will now be discussed in further detail.

Model Mis-specification

In any statistical analysis, including all of the relevant explanatory variables as well as the proper outcome variable(s) is of prime importance in achieving results that can be validated. Reynolds et al. (2000) note that, "…a major criticism of the early school effects literature was that school/classroom processes were not adequately measured, and that this contributed to school level variance being attributed to family background variables rather than educational processes." (pp. 8-9). In addition to the need to probe deeper into educational process variables, Madaus, Kellaghan, Rakow, & King (1979) explicated the importance of using curriculum specific tests as a measure of student achievement. Their research suggested that "examinations geared to the curricula of schools are more sensitive indicators of school performance than are conventional norm-referenced tests." (Madaus et al., 1979, p. 223). Consequently, school factors were able to explain more of the variance in achievement on tests that were specific to the schools curriculum than if the test was a generalized measure of achievement.

The problem of model mis-specification was particularly evident in early studies of school effectiveness such as the Coleman report. While the Coleman Report was pioneering for its time in that it measured educational inputs as well outputs, the study did not collect any data on educational process measures. In retrospect, the Coleman Report was of limited use for assessing school effectiveness because it used only a brief and generalized measure of achievement employing a series of multiple-choice items that were not aligned to any particular set of curricular goals. The problem of model mis-specification was rampant in early school effectiveness research and addressing this issue required the field to move toward studies incorporating inputs, processes, and outputs.

A problem related to model mis-specification is the problem of multicollinearity among the explanatory variables. Multicollinearity concerns the extent to which explanatory variables are highly correlated with one another. If two explanatory variables are highly correlated with one another and both variables explain a great deal of variability in the outcome measure, it is difficult to draw any solid conclusions about the importance of one variable over the other in explaining student outcomes. This issue has been particularly troublesome for school effectiveness research due to the high degree of association between student background factors (e.g., SES) and school process variables. Perhaps the classic study illustrating the problem of multicollinearity was conducted by Brookover et al. (1978). This study found that when school climate factors were entered into a regression equation first, they accounted for 73 percent of the school level variance in student achievement, however, school climate factors only accounted for 4 percent of the variance when entered into the regression equation last. While most studies have allayed the problem of model mis-specification by introducing IPO models, multicollinearity is perhaps the most incessant and persistent issue in the field and will only be resolved through the further development of theoretical models that explore the causal mechanisms underlying school effectiveness.

The Unit of Analysis Problem

The second major issue to confront school effectiveness researchers is known as the unit of analysis problem. In 1976, Cronbach published a paper warning of the statistical problems associated with using aggregated data to answer questions targeted at lower units of analysis. This article was soon followed by others (see Burstein, 1980; Haney, 1980) also warning of the potentially misleading effects of ignoring the unit of analysis in school effectiveness research. Reynolds (2000) notes that the issue of aggregation bias, known as the Robinson effect, considerably slowed the development of the field of school effectiveness research in the United States. The field went through a period of arrested development in the 1980s as researchers were aware of the Robinson effect but had no recourse against it. It was only with the advent of multi-level modeling techniques that the field was able to continue its evolution.

Many of the early developments in multi-level modeling occurred in the UK and as a result, a large proportion of school effectiveness research conducted during the 1980s was carried out by British researchers (see Creemers & Scheerens, 1989; Mortimore, Sammons, Stoll, Lewis, & Ecob 1988a, 1988b; Willms, 1987a). By the late 1980s and early 1990s, HLM techniques were being embraced in the United States, often employed in analyses of the National Assessment of Educational Progress (NAEP) and the National Educational Longitudinal Study (NELS) (see Arnold, Kaufman, & Sedlacek, 1992; Arnold & Sedlacek, 1995; Bryk & Raudenbush, 1989; Mullis et al., 1994). Even today, the use of hierarchical modeling techniques represents the state of the art for analyzing data on school effectiveness internationally (see Martin et al., 2000). HLM is now widely accepted as an effective solution to the unit of analysis problem.

External Validity

During the 1970s and 1980s, the heavy emphasis on case study research often led investigators in the field to neglect the issue of external validity. Much of the case study research was effective in convincing policy makers that schools could indeed have a larger impact on student achievement than previously thought. Indeed this research was so successful that overzealous researchers, administrators, and policy-makers sometimes rushed into crafting general policies based upon the findings of the case study research. Policy-makers soon found out the value of external validity in research studies when the findings from these case studies (conducted mostly in urban and impoverished contexts) did not translate well in other contexts. Consequently, the 1980s saw the development of models involving an increased sensitivity to context, resulting in theoretical models that accounted for context, as well as inputs, processes, and products. Many researchers have observed that the generalizability of findings into other contexts is currently one of the most important areas in the field of school effectiveness research and call for further work on this important topic to be carried out (Reynolds, 2000; Teddlie, Reynolds, & Pol, 2000).

Longitudinal v. Cross-sectional Research Designs

Almost all research that is conducted in an educational setting is necessarily quasi-experimental (see Cook & Campbell, 1979). It is very rare in education to see a research design that employs the random assignment of students into treatment and control groups because educational systems are simply not structured that way. Instead, most studies draw students from intact classrooms within a school system. The students in these classroom are not necessarily a random sample of students in the district, state, or nation. Even if random selection could be attained, however, most researchers would find it unethical to perform experiments with school curricula by assigning the experimental curriculum to some and withholding it from others. Because of this lack of experimental control in educational research, most of what is known is based upon correlational research. Although it is tempting at times to draw causal inferences from this research, it must always be recognized that an observed association can be caused by any number of mediating variables and that the causal arrow could point in either direction in many cases.

In designing studies that explore school effectiveness, a common and often powerful approach is to employ a design that uses cross-sectional data. In this design, measurements are made on a particular group of people that are ideally selected through random sampling procedures. This group of people is taken to represent the entire population under study at a particular point in time. Gray (1989) postulates that there is a greater willingness among American researchers to employ essentially cross-sectional strategies; noting in particular the Equality of Educational Opportunity Survey (Coleman et al., 1966) and many of the analyses of the High School and Beyond data (Coleman, Hoffer, & Kilgore, 1982). The major benefit to cross-sectional studies is that they typically have high response rates and allow for generalizability of the results. Much of the research on school effectiveness has been drawn from studies with cross-sectional designs (see for example Arnold & Sedlacek, 1995; Martin, et al., 2000; Mullis et al., 1994). Cross-sectional data provide a snapshot of achievement at one point in time and can be very powerful when proper sampling procedures are employed. The data used in the present dissertation study are drawn from a cross-sectional sample of students in grade four mathematics classes in 26 countries.

A second approach to studying school effectiveness has been to examine results across time. Some researchers have argued that studies using longitudinal design for studying school effects yield more valid results than studies using a cross-sectional design (see Sammons, 1999; Teddlie & Reynolds, 2000). The major benefit to a longitudinal design is that it allows the researcher to reduce the error variance that arises from dealing with two or more different samples. Consequently, because repeated measurements are made upon the same unit of analysis over time, causal inferences may be more tenable based on the results of a longitudinal study. Yet some researchers have pointed out that longitudinal designs have their own set of problems that may limit their utility (see Sammons, 1999). For example, in the volatile world of educational policy, policy shifts happen frequently. Longitudinal designs and causal inference are powerful only to the extent that policies are held constant over time. Furthermore, while many longitudinal research designs are based upon representative sampling for their first measurement, as time goes on and attrition occurs, the representativeness of the sample is no longer assured, leading to lower levels of external validity. Both cross-sectional and longitudinal research designs have strengths and limitations. Consequently, the field of school effectiveness research will continue to benefit from the presence of research studies using each of these designs.

UNDERSTANDING THE CONTEXT AND CONTENT OF TIMSS

The International Association for the Evaluation of Educational Achievement (IEA) was established in 1961 with the purpose of promoting the development of comparative educational studies in an international context. Since then, the IEA has sponsored a series of studies that have provided policy makers, educators, researchers, and practitioners with information about educational achievement and learning contexts. The most recent of these studies is known as the Third International Mathematics and Science Study (TIMSS) (Mullis et al., 1997). The first international study of comparative achievement sponsored by the IEA, known as the First International Mathematics Study (FIMS), was conducted in 1964. This study was soon followed by the First International Science Study (FISS) conducted from 1970-1971. Building upon the knowledge gained from FIMSS and FISS, the Second International Mathematics Study (SIMS) was conducted from 1980-1982, with the Second International Science Study (SISS) being conducted from 1983-1984. For TIMSS, however, it was decided that the subjects of mathematics and science were related and that a better approach would be to conduct the studies together as an integrated effort (Mullis et al., 1997).

Students representing five different grade levels from three different populations were tested on TIMSS (see Martin & Kelly, 1996 for more information about the grades tested in specific countries). Population 1 consisted of all students enrolled in the two adjacent grades that contained the largest proportion of nine year-olds at the time of testing. Across most countries, this corresponded to grades three and four. Population 2 consisted of all students enrolled in the two adjacent grades that contain the largest proportion of 13 year-olds at the time of testing. Across most countries this corresponded to grades seven and eight. Finally, Population 3 consisted of students enrolled in their final year of secondary education. Although the amount of schooling received by students in this population varied within countries, in a number of countries, including the U.S., this corresponded to grade 12. Participation in the study required all countries to administer the survey to students at Population 2, while participation was voluntary at the other two populations. This dissertation study explores the results from those countries who participated at Population 1. More specifically, the focus of the study is on those students who were in the upper grade of Population 1, corresponding to the fourth grade in most countries.

TIMSS was based on a set of underlying curriculum frameworks for both mathematics and science. These frameworks were developed by groups of mathematics and science educators, with input from the TIMSS National Research Coordinators (NRCs). Working within the curriculum frameworks, test specifications were developed for both mathematics and science that included items representing a wide range of topics and eliciting a range of skills from the students. Approximately one-fourth of the items on the tests were in the free-response format requiring students to generate their own answers while the remaining questions used a multiple-choice format. The tests were developed through an international consensus involving input from experts in mathematics, science, and educational measurement. Every effort was made to help ensure that the tests represented the curricula of the participating countries. In the end, the final forms of the tests were endorsed by the NRCs of the participating countries (Mullis et al., 1997).

The mathematics test at Population 1 (third and fourth grade in most countries) covered six content areas and four performance expectation. Table 2.1 presents the distribution of mathematics items by content reporting category and performance expectations.

Table 2.1 Distribution of Mathematics Items by Content Reporting Category and Performance Expectation – Population 1*

Content Category	Percentage of items	Number of items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items
Whole Numbers	25%	25	19	5	1
Fractions and Proportionality	21%	21	15	2	4
Measurement, Estimation, and Number Sense	20%	20	16	3	1
Data Representation, Analysis, and Probability	12%	12	8	2	2
Geometry	14%	14	12	2	0
Patterns, Relations, and Functions	10%	10	9	1	0
Total**	102%	102	79	15	8

Performance Expectation	Percentage of Items	Number of Items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items
Knowing	41%	42	35	7	0
Performing Routine Procedures	16%	16	13	3	0
Using Complex Procedures	24%	24	21	2	1
Solving Problems	20%	20	10	3	7
Total**	101%	102	79	15	8

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-1995.

* Third and fourth grade in most countries

** Because results are rounded to the nearest whole number some totals may appear inconsistent

Table 2.2 presents the distribution of science items by content reporting category and performance expectation. The science test covered four content areas and five performance expectations.

Table 2.2 Distribution of Science Items by Content Reporting Category and Performance Expectation – Population 1*

Content Category	Percentage of items	Number of items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items
Earth Science	18%	17	13	2	2
Life Science	42%	41	33	5	3
Physical Science	31%	30	23	4	3
Environmental Issues and the Nature of Science	9%	9	5	2	2
Total**	100%	97	74	13	10

Performance Expectation	Percentage of Items	Number of Items	Number of Multiple-Choice Items	Number of Short-Answer Items	Number of Extended-Response Items
Understanding Simple Information	45%	44	42	1	1
Understanding Complex Information	31%	30	21	5	4
Theorizing, Analyzing, and Solving Problems	14%	14	3	6	5
Using Tools, Routine Procedures, and Science Processes	6%	6	5	1	0
Investigating the Natural World	3%	3	3	0	0
Total**	99%	97	74	13	10

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-1995.
* Third and fourth grade in most countries
** Because results are rounded to the nearest whole number some totals may appear inconsistent

In addition to collecting information on student achievement, considerable efforts were made to collect background information that could help to put student achievement into the appropriate context. Toward this end, students, their teachers, and school administrators were all given background questionnaires to complete.

The Population 1 student questionnaire was composed of 23 questions, many of which had multiple parts. These questions asked students about basic demographic information, the kinds of activities they enjoy, their attitudes toward and beliefs about mathematics and science, the kinds of activities that typically happen in their schools (e.g., something of mine was stolen in the past month), and the kinds of instructional strategies that are employed in their mathematics and science classrooms (e.g., how often do you copy notes from the board).

The Population 1 teacher questionnaire was composed of 24 general questions, followed by a series of questions specifically related to the subject area they taught (i.e., either mathematics or science). In the upper grade of Population 1, it was often the case that the students had the same teacher for both mathematics and science. In that case, the teacher would fill out questions pertaining to both subject areas. The general questions asked about basic demographic information such as age, gender, and years of teaching experience. In addition, the general questions asked about the kinds of school-related activities teachers engage in (e.g., grading, meetings), their beliefs about and attitudes towards mathematics, the factors that limit their teaching ability, their familiarity with and reliance on curriculum frameworks, and their classroom planning strategies. The subject-specific portions of the survey requested detailed information on when various subjects were or would be taught, the flow of the lesson plan, and many detailed questions about classroom instructional practices (e.g., how often they ask students to explain the reasoning behind an idea, how often do they ask students to work in groups, etc.).

Finally, the Population 1 school background questionnaire was composed of 27 questions, many of which had multiple parts. Questions on this survey were targeted at gaining a clear picture of the context of the school (e.g., type of community school is locate in, number of various staff members such as teachers and principals), assessing the school climate (e.g., how long pupils stay with the same teacher, school policies on teacher collaboration), gauging the political dynamic of the school (e.g., who has the primary responsibility for hiring teachers, how much influence parents have upon the curriculum), and understanding the internal dynamic of the school (e.g., how frequently various behavior problems were dealt with).

As a result of the wealth of policy-relevant data that was collected on TIMSS, many researchers have conducted secondary analyses on the TIMSS data. Interestingly, however, the majority of these analyses have focused on the data collected at Population 2, and in particular the students in the upper grade of Population 2, which corresponds to the eighth grade in most countries (see Boss, Kuiper, & Plomp, 1999; Jones, 1998; Lee, 1998; Martin et al., 2000; Tarr, Mittag, Uekawa, & Lennex, 1999; Wilson & Blank, 1999). By contrast, there have been fewer studies targeted at understanding the educational practices that have an impact on students at Population 1 (Frase, 1997; Tamir, & Zuzovsky, 1999). The present study will fill a gap that presently exists in the literature by exploring variables related to achievement at the upper grade of Population 1, which corresponds to the fourth grade in most countries. The findings of this dissertation will be especially useful in light of the fact that TIMSS 2003 is presently being planned and will once again include an investigation of student achievement at grade 4. Consequently, the results of this study will serve as a baseline against which comparisons may be made in the future.

HOW CAN TIMSS CONTRIBUTE TO SCHOOL EFFECTIVENESS RESEARCH?

The TIMSS database provides access to data from the largest study of comparative educational achievement to date. In addition, the design of TIMSS is advantageous in that it overcomes many of the problems of early school effectiveness research discussed in the previous section. Learning from prior research, TIMSS attempted to combat the problem of model mis-specification of the outcome variables of mathematics and science achievement by conducting a thorough curriculum analysis in each subject for each country. The tests were designed based upon an international consensus of mathematics and science educators in each participating country.

In addition to the standard reporting of student achievement, TIMSS took on an extra analysis known as the test-curriculum matching analysis (TCMA). The TCMA allowed the National Research Coordinator (NRC) for each country to specify which items were not appropriate to their curriculum and to essentially compare themselves with other countries on the basis of what each country NRC reported as appropriate items (see Beaton & Gonzalez, 1997 for further details). The results of the TCMA analysis did not differ markedly from the regularly reported results. Consequently, a strong case can be made that the results from the mathematics and science assessments do adequately reflect the content domains covered by the curricula of the participating countries.

TIMSS also took special efforts to collect detailed information on educational processes by administering questionnaires to the students taking the exam, their teachers, and their school administrators. Across all populations, data were collected on over 1,500 variables, making the data set a rich source for exploring the effects of educational processes.

Another benefit to TIMSS is that it was designed to allow researchers to effectively deal with the unit of analysis problem. The sampling design allows schools, classrooms, and students to all be potential units of analysis (Foy et al., 1996). In order to accomplish this, each had to be considered as sampling units in the sample design in order to obtain the maximum sampling precision for each unit. Consequently, each country participating in TIMSS was required to include at least 150 schools per target population with an effective sample size of at least 400 students for the main criterion variables (Foy et al., 1996). "A sample of 150 schools yields 95% confidence limits for school- and classroom-level mean estimates that are precise to within +/- 16% of their standard deviation." (Foy et al., 1996, p. 4-7). Additionally, an effective sample size of 400 students results in 95% confidence limits that are precise to within m +/- 0.1s (where m is the means estimate and s is the estimated standard deviation for students) (Foy et al., 1996). For example, at the upper grade of Population 1, New Zealand typified the sampling design selecting 149 schools with 2,421 students participating.

Finally, the TIMSS database allows for an unprecedented opportunity to explore the extent to which an explanatory model is stable across contexts. Twenty-six countries met the guidelines for participation at Population 1 (third and fourth grade in most countries), allowing them to be included in comparative analyses. Although TIMSS is a cross-sectional study, the power of the design is that students from each of the participating countries may truly be described as a representative sample of students in the target population. The power of the sampling design and the international nature of the study give findings from the TIMSS database the potential to possess a high degree of external validity, provided the data are analyzed appropriately. Let us now turn to the issue of choosing the proper analytic technique for this study.

WHY USE HIERARCHICAL LINEAR MODELING (HLM) FOR ANALYSIS?

The most advanced and most recently developed method for analyzing school effectiveness data is through the use of multi-level modeling (Bryk & Raudenbush, 1992; Goldstein, 1995; Kreft & DeLeeuw, 1998; Pedhazur, 1997). Multi-level modeling is an extension of regression analysis and was developed in order to help combat the unit of analysis problem that has traditionally plagued school effectiveness research. In short, the unit of analysis problem arises from the fact that researchers often wish to examine the effects of data drawn from two or more levels of a hierarchy. "For example, when the object of analysis is the assessment of the importance of school-level variables, but we have individual (pupil-level) outcomes, should these be aggregated to the school level for the analysis, or should we analyse the individual pupil outcomes? If the latter, how should the school structure be represented in the model?" (Aitkin & Longford, 1986, p. 1). Aitkin & Longford (1986) go on to note that aggregating student level data up to the school level still allows variance to be explained between schools. Variables measured at the school level, however, cannot be used to explain variance at the student level because variables measured at the school level are constant within the school (i.e., all students have the same value on that variable). Traditional regression models such as Ordinary Least Squares (OLS) can only handle measurements made at one level of a hierarchy so the researcher must decide whether to analyze data at the school level or at the student level.

The major benefit of multi-level modeling over OLS regression is that multi-level models allow researchers to simultaneously handle measurements made at various levels of a hierarchy and the variance component can then be broken down by hierarchical level (Kreft & DeLeeuw, 1998). Plewis (1986) notes that, "If there is indeed variability between schools in the regression coefficients relating intake to outcome, our ideas about school effectiveness will need to be modified. No longer will it be possible to talk about one school being uniformly more effective than another, as it is with parallel regression lines. Indeed, we might find that school A is more effective than school B for children with low attainment scores at intake, but less effective for children with high intake scores." (p. 27). In concurrence, Raudenbush & Bryk (1986) have declared that, "By its very nature, school-effects research requires exploration of hierarchical data – a search for statistical associations between school factors on the one hand and student-level variables on the other." (p. 13).

HLM is an especially useful technique for the analysis of data that has come from a cluster sample design rather than simple random sampling. This is a particularly important issue in TIMSS given that the students in the study were sampled using a two-stage stratified cluster sampling approach (Martin & Kelly, 1996). Although cluster sampling techniques are more practical than simple random sampling, using OLS analyses on clustered data have led to analyses that have been problematic in the past (studies applying OLS to multi-level problems include Coleman et al., 1982; Page & Keith, 1981; McPartland & McDill, 1982; Alexander & Pallas, 1983). This is because students tend to be more homogeneous within schools than they are in a true random sample (Raudenbush & Bryk, 1986; Kreft & DeLeeuw, 1998; Pedhazur, 1997). This homogeneity is a violation of the assumption of independent error terms required for OLS regression and leads to biased parameter estimates for the error term, which results in a higher rate of Type I error than would be observed with data from a simple random sample (Aitkin & Longford, 1986; Bryk & Raudenbush, 1992; Haney, 1980; Kreft & DeLeeuw, 1998; Raudenbush & Bryk, 1986). This problem, introduced by the correlations among observations resulting from the multi-level structure, has remained intractable until recently (Aitkin & Longford, 1986).

Even questions that do not appear on their surface to require a multi-level model in order to be answered may, in fact, be well suited to such a model. For example, take the question “Is the number of hours of homework completed a good explanatory variable for high math test scores?”. At face value it is not a multi-level question, but there are good reasons for employing multi-level analysis in this case. First, intra-class correlations may be present since students are nested within schools, and sampled from within schools. Second, we expect school effects to be present (i.e., the relationship between homework and math score is not consistent across all schools) (Kreft & De Leeuw, 1998). These complex questions cannot be answered using traditional regression analyses. Consequently, most current studies of school effectiveness choose to employ the more powerful multi-level techniques.

PARTITIONING THE VARIANCE IN ACHIEVEMENT

Many recent studies of school effectiveness have enjoyed the benefits of using HLM in their analyses. Much of the mutli-level research in this area has provided strong evidence of the existence of differences between schools in their overall effectiveness in promoting pupils' academic attainments (see Mortimore et al., 1988a, 1988b; Bondi, 1991; Willms, 1986). Most recent studies of school effectiveness have been conducted using a two-level model where the variance in achievement can be attributed to students at one level and to differences between schools at a second level. Researchers interested in exploring school effectiveness are more interested in the amount of variance that lies between schools than the variance that lies between students. It is frequently observed, however, that the within-school variance is higher than the between-school variance, indicating that individual students differ more from each other than schools do (Kreft & De Leeuw, 1998).

Perhaps the least optimistic results were found in an analysis of school effectiveness conducted by Aitkin & Longford (1986). In their analysis, the between school variance for the 18 schools in the sample amounted to seven percent after allowing for Verbal Reasoning (VRQ). When the two selective schools in the sample were removed from the analyses, the overall variance between schools fell to only two percent for the more homogeneous sub-sample of 16 schools. As Aitkin & Longford point out, “This point is worth emphasizing: in the search for relevant explanatory variables at the school level, it must be recognized that such variables can explain at most 2 per cent of the total variance of individual pupil outcomes not explained by intake score VRQ.” (1986, p. 15).

Similarly, Willms (1987a) reanalyzed data from one Scottish administrative division in which both prior attainment and background data were available. The data he reports for the 21 schools concerned suggest that the between-school component amounted to around three percent, when two types of control variables (prior attainment and background) were applied (Gray, 1989). Willms (1987b) conducted another study of the Scottish educational system using data for the whole of Scotland. The results of that study were slightly more optimistic in that the variance between pupils was around 90 percent, depending on the model. Variance between schools was on the order of 10 percent, reducing to around six percent in the model that fit school mean SES as a fixed effect.

Gray, Jesson, & Jones (1986) conducted a study of school effectiveness similar to those done by Willms, using data from another school system (LEA 3), and a fairly similar story emerged. In the local education authority they examined, they were able to obtain complete data on all the pupils (n > 4,000) in the year cohort rather than just a sample. Nonetheless, they found that between-school variation amounted to just five percent.

Sammons (1999) cites an AMA study in which 19.1 per cent of the total variation in students' GCSE performance scores was found to lie between schools. She also reports the results of her own study that included 58,628 cases covering 418 secondary schools. In that study, "Almost exactly a quarter of the total variation in students' total GCSE performance scores was found to lie between schools (was attributable to differences between schools rather than to differences between individuals)." (p. 42). Although the percent of variance that lies between schools is rather small, Kilgore & Pendleton (1986) postulate that when we observe very low variability in achievement, it may be attributable to national policies driving the curriculum that is taught in the schools.

In the rare instances when three-level models have been employed in school effectiveness research, the third level tends to be repeated measurements over time (see Bryk & Raudenbush, 1989). Two notable exceptions are found in the research by Tymms (1993) and the research by Hill & Rowe (1996). The Tymms study was innovative in that it examined students within classrooms within departments at the secondary level of education in the UK. These results were then compared to the results of a two-level model of the same data with students and schools representing the two levels. Tymms (1993) demonstrated that while the variance in achievement attributable to schools in a two-level model was only seven percent, the variance in achievement attributable to class ranged from 9 to 25 percent and by department ranged from 8 to 24 percent when using a three-level model. Hill & Rowe (1996) observed a similar trend in their study in that 18 percent of the variance in student achievement in their study was accounted for by between-school differences when a two-level model was employed, however, that number dropped to between five and six percent when a three-level model was run. What these studies suggest is that schools are having an effect upon student achievement, but that effect is greatest at the level of the classroom.

THEORETICAL MODELS FOR EXPLORING SCHOOL EFFECTIVENESS

This dissertation study will present a two-level model of school effectiveness in mathematics at the fourth grade and a two-level model of school effectiveness in science at the fourth grade. Although the amount of variance that lies between schools appears to be much smaller than the variance between individuals, it is still instructive from a school effectiveness perspective to attempt to explain as much of the between school variance as possible in order to implement more effective educational policies. Toward that end, researchers have worked hard to develop more precise techniques for analyzing correlational data that come from hierarchical structures. Several studies of school effectiveness have been conducted recently that use such multi-level modeling techniques to increase the precision of the parameter estimates for predictor variables (Arnold & Sedlacek, 1995; Burstein, 1980; Martin et al., 2000; Mullis et al., 1994). These studies, and many others, have been successful in identifying a number of policy-relevant variables associated with student achievement.

Although the specific explanatory models tested vary from study to study, each model is constructed using variables from five major categories. These categories include sets of variables that are classified as:

1. Dependent variables
2. Student background variables
3. Student-level explanatory variables
4. Teacher or Classroom-level explanatory variables
5. School-level explanatory variables

Dependent Variables

In school effectiveness research, the dependent variable has the special role of operationally defining the term “effective”. Some dependent variables that have been used to explore and define school effectiveness include measures of:

- Student achievement (see for example Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Bennett, 1976; Martin et al., 2000; Mullis et al., 1994; Mortimore et al., 1988a, 1988b; Rutter, Maughan, Mortimore, & Ouston, 1979; Steedman, 1983)
- Student attitudes (Mortimore et al., 1988a, 1988b)
- Student attendance (Rutter et al., 1979; Mortimore et al., 1988a, 1988b; Smith, Tomlinson, Tomes, & Bannerjea, 1984)
- Discipline problems (Rutter et al., 1979; Mortimore et al., 1988a, 1988b; Smith et al., 1984)

While it is advisable to use multiple measures in defining “effectiveness” (especially when the classification has high-stakes attached), most studies of school effectiveness use a measure of achievement as the main dependent variable in their study.

The reason for this is generally because of its capacity for reliable measurement. Within the construct of achievement, however, there are several content areas that have been the focus of school effectiveness research including:

- Mathematics achievement (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Bennett, 1976; Steedman, 1983; Martin et al., 2000; Mullis et al., 1994)
- Science achievement (Arnold et al., 1992; Martin, et al., 2000)
- English achievement (Bennett, 1976; Steedman, 1983);
- Reading achievement (Bennett, 1976);

For the purposes of the present dissertation study, two dependent variables will be explored. These dependent variables are mathematics achievement and science achievement on TIMSS. Using two dependent variables will allow the explanatory power of a single theoretical model of school effectiveness to be compared across content domains.

Student Background Variables

Nearly all studies of school effectiveness that have been conducted use one or more measures of student background as control variables. Student background variables are chosen for the purposes of statistical control for two reasons. First, these variables are not typically affected by changes to educational policy; and second because it has been established in the literature that these variables are strongly related to educational achievement. As a consequence, it is important to recognize and control for student background factors in order to ensure that our model has been correctly specified.

The most commonly employed student background variable is socioeconomic status (SES). Interestingly, however, the construct of SES, is not measured in exactly the same way across studies. The SES variable is a composite variable typically calculated using some combination of measures such as:

- Parent's level of education (Martin et al., 2000; Mullis et al., 1994; Scheerens, Vermeulen, & Pelgrum, 1989)
- Number of books in the home (Arnold & Sedlacek, 1995; Mullis et al., 1994; Martin et al., 2000)
- Number of family members living together in the household (Arnold & Sedlacek, 1995; Davie, Butler, and Goldstein, 1972; Douglas, 1964; Essen & Wedge, 1982; Martin et al., 2000; Mortimore & Blackstone, 1982; Mullis et al., 1994; Rutter & Madge, 1976; Sammons, 1999)
- Possessions in the home (Arnold & Sedlacek, 1995; Martin et al., 2000)
- Parental occupation (Lockheed & Komenan, 1989; Scheerens et al., 1989; Sammons, 1999)

Out of all the variables included in the assortment of school effectiveness research literature, SES tends to explain the largest share of the variance in most studies. In addition, other measures of prior achievement are also extremely powerful for explain the variability in achievement. Gray (1989) notes that controls for differences in pupils' backgrounds appear to explain between 20-30 percent of the variation in pupils' outcome. Furthermore, studies that have some measure of prior achievement typically account for upwards of 50 percent of the variation in pupil outcome (Gray, 1989).

Sammons (1999) recounts an AMA study in which 19.1% of the total variation in students' GCSE performance scores was found to lie between schools. Of that 19.1%, individual-level prior attainment and background data accounted for 64.5% of the between school-level variation. In another study by Sammons (1999), a variable indicating school type was included as a surrogate for prior achievement. When this variable was included, the prior achievement explained 75.6% of the between-school variation.

In their study of school effectiveness, Aitkin & Longford (1986) constructed a block of variables representing student intake variables (SES, ability, sex, ethnicity). They also included a block of variables representing school intake variables (financial and other resources, class size, pupil/staff ratio, teachers' attitudes, curriculum, school size, and proportion of females in the school). Their study found that 27% of the variance in achievement was between schools (and therefore was explainable), and that simply adding a measure of prior achievement to the model explained 91% of the between-school variance.

Given the explanatory power of measures of prior achievement, many authors have argued that the availability of baseline prior achievement data is crucial for the purpose of comparing school effectiveness (see Cuttance, 1986; Mortimore et al, 1988b; Jesson & Gray, 1991; McPherson, 1992; and Goldstein et al., 1992 for a further discussion of the issue of valid comparison of schools). Yet Sammons, Mortimore, and Thomas (1993) have noted that, "Where no or only crude measures of prior attainment are available, the estimates of the impact of background factors upon pupils' later attainment are likely to remain large." (p. 389).

In addition to including controls for SES and prior achievement, Sammons, Nuttall, & Cuttance (1993) have noted that measures of sex and ethnicity have been shown to be important predictors of attainment at both the primary and secondary levels. These measures are also related to prior achievement and, in the absence of measures of prior achievement, assume much greater importance as control measures in the analysis of schools' educational outcomes.

Other student background variables that researchers have considered important to control include the student's:

- Gender (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Lockheed & Komenan, 1989; Sammons, 1999)
- Ethnicity (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Ma & Kishor, 1997; Sammons, 1999)
- Age (Aitkin & Longford, 1986; Lockheed & Komenan, 1989)
- Prior achievement (Cuttance, 1986; Mortimore et al., 1988b; Goldstein et al., 1992)

Once student background has been statistically adjusted, it is then possible to gain a more accurate perception of the importance of variables that are affected by educational policy. Variables that are affected by policy changes may be found at the student level, the classroom and teacher level, and at the school level.

Student-level Explanatory Variables

Research has demonstrated that the kinds of behaviors in which students engage and the beliefs that they hold may have an influence upon their subsequent achievement. Consequently, school effectiveness studies have explored a wide variety of student-level explanatory variables. Some of the more typical of these include:

- Attitudes toward subject (Aiken, 1970; Arnold & Sedlacek, 1995; Maker, 1982; Martin et al., 2000)
- Time spent on homework (Arnold & Sedlacek, 1995; Creemers & Osinga, 1995; Martin et al., 2000; Mullis et al., 1994; Scheerens et al., 1989)
- Student Locus of Control (Fenn & Iwanicki, 1983; Reynolds et al., 2000; Sterbin & Rakow, 1996)
- Achievement orientation (Creemers & Osinga, 1995; Lockheed & Komenan, 1989; Scheerens et al., 1989)
- Perceived parental support (Lockheed & Komenan, 1989)

Perhaps the strongest and most consistent association with achievement reported in the literature relates to students locus of control. The construct of locus of control was developed by Rotter (1966) in the context of social learning theory. Rotter defined the locus of control such that when students perceive that the results following their actions are related to some external force such as luck, fate, or powerful others, then students are said to have an external locus of control. To the extent that students see a direct connection between outcomes and their own actions (such as hard work) or relatively personal characteristics (such as ability), these students are said to have an internal locus of control (Benham, 1995). The idea of locus of control was expanded and incorporated into the attributional theory of motivation developed by Weiner (1972). Their attributional theory describes ability as a stable attribution of internal locus of control, effort as an unstable attribution of internal locus of control, task difficulty as a stable attribution of external locus of control, and luck as an unstable attribution of external locus of control.

Sterbin & Rakow (1996) investigated the association between standardized test scores on the National Educational Longitudinal Study (NELS) 1994 database and a number of affective variables, including student locus of control and student self-esteem. Their results showed that internal locus of control was highly statistically significantly correlated with student achievement ($r = 0.29$). Additionally, Fenn & Iwanicki (1983) conducted an investigation of the relationship between student affective characteristics and student achievement within more and less effective schools. After accounting for differences in student background, the findings revealed that, on all measures, the affective disposition of students in more effective schools was consistently more positive than that of students in less effective schools (Fenn & Iwanicki, 1983).

Many of the other variables included in this block are theoretically interesting, however, their relationship to achievement has been shown to be more precarious. For example, Ma & Kishor (1997) in their meta-analysis of the relationship between attitude toward mathematics (ATM) and achievement in mathematics (AIM) concluded that, "The research literature has failed to provide consistent findings regarding the relationship between ATM and AIM." (p. 27). Even when research has found the relationship to be statistically significant, Aiken (1970) points out that, "…the correlations between attitude and achievement in elementary school, though statistically significant in certain instances, are typically not very large." (p. 559). Nevertheless, it is important to examine the relationship between cognitive variables and affective variables because, as Maker (1982) points out, "It is impossible to separate the cognitive from the affective domains in any activity…there is a cognitive component to every affective objective and an affective component to every cognitive objective." (p. 30-31).

Similarly, the relationship between time spent on homework and achievement is also not yet fully understood. It may be the case that students who are spending a lot of time on their homework are the most conscientious students, and therefore will also be the highest achieving. On the other hand, there is another argument to be made that the students who spend the most time on their homework are those who are struggling with the material, and will subsequently be the lowest achievers. Recent research by Martin et al. (2000) has shown that the amount of homework that students do is significantly positively related to student achievement at the school level in a number of countries in eighth grade mathematics and science.

Teacher-level/Classroom-level Explanatory Variables

Another important set of policy-relevant variables that have been explored are teacher and classroom-level explanatory variables. Some of the more typical of these include:

- Teacher education (Fuller, 1987; Walberg, 1991)
- Teacher specialization (Martin et al., 2000)
- Teacher experience (Creemers & Osinga, 1995; Fuller, 1987; Lockheed & Komenan, 1989; Martin et al., 2000; Mayer, Mullens, & Moore, 2000; Reynolds et al., 2000; Scheerens et al., 1989)
- Gender of teacher (Scheerens et al., 1989)
- Teacher salary (Fuller, 1987)
- Frequency of group work (Arnold & Sedlacek, 1995)
- Frequency of classroom tests (Arnold & Sedlacek, 1995; Mullis et al., 1994)
- Frequency of computer use (Arnold & Sedlacek, 1995; Mayer et al., 2000; Wenglinsky, 1998)
- Frequency of calculator use (Arnold & Sedlacek, 1995; Mullis et al., 1994)
- Homework frequency (Fuller, 1987)
- Minutes of subject instruction per week (Lockheed & Komenan, 1989; Scheerens et al., 1989)
- Teacher expectations of students (Fuller, 1987; Reynolds et al., 2000; Scheerens et al., 1989)
- Use of commercially published teaching materials (Lockheed & Komenan, 1989; Scheerens et al., 1989)
- Use of personally produced teaching materials (Lockheed & Komenan, 1989; Scheerens et al., 1989)
- Class size (Arnold & Sedlacek, 1995; Fuller, 1987; Lockheed & Komenan, 1989; Martin et al., 2000; Mayer et al., 2000; Scheerens et al., 1989)

Murnane & Phillips (1981) found that after controlling for student and teacher SES and race/ethnicity, students taught by teachers with at least five years of experience made three to four months' more progress in reading skills during a school year than students who are taught by a first-year teacher. In addition, a recent review of educational indicators associated with school effectiveness in the United States reported that, "Substantial research suggests that school quality is enhanced when teachers have high academic skills, teach in the field in which they are trained, have more than a few years of experience, and participate in high-quality induction and professional development programs." (Mayer et al., 2000, p. i). Subsequent research, however, suggests that the benefits of teacher experience may tend to level off after about five years (Darling-Hammond, 2000).

In addition, several theories of school effectiveness have emphasized the importance of instructional practices. For example, Bierhoff, (1996) argues that the high mean scores and low variability in achievement in mathematics and science that is observed in Switzerland may be attributed to such factors as teacher's heavy reliance on teaching from nationally developed textbooks that are tied directly to the curriculum (see Reynolds, 2000). In addition, most of the lesson time in Switzerland involves interactive teaching as opposed to teacher lectures, and he argues that this too may contribute to high scores.

In many quantitative studies of school effectiveness, it is difficult to gain a full appreciation for variables such as how accepted the students feel in the classroom, and the extent to which classroom routines are well understood. Nevertheless, prior research has indicated that some factors related to specific instructional practices such as the use of technology in the classroom, the amount and frequency of homework assigned, and the frequency of group work and/or teacher lectures are associated with student achievement.

Perhaps the most consistently significant variable in this set is the use of student calculators. Interestingly, however, the use of calculators has been shown to have a statistically significant positive relationship with average achievement at grades 8 and 12 (Mullis et al., 1994), but has been shown to have a significantly negative association with average school achievement at grade 4 (Arnold & Sedlacek, 1995).

The frequency of computer use in the classroom has shown a precarious relationship with school effectiveness. Wenglinsky (1998) suggests that part of the reason for this is because the variable of interest should not simply be how often computers are used, but also how they are used. His recent analysis of NAEP data revealed that students using computers for lower-order instruction tended to score lower on tests of achievement while students using computers for higher-order thinking tended to score higher on achievement tests. For example, in the eighth grade, "Black students were less likely to be exposed to higher-order uses of computers and more likely to be exposed to lower-order uses than Whites. Similarly, poor, urban, and rural students were less likely to be exposed to higher-order uses than non-poor and suburban students." (Wenglinsky, 1998, p. 3). These findings suggest that when computers use is negatively associated with school effectiveness, it is likely due to the way that the technology is being used. Yet even this relationship is debated. Mayer et al. (2000) claim that, "Numerous studies conducted in the elementary and secondary grades have concluded that student learning is enhanced by computers when the computer is used to teach discrete skills in the style referred to as 'drill and practice.'" (p. 27).

Russell (2000) suggests that perhaps part of the reason that the relationship between computer use and achievement is so capricious is because the number and type of items used to test students' achievement in certain areas may be insufficient for assessing the impact of computer use on those skills. Nevertheless, an examination of the nature of the relationship between the frequency of computer use and school effectiveness may serve to highlight areas for further exploration within countries.

School-level Explanatory Variables

Finally, while not all school effectiveness research includes classroom and teacher variables, all school effectiveness research includes some measure of school-level explanatory variables. Some of the more typical of these include:

- Financial and other school resources (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Fuller, 1987; Martin et al., 2000)
- School size (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Fuller, 1987; Lockheed & Komenan, 1989; Martin et al., 2000; Mullis et al., 1994)
- Percentage of females in school (Aitkin & Longford, 1986; Lockheed & Komenan, 1989)
- Percentage minority students in school (Aitkin & Longford, 1986; Arnold & Sedlacek, 1995)
- Percentage of students with special needs (Aitkin & Longford, 1986)
- Serious behavior problems (Martin et al., 2000; Mayer et al., 2000)
- Minor behavior problems (Arnold & Sedlacek, 1995; Martin et al., 2000; Mayer et al., 2000)
- Instructional leadership from principal (Creemers & Osinga, 1995; Reynolds et al., 2000)
- Broadly understood instructional focus (Reynolds et al., 2000)

- Safe and orderly school environment (Creemers & Osinga, 1995; Reynolds et al., 2000; Scheerens et al., 1989)
- Length of school year (Lockheed & Komenan, 1989)
- Library size and activity (Fuller, 1987; Walberg, 1991)
- Instructional media (computers, overheads) (Arnold & Sedlacek, 1995; Fuller, 1987)
- Instructional materials (Fuller, 1987)
- Student body stability (Arnold & Sedlacek, 1995)
- Instructional time in subject (Arnold & Sedlacek, 1995)
- Aggregate SES of the school (Arnold & Sedlacek, 1995; Martin et al., 2000)

School climate is a broad concept that often is intended to incorporate the total environment of the school, including factors related to parents and communities. Anderson (1982) speaks to four groups of climate factors that include (i) ecology (physical and material aspects); (ii) milieu (the composition and population of a school); (iii) social system (relationships between persons); and (iv) culture (belief systems, values). Although there are many important variables that constitute each of these four dimensions, what will be done in the present study is to separate school-level explanatory variables into two types: (i) school resources, and (ii) school climate.

School resources are analogous to Anderson's (1982) ecology dimension and deals with the physical aspects and resources available to a school such as financial resources, school size, and instructional materials. School climate, as traditionally defined in school effectiveness research tends to incorporate factors from Anderson's milieu, social system, and culture dimension. The factors related to milieu include the percentage of males and females in each school and the student body stability, while factors related to social systems include major and minor behavioral problems. Students' perceptions of peers' attitudes toward the subject area will be used as a proxy for culture.

Several of the variables listed above have been shown to be significantly related to school effectiveness. In particular, the stability of the student body was shown by Mullis et al. (1994) to be a significant predictor of school effectiveness in mathematics at the fourth grade such that effective schools were more likely to have students who had changed schools fewer times in the past two years. Other school-level predictor variables that have been shown to be significantly related to school effectiveness in mathematics at the fourth grade include the percentage of minority students in a school and the average SES of the school. A study by Arnold & Sedlacek (1995) found a highly significant positive relationship between the average SES level of the school and the average mathematics achievement of the school. In addition, Anderson (1982) notes several showing a positive association between the extent to which students in a school value the subject area and levels of academic achievement.

SUMMARY

Perhaps one of the most important contributions of this study derives from its large-scale, cross-sectional nature. Reynolds has commented that,

> The absence of cross national perspectives and relationships between school effectiveness researchers, the neglect of internationally based research on educational achievement and the lack of interaction within societies with the comparative education discipline are all features of the present state of the school effectiveness discipline that must be seen as increasingly proving intellectually costly (2000, p. 233).

The major contribution of this dissertation will be to provide an international perspective on school effectiveness in mathematics and science at the fourth grade. In addition, while many studies of school effectiveness have examined school effectiveness in mathematics, few have explored school effectiveness in science.

This chapter has provided a theoretical grounding for this dissertation study of school effectiveness. The controversy surrounding the construct of school effectiveness was presented and discussed along with an overview of the historical evolution of school effectiveness as a formal field of research. The chapter then moved on to discuss the strength of the TIMSS database and its relevance to the present study of school effectiveness. A brief introduction to the technique of hierarchical linear modeling and a discussion of its importance to the present study was presented. Finally, several key student-level, teacher-level, and school-level variables that have been employed in prior school effectiveness research were presented and discussed.

CHAPTER 3 - METHODOLOGY

OVERVIEW

The literature reviewed in chapter two established the theoretical basis for including certain variables in an explanatory model of school effectiveness. This chapter moves away from the theoretical issues in model development and toward a discussion of the empirical guidelines for variable selection that will be used in this dissertation study. These guidelines relate to response rates, observed variability, measurement error, and the efficiency of hierarchical linear modeling. After discussing why certain sets of variables will be excluded from the study, the chapter will then go on to document exactly which sets of variables will be retained for the explanatory model and why.

In addition to establishing guidelines for the inclusion of particular variables, it was also necessary to establish certain criteria for the inclusion of countries in the analysis. The second segment of this chapter will present the criteria used for the inclusion of countries in this study. These criteria relate specifically to the adequacy of the sampling procedures employed, the between-school variability in achievement observed within each country, and the extent of missing data on the explanatory variables of interest.

The third section of this chapter will serve to document some of the important nuances of working with the TIMSS database. Measuring student achievement using plausible values will be discussed along with the use of sampling weights and the procedures for estimating variability in achievement.

The final section of this chapter will discuss the analytic techniques that will be used to test the research hypotheses. In particular, the two-level hierarchical linear model will be explained along with the decisions that were made with regard to the use of fixed v. random coefficients and variable centering.

CRITICERIA FOR VARIABLE INCLUSION

Over 900 potential explanatory variables are included in the student, teacher, and school questionnaires at Population 1 on TIMSS. As such, it is necessary to impose some criteria for variable selection in creating the explanatory model used in this study. The five major criteria used to select variables for inclusion will be: (i) theoretical importance, (ii) response rates, (iii) variability, (iv) model parsimony, and (v) measurement error. Each of these criteria will now be presented and discussed.

Theoretical Importance

The most important criterion for including a variable in the explanatory model of school effectiveness developed for this study is its theoretical importance. By definition, a major aim in developing an explanatory model is to select independent variables that will explain as much variation in the dependent variable (i.e., average subject matter achievement on TIMSS) as possible. Explaining variance does little to advance our understanding of school effectiveness, however, if the variables that explain the most variance are not open to policy manipulation. For example, knowing that socioeconomic factors often explain up to 80% of the variation in school performance (Coleman et al., 1966; Jencks et al., 1972; Teddlie & Reynolds, 2000) is not an entirely useful finding from an educational policy standpoint because very little can be done to influence the SES of the students. By contrast, if it was discovered that class size is capable of explaining 10% of the variation in school performance, this would be considered an educationally significant finding because class size is open to policy manipulation. Consequently, judgments regarding the educational significance of findings must be interpreted in light of the amount of variance explained by the variable as well as the theoretical importance of the variable.

A major goal of this study is to develop an explanatory model that incorporates theoretically important variables that are potentially influenced by educational policies. Variables that have been explored in prior studies of school effectiveness provide some confirmatory evidence that other researchers in the field find particular variables to be of theoretical importance in school effectiveness research. As such, the variables listed in chapter two represent a large pool of explanatory variables with potential relevance to the present investigation.

Given that the present study is concerned with the achievement of fourth-grade students, all of the variables that were specifically related to third-grade students were discarded during the initial screening of variables while questions referring to students in the “upper grade of Population 1” (i.e., grade 4) were retained. In addition, many of the questions answered by the students, teachers, and administrators had elements that were very specific and also had elements that were more general. For example, several specific questions asked in the student survey were: Does you grandmother live at home with you? Does your mother live at home with you? How many sisters live at home with you? By contrast, a more general form of this question was also asked: How many people live at home with you? In the present study, the more general variables were investigated unless there was a theoretical reason to expect a relationship between a specific variable and school effectiveness. Finally, while many of the variables in the student, teacher, and school questionnaires are interesting in their own right, the literature reviewed in chapter two revealed that some of the variables were simply not theoretically relevant to a study school effectiveness.

Response Rates

Not all of the variables that are theoretically relevant have the capacity to be explored in this study, however. Some items had to be withdrawn from consideration due to low response rates. Because TIMSS is a truly random sample of students and not necessarily teachers (Gonzalez & Smith, 1997), the student-level variables chosen for this study had to have a response rate that led to coverage for more than 90% of the students in a majority of countries in order to be included. It should be noted, however, that the response rates to each variable varied considerably across countries. Countries with low response rates on particular variables are noted in the text and the full descriptive statistics on each variable in each country are presented in Appendix A.

The program that will be used to analyze the data in this study is called HLM4 (Raudenbush, Bryk, Cheong, & Congdon, 2000). The program was designed in such a way that it requires complete data at level two of the analysis in order to generate estimates (in other words, missing data must be dealt with prior to running the program). Consequently, those respondents with missing data at level two were deleted from the analysis on a listwise basis. In practice, this meant that the data used in the present study were drawn from schools with a complete set of data (i.e., students had no missing data, teachers had no missing data, and administrators had no missing data on the level two variables of interest).

Variability

Another important component of the variable selection procedure had to do with the observed variability of each item. In order to use a variable to explain differences in achievement, it was necessary for students to have responded to the item(s) in a non-uniform way. In other words, all or most of the response categories should have been represented for any given variable. Unfortunately, the response categories for some important variables were so narrow that all student responses were able to fit into one or two categories. When most students answer a question the same way, the variable cannot covary with achievement because students do not differ on the predictor variable. Consequently, any variables that had 90% or more student responses falling into just one category in a majority of countries were discarded from analysis.

Model Parsimony

In most explanatory models constructed using regression techniques, there is a tension between the simplicity of the model and the percentage of variance that can be explained by the model. The goal of the present analysis was to explain as much variation in the dependent variable as possible while using as few independent variables as possible, an idea known as scientific parsimony.

One tool for achieving parsimony is through the use of index variables. Index variables are composed of a linear combination of variables that all represent the same construct. In this way, the influence of several related variables can be adequately described by one composite variable that can in turn be incorporated into the explanatory model in a more efficient way.

When possible, variable indices were created in an effort to maximize the amount of information while minimizing the number of variables included in the analysis. Furthermore, there are limits to the number of variables that may be included in the HLM4 program, and a model that includes about 20 variables begins to approach that limit (Arnold & Sedlacek, 1995; Raudenbush et al., 2000).

Measurement Error

Finally, several of the variables surviving the initial screening had to be excluded from this study based on the fact that the items were poorly constructed on the questionnaire. Poorly constructed items are problematic because they introduce an unknown degree of measurement error. Unlike sampling error, which may be estimated statistically, there is no way to directly estimate the amount of measurement error contained in a response. For example, it was determined that a subset of items constructed to address students' involvement in outside activities were likely to contain high levels of measurement error, and were therefore discarded from further consideration in this analysis. In particular, the questions asked students how frequently they participated in various activities such as: attending a concert, going to the theater, and going to the movies. Unfortunately, the student survey was not clear about the operational definition of each of these activities, leading to a very high potential for misinterpretation of the question. To further complicate matters, the response categories given for students to answer were: once per day, once per week, once per month, and rarely. An activity such as going to the theater does not happen every day and it is unclear how frequently such an occurrence would need to happen in order for it to be considered a cultural event. Furthermore, it is impossible to predict how respondents interpreted the questions and therefore any inferences made from these questions would be dubious indeed.

CONSTRUCTING EXPLANATORY MODELS OF SCHOOL EFFECTIVENESS

The approach to model building employed in this study is based upon the methodological approach employed in several recent studies of school effectiveness (see Arnold & Sedlacek, 1995; Hoyle, 2001; Martin et al., 2000; Mullis et al., 1994). In each of these studies, a series of separate, but cumulative explanatory models were constructed and compared. Variables were grouped into explanatory "blocks", or groups of variables that were theoretically connected. For example, a *School Structure* block might consist of three variables related to the physical dimensions of the school such as urban location, school size, and average class size.

This dissertation study will use a total of 20 explanatory variables across six blocks to compare seven separate hierarchical linear models for each subject area. A bottom-up approach to model building with an eye toward variables that could potentially be influenced by specific educational policies will be used. The six blocks will be: (i) student involvement, (ii) instructional methods, (iii) school climate, (iv) school resources, (v) student background, and (vi) mean SES. The block of variables directly related to student attitudes and behaviors will be entered into the equation. After controlling for SES at the student level, the first model will include only those variables in the *Student Involvement* block. Again, after the student level control for SES, the second model will include variables in both the *Student Involvement* and the *Instructional Methods* blocks. The remaining models will be extended in a similar fashion with each model adding one more explanatory block until all six blocks have been included. Finally, a separate model will be constructed that controls for student level SES and uses only the mean school SES as an explanatory variable. This final model will allow the explanatory power of each of the previous six models to be judged in relation to a model with only SES variables.

Based on the criteria for variable inclusion discussed in this chapter, two sets of explanatory models of school effectiveness have been developed for use in this dissertation study, one set for mathematics and one set for science. Table 3.1 presents the sets of explanatory models of school effectiveness in mathematics that will be tested.

Table 3.2 presents the sets of explanatory models of school effectiveness in science that will be tested in this study.

Table 3.1 Variables associated with each explanatory block - Mathematics

EXPLANATORY BLOCK	KEY VARIABLES	VARIABLE CODE	VARIABLE DRAWN FROM
CONTROL	Student-level SES	ASBGHOME + ASBGBOOK + ASBGPS01-16	Student survey
STUDENT INVOLVEMENT	Time on mathematics homework	ASBMDAY7	Student survey
	Likes mathematics	ASBMLIKE **(R)**	Student survey
	Locus of control in mathematics index	ASBMDOW1 **(R)** + ASBMDOW2 + ASBMDOW3 **(R)**	Student survey
INSTRUCTIONAL METHODS	Frequency of worksheet homework in mathematics class	ASBMWSHT **(R)**	Student survey
	Frequency of testing in mathematics class	ASBMTEST **(R)**	Student survey
	Frequency of calculator use in mathematics class	ASBMCALC **(R)**	Student survey
	Frequency of computer use in mathematics class	ASBMCOMP **(R)**	Student survey
CLASSROOM ORGANIZATION	Frequency of problem solving in mathematics class	ASBMPROB **(R)**	Student survey
	Frequency of note taking from the board in mathematics class	ASBMNOTE **(R)**	Student survey
	Frequency of small group work sessions in mathematics class	ASBMSGRP **(R)**	Student survey
SCHOOL CLIMATE	Stability of student body	ACBGENDY	School survey
	Index of minor behavior	ACBGUP01-07	School survey
	Index of major behavior	ACBGUP08-13	School survey
	Instructional leadership from principal	ACBGAC05	School survey
	Teachers years of experience	ATBGTAUG	Teacher survey
	Perception of peer attitudes toward math (1=Positive, 0=Negative)	ASBMFIP2 **(R)**	Student survey
SCHOOL STRUCTURE	Urban location (1=Urban, 0=Other)	ACBGCOMM	School survey
	School size	ACBGBENR + ACBGGENR	School survey
	Average class size in upper grade	ACBGUSIZ	School survey
MEAN SES	School-level SES	ASBGHOME + ASBGBOOK + ASBGPS01-16	Student survey

(R) = Variable will be reverse coded

Table 3.2 Variables associated with each explanatory block - Science

EXPLANATORY BLOCK	KEY VARIABLES	VARIABLE CODE	VARIABLE DRAWN FROM
CONTROL	Student-level SES	ASBGHOME + ASBGBOOK + ASBGPS01-16	Student survey
STUDENT INVOLVEMENT	Time on science homework	ASBSDAY8	Student survey
	Likes science	ASBSLIKE **(R)**	Student survey
	Locus of control in science index	ASBSDOW1 **(R)** + ASBSDOW2 + ASBSDOW3 **(R)**	Student survey
INSTRUCTIONAL METHODS	Frequency of worksheet homework in science class	ASBSWSHT **(R)**	Student survey
	Frequency of testing in science class	ASBSTEST **(R)**	Student survey
	Frequency of calculator use in science class	ASBSCALC **(R)**	Student survey
	Frequency of computer use in science class	ASBSCOMP **(R)**	Student survey
CLASSROOM ORGANIZATION	Frequency of problem solving in science class	ASBSPROB **(R)**	Student survey
	Frequency of note taking from the board in science class	ASBSOTE **(R)**	Student survey
	Frequency of small group work sessions in science class	ASBS SGRP **(R)**	Student survey
SCHOOL CLIMATE	Stability of student body	ACBGENDY	School survey
	Index of minor behavior	ACBGUP01-07	School survey
	Index of major behavior	ACBGUP08-13	School survey
	Instructional leadership from principal	ACBGAC05	School survey
	Teachers years of experience	ATBGTAUG	Teacher survey
	Perception of peer attitudes toward science (1=Positive, 0=Negative)	ASBMFIP2 **(R)**	Student survey
SCHOOL STRUCTURE	Urban location (1=Urban, 0=Other)	ACBGCOMM	School survey
	School size	ACBGBENR + ACBGGENR	School survey
	Average class size in upper grade	ACBGUSIZ	School survey
MEAN SES	School-level SES	ASBGHOME + ASBGBOOK + ASBGPS01-16	Student survey

(R) = Variable will be reverse coded

Recoding

A number of variables will be recoded in order to ease their interpretation. For items with categorical response options, variables will be recoded so that the higher values represented the strongest level of agreement with the prompt. For example, one question asked students how often they had a quiz or a test in their mathematics lesson. The response options for this item were 1 = most lessons, 2 = some lessons, and 3 = never. This variable will be reverse coded so that the most frequent occurrence corresponded with the highest score.

In some cases, variables had only two response categories and were therefore recoded into dummy variables. For example, the item regarding the perception of peers attitudes toward subject matter variable had only two categories and was recoded so that the favorable perceptions were the reference group (receiving a value of one). Variables that were reverse coded and variables that were dummy coded are noted in Table 3.1 and Table 3.2.

Standardization of Explanatory Variables

The dependent variables (mathematics and science scores) will remain in their original units, however, the explanatory variables used in this study will be standardized in order to permit direct comparison among the variables within the model. Each of the variables will be standardized within the country and the mean for each variable will be set to zero with a standard deviation of one. Consequently, the coefficients from the HLM analyses may be interpreted in the same way as Beta coefficients in a typical regression analysis. That is, the coefficients show the amount of change in the dependent variable that may be expected for every one standard deviation unit of change in the explanatory variable.

Control Block

The strength of the relationship between student SES and achievement has been firmly established in the literature (Coleman et al., 1966; Jencks et al., 1972; Plowden Committee, 1967). As such, in order for a model of school effectiveness to be valid, it is critical that the model make an effort to control for differences in student SES before attempting to assess the impact of various explanatory variables upon achievement. For the purposes of this study, a SES index will be created from a linear composite of the following variables:

- Number of people living at home
- Number of books in the home
- Number of country specific possessions (16 items)

Each of the variables in this index will first be standardized to make its mean value equal to zero and standard deviation equal to one within each country. The SES variable will then be created through a linear composite of these standardized variables.

Block 1 - Student Involvement

After controlling for SES, the first block of variables that will be entered into the model are those dealing with student involvement. This explanatory block consists of three variables, all drawn from the student questionnaire. The first variable addresses the amount of time the student spends studying a specific subject area on a normal day. In a study conducted by Martin et al. (2000), the amount of time students spent studying a particular subject was found to be statistically significantly related to school effectiveness in several countries in both grade eight mathematics and science, although it should be noted that the relationship was a negative one (i.e., schools with high average achievement had students who spent less time on homework).

The next variable used in this block addresses the students' attitudes toward the subject (mathematics or science). While Ma & Kishor (1997) report that the relationship between attitude towards mathematics and achievement in mathematics is inconsistent in the literature, they conclude that continued study of the relationship it is still worthy of attention on purely theoretical grounds.

Finally, a locus of control index will be created based upon a linear combination of the following three elements:

- To do well in <u>subject</u> (e.g., mathematics) you need a lot of natural ability (reverse coded)
- To do well in <u>subject</u> (e.g., mathematics) you need good luck
- To do well in <u>subject</u> (e.g., mathematics) you need lots of hard work studying at home (reverse coded)

Several studies have shown a strong relationship between student locus of control and academic achievement (see Benham, 1995 for an excellent review of these studies). Locus of control is a construct that was first defined by Rotter (1966) in relation to social learning theory. In essence, the construct deals with the how students explain the relationship between their own efforts and some outcome (such as achievement). To the extent that the student sees a high relationship between their own efforts and abilities and achievement, that student is said to have a high locus of control. To the extent that the student feels that outcome variables (such as achievement) are influenced by factors beyond their immediate control (e.g., luck, powerful others), that student is said to have a high external locus of control.

In order to create the locus of control index, the first and third items in the index (natural ability to do well; hard work to do well) will be recoded so that high scores on all variables are associated with higher levels of internal locus of control. After the recoding, the locus of control index will be created in the same manner as the SES index, with each separate variable first being standardized with a mean of zero and a standard deviation of one within each country. The index will then be formed through a linear combination of each of the variables, with a high value on the index being equivalent to a high level of internal locus of control.

Block 2 - Instructional Methods

This block consists of four variables related to the instructional techniques employed within the classroom. While there were approximately 20 potential instructional variables to choose from on the teacher and student surveys, those retained in this model have been shown to be related to achievement in prior studies of school effectiveness (see Arnold & Sedlacek, 1995; Mullis et al., 1994). They are:

- How often do you work from worksheets or textbooks alone in your subject area lesson? (reverse coded)
- How often do you have a quiz or test in your subject area lesson? (reverse coded)

- How often do you use calculators in your subject area lesson? (reverse coded)
- How often do you use computers in your subject area lesson? (reverse coded)

In particular, the use of calculators has been shown to be significantly related to school effectiveness, however the relationship has been shown to be negative at grade 4 and positive at grades 8 and 12. Note that the questions employed in this study relate to the frequency of calculator and computer use, and not necessarily to the particular ways in which computers and calculators are used. It is important to note that frequency of computer use is only one of the key elements involved in understanding the relationship between educational technology and school effectiveness. Wenglinsky (1998) has noted that the nature of the relationship between frequency of computer use and achievement is directly related to how the computers are being used for instruction. More specifically, eighth-grade students who reported using computers for higher-order instructional tasks had higher achievement on NAEP while those students who reported using computers for lower-order instructional tasks related to drill and practice tended to have lower scores on NAEP.

It is important to note that although these questions deal with the instructional methods employed at the classroom level, the responses were drawn from the student questionnaires and therefore represent the students' perceptions of the instructional methods employed in the classrooms rather than the teachers' perceptions. Although similar questions were asked on both the teacher and the student questionnaires, it was decided that student responses on this item would more closely approximate what was really going on in the classroom and would suffer from less respondent bias than the teacher responses. Note that each of the variables used in this block will be reverse coded so that the highest value will correspond to the strongest level of agreement with the prompt.

Block 3 – Classroom Organization

The classroom organization block is composed of three variables dealing with the typical structure of the classroom environment. Each of these variables was drawn from the student questionnaire. They are:

- How often does the teacher show you how to do subject area problems in your subject area lesson? (reverse coded)
- How often do you copy notes from the board in you subject area lesson? (reverse coded)
- How often do you work together in pairs or small groups in your subject area lesson? (reverse coded)

The reason that the variable relating to group work is drawn from the student survey is because the issue of small group work was asked in relation to the particular subject area (i.e., mathematics and science) on the student survey, whereas the question was posed in more global terms without reference to any specific subject on the teacher survey. Again, note that each of the variables used in this block will be reverse coded so that the highest value will correspond to the strongest level of agreement with the prompt.

Block 4 - School Climate

The explanatory block that is being called *School Climate* does not capture the full complexity of the construct. For the purposes of this study, school climate factors will be restricted to those that many be classified as relating to the milieu and social systems factors set forth by Anderson (1982). The variables related to these blocks are student stability, major and minor discipline problems in the school, instructional leadership from the principal, teachers years of experience, and students perceptions of peer attitudes toward the subject.

The school climate block consists of six variables. The first variable deals with the stability of the student body and was shown by Mullis et al. (1994) to be positively related to achievement. In other words, higher performing schools tended to have fewer students transferring in and out of their school throughout the academic year. In order to avoid dirty data due to coding errors, the minimum number of students who both began and finished the school year at a particular school was set to 45%. It was considered highly unlikely that a school would have more than 55% of its student body move out of the school during the year so that any value exceeding 55% mobility on the variable was more likely than not a reflection of coding error or a misread of the item.

The next two variables in this block are indices of behavior problems within schools shown to have negative relationships with achievement (Barton, Coley, & Wenglinsky, 1998; Martin et al., 2000). Both indices, one of minor administrative problems and the other of more serious student discipline problems, are made from a linear combination of questions. The index of minor administrative behavior problems was based on research by Martin et al. (2000) and consists of a linear combination of the following items (recall that the term “upper grade” refers to the fourth grade in most countries):

- How often does the school administration or staff have to deal with upper grade students arriving late at school?
- How often does school administration or staff have to deal with upper grade students’ absenteeism without excuse?
- How often does school administration or staff have to deal with upper grade students skipping class periods?
- How often does school administration or staff have to deal with upper grade students violating the dress code?
- How often does school administration or staff have to deal with classroom disturbance by upper grade students?
- How often does school administration or staff have to deal with cheating by upper grade students?
- How often does school administration or staff have to deal with the use of profanity by upper grade students?

The index of major student discipline problems was based on research by Martin et al. (2000) and by Arnold & Sedlacek (1995). This index consists of a linear combination of the following items:

- How often does school administration or staff have to deal with vandalism by upper grade students?
- How often does school administration or staff have to deal with theft by upper grade students?
- How often does school administration or staff have to deal with intimidation of students by upper grade students?
- How often does school administration or staff have to deal with physical injury to students caused by upper grade students?
- How often does school administration or staff have to deal with intimidation of teachers or staff by upper grade students?
- How often does school administration or staff have to deal with physical injury of teachers or staff caused by upper grade students?

For each of these indices, the first step was to standardize each of the individual variables. The index was only created if respondents answered three or more of the items in the index. The mean value of the answered items was then used as the respondents score on the index.

The fourth explanatory variable in this block is drawn from the school background survey and relates to the instructional leadership of the school principal. This variable was included because prior research has shown a positive relationship between the frequency of instruction by the principal and the average achievement in the school (Creemers & Osinga, 1995; Reynolds et al., 2000).

Two other variables that fit under the theoretical umbrella of school climate include the teacher's years of experience (Fuller, 1987; Lockheed & Komenan, 1989; Martin et al., 2000; Reynolds et al., 2000; Scheerens et al., 1989) and the students' perceptions of peer attitudes toward the subject (Brookover et al., 1978; Henderson, Mieszkowski, & Sauvageau, 1978). The number of years experience teaching and the perception of peers attitudes toward the subject both have demonstrated slightly positive relationships with achievement. Note that the variable related to peers' attitudes toward the subject was recoded into a dummy variable.

Block 5 - School Resources

The fifth block of explanatory variables used in this study address school structure. This block consists of three variables drawn from the school background questionnaires. The variables included are urban location (Martin et al., 2000), school size (see Aitkin & Longford, 1986; Arnold & Sedlacek, 1995; Fuller, 1987; Martin et al., 2000)), and average class size (see Arnold & Sedlacek, 1995; Fuller, 1987; Lockheed & Komenan, 1989; Martin et al., 2000; Scheerens et al., 1989). All three variables have been shown to be negatively associated with school effectiveness in the United States. In other countries, however, an urban location can provide the school with more access to academic resources than a rural location and may therefore be positively related to achievement. Note that the urban location variable was recoded into a dummy variable.

PRELIMINARY SCREENING ANALYSIS

After constructing the theoretical model, the next step is to lay out some criteria for the inclusion of countries in the analysis. These criteria specifically deal with the nature of the sampling procedures employed in each country, the extent to which there is variability in achievement among schools within each country, and the extent to which countries are missing data on the level two explanatory variables. Table 3.3 presents the countries that will be included in the present study and the reasons that other countries will be excluded from investigation based upon the various selection criteria. Each of the selection criteria will now be presented and discussed further.

Table 3.3 Countries Meeting Criteria for Inclusion

	YES	NO
Australia	X	
Austria		Missing Background Data
Canada	X	
Cyprus	X	
Czech Republic	X	
England		Missing Background Data
Greece	X	
Hong Kong*	X	
Hungary		Sampling
Iceland		Missing Background Data
Iran, Islamic Republic	X	
Ireland	X	
Israel		Sampling
Japan		Missing Background Data
Korea	X	
Kuwait		Sampling
Latvia	X	
Netherlands		Sampling
New Zealand	X	
Norway		Missing Background Data
Portugal	X	
Scotland		Missing Background Data
Singapore		Missing Background Data
Slovenia	X	
Thailand		Sampling
United States	X	

* Mathematics only

Sampling Procedures

Given that the data collected from TIMSS was cross-sectional in nature and that the goal of this investigation is to generalize to the entire sample of schools within each participating country, adherence to the prescribed sampling procedures is of paramount importance. The two-stage stratified cluster sampling design was employed for both practical and technical reasons. Countries who followed the design would have results that could be generalized at the level of the student, the teacher, and the school. This generalizability was fully realized only when countries adhered to the prescribed sampling plan. Consequently, those countries that did not meet the sampling requirements of TIMSS were excluded from the present analysis.

In order to meet the sampling guidelines, countries had to achieve a 75% participation rate of both students and schools, or a combined rate (the product of school and student participation) of 75% with or without replacement schools (see Mullis et al., 1997, Appendix A). Countries that failed to reach at least 50% school participation without the use of replacement schools, or that failed to reach the sampling participation standard even with the inclusion of replacement schools will be excluded from this study. In addition, countries with unapproved sampling procedures at the classroom level and/or not meeting other guidelines will also be omitted. Countries that will be excluded from the present analysis on the basis of their sampling include Austria, Netherlands, Hungary, Israel, Kuwait, and Thailand.[1]

Between-School and Within-Country Variability in Achievement

In order to explain why some schools have higher average student achievement than others, it is first necessary to determine the extent to which schools differ in the average achievement of their students. The extent to which schools differ in the average achievement of their students is known as between-school variance. If all schools have

[1] It should be noted that although Australia and Latvia (LSS) fell below the technical cutoff of 70% overall participation rates, both countries had participation rates in the sixty percent range even before replacement

the same average student achievement, then there is no between-school variance and all of the variability observed in student test scores can be attributed to within-school differences (i.e., differences in student ability). In a two-level hierarchical model, variability in achievement can be estimated and partitioned into a within-school component and a between-school component. It is the between-school variability that is the focus of the present investigation. Consequently, one of the major criteria for inclusion in this study is that at least 10% of the variation in achievement in the country be attributable to between-school differences.

Missing Questionnaire Data

Despite the fact that 26 countries participated in the student testing component of TIMSS at Population 1, not all countries participated in particular aspects of the background questionnaire component of the study. In attempting to build an explanatory model of school effectiveness, explanatory data were drawn from the student, teacher, and school background questionnaires. Consequently, it was of paramount importance that the countries involved in the present analysis have sufficient data on each of the explanatory variables used in constructing the models. Many times, however, respondents from particular countries did not fill out a particular section of the questionnaires (e.g., no teachers in Japan filled out a background questionnaire). Only countries with sufficient data on the explanatory variables will be used.

---

schools and were just below the cutoff after the use of replacement schools. Given that the present study is somewhat exploratory, it was deemed acceptable to include the results from these countries in this study.

Countries Under Investigation

After conducting the preliminary screening analysis based on the selection criteria listed above, 14 of the 26 countries participating in TIMSS at the fourth grade met the criteria for inclusion in the present investigation. Although the fact that only 14 out of 26 participating countries qualify for the present analysis may seem relatively few, it should be remembered that the use of 14 countries, with 75 schools in each country being evaluated, still qualifies this dissertation study as one of the largest and most diverse studies of school effectiveness conducted to date. Table 3.3 gives a complete listing of all countries that participated in TIMSS and highlights the countries that met the criteria for inclusion in this study.

DATABASE CHARACTERISTICS

The purpose of this section is to discuss the idiosyncrasies involved in working with the TIMSS database that will have an impact upon the interpretation of the results of this analysis. In particular, three issues that are unique to large-scale assessment will be discussed: (i) the use of plausible values in reporting student achievement, (ii) the use of sampling weights in obtaining accurate parameter estimates, and (iii) the methods for calculating sampling error on data from a cluster sampled design.

Plausible Values

The TIMSS tests were designed to cover a broad range of content across the domains of mathematics and science. When testing students at grade four, however, there was a tension between the breadth of knowledge that could be assessed and the amount of time that could realistically be allotted to testing. Consequently, rather than give the same test to all students, which would result in reliable individual estimates but would necessarily limit the coverage of the content domain, TIMSS chose to create eight different versions of the mathematics and science tests at Population 1 in order to expand the domain of knowledge that could be assessed at the country level. Twenty-six sets of items were created and these sets were rotated throughout the eight different test booklets. This design was elegant in that it allowed all items to be linked to one another on a linear scale. This approach also allows for very efficient estimates of population parameters to be made (Adams & Gonzalez, 1996).

The drawback to the design is that because students are each given only a fraction of the entire test, it is difficult to obtain reliable estimates of student proficiency. In order to achieve reliable indices of student proficiency in this situation it was necessary to make use of multiple imputation or 'plausible values' methodology (Adams & Gonzalez, 1996). Given that no one student was able to take all of the items constructed, the plausible value is an estimate of how the individual student would have performed on a test that included all possible items in the assessment. The plausible value is estimated based upon the of the students' responses to the items that were included in the test booklet that the student actually took and the performance of students with similar characteristics. “The proficiency scale scores or plausible values assigned to each student are actually random draws from the estimated ability distribution of students with similar item response patterns and background characteristics. The plausible values are intermediate values that may be used in statistical analyses to provide good estimates of parameters of student populations.” (Gonzalez & Smith, 1997, p. 5-1)

The imputation process involves a degree of error as well, however. Consequently, TIMSS produced not one but five imputed values for each student in mathematics and science. “Since a plausible value is an imputed score that includes a random component, it is customary when using this methodology to draw a number of plausible values for each respondent (usually five). Each analysis is then carried out five times, once with each plausible value, and the results averaged to get the best overall result.” (Gonzalez, 1997, p. 147). The program used in this study will be HLM version 4 (Bryk, Raudenbush, & Congdon, 1996). This program, “…takes the plausible values into account in generating the HLM estimates. For each HLM model, the program runs each of the five plausible values internally, and produces their average value and the correct standard errors.” (Bryk et al., 1996, p. 156).

Sampling Weights

The need for sampling weights arises from the fact that TIMSS employed a two-stage stratified cluster sample design. In this design, each student has a known probability of selection, however, the probability of selection is not equal (as in the case of a simple random sample). Because the probabilities of selection were proportional to school size, rather than equal and random, the use of sampling weights is critical for making accurate estimates of the population parameters. A sampling weight is simply the inverse of the probability of selection of the sampling unit. In TIMSS, there were three levels of sampling unit, the first level was the school, the second level was the classroom and the third level was the student. Each of these sampling units has a corresponding sampling weight. An adjustment for non-response at each level was also factored into the weighting scheme (for more details see Foy, 1997). In most cases, different schools will have different weights, however, classroom weights were often unnecessary since only one classroom in the target grade was selected in most countries. Furthermore, in the same classroom all received the same student weight because the majority of countries selected all students from a particular classroom rather than selecting a subsample of students from a particular classroom. "The overall sampling weight attached to each student record is the product of the intermediate weights: the first stage (school) weight, the second stage (classroom) weight, and the third stage (student) weight." (For further details on the calculation of sampling weights, see Foy, 1997, pp. 71-72).

Gonzalez & Smith (1997) caution that, "Appropriate estimation of population characteristics based on the TIMSS samples requires that the TIMSS sample design be taken into account in all analyses. This is accomplished in part by assigning a weight to each respondent, where the sampling weight properly accounts for the sample design, takes into account any stratification or disproportional sampling of subgroups, and includes adjustments for non-response." (p. 3-12).

Intra-class Correlation

It has already been mentioned that students in the same class tend to be more like each other than if a random sample of students from the entire school was drawn and compared. This is because students in the same classroom share the same learning environment. This shared environment creates more systematic error variance than would be observed in a random sample of students. “This error variance represents the effect of all omitted variables and measurement errors, under the assumption that these errors are unrelated. In traditional linear models, omitted variables are assumed to have a random and not a structural effect, a debatable assumption in data that contain clustered observations.” (Kreft & DeLeeuw, 1998, p. 9). A numeric estimate of the extent to which students are similar in terms of achievement is known as the intra-class correlation.

If intra-class correlation is present, then the assumption of independent observations in the traditional linear model is violated. Practically speaking, this means that 30 students in the same school class are no longer 30 independent observations, but are less than that. How much less depends on the degree of similarity between the group members, and it is this similarity that is estimated from the intra-class correlation (Kreft & DeLeeuw, 1998). Because traditional techniques for analyzing data assume independent observations for the purposes of calculating statistical significance, the presence of an intra-class correlation can make the results from these techniques especially vulnerable to a Type I error. The intra-class correlation for data having a two-level hierarchical structure, “is defined as the proportion of the variance in the outcome that is between second-level units.” (Kreft & DeLeeuw, 1998, p. 9). In practice, the greater the value of the intra-class correlation, the greater the risk of committing a Type I error when using traditional analytic models (e.g., OLS).

Another way to quantify the reduction in design efficiency is through the estimation of the design effect (Kish, 1965). "The design effect for a variable is the ratio of two estimates of the sampling variance for a particular sample statistic: one computed using a technique such as the jackknife that takes all components of variance in the sampling design into account, and the other computed using the simple random sampling formula." (Gonzalez & Foy, 1997, p. 86). The closer this ratio is to 1, the more students from a cluster sample approximate the variability expected of students from a simple random sample.

HIERARCHICAL LINEAR MODELING

The most straightforward approach to studying school effectiveness is to compare schools on the mean achievement of their students. This practice can only be justified, however, if there are few differences in the composition of students in each school. Prior literature has shown that differences in student background are strongly related to differences in student achievement (Coleman et al., 1966; the Plowden Committee, 1967; Jencks et al., 1972). Consequently, to the extent that students from different schools differ in their backgrounds, studies of school effectiveness must attend to this. As Martin et al. (2000) note,

> Schools with a high proportion of well-prepared students from homes and communities with strong support for learning are already well on the way to high achievement levels, regardless of the contribution of the school in terms of

> instruction, facilities, and support. Schools in less-advantaged circumstances face a more difficult challenge. Accordingly, studies of school effectiveness typically attempt to disentangle the organizational and instructional practices of the school from the effects of the abilities and levels of preparation of the student body prior to entering the school (p. 3).

In order to disentangle the effects of student SES and school resources on achievement, the technique of HLM will be used. After statistically adjusting for the effects of student SES, six separate but related models designed to explain the between-school variability in mathematics achievement will be constructed and compared using HLM, with each model being an extension of a previous model. For example, Model 1 will investigate the extent to which three variables related to *Student Involvement* can sufficiently explain the between school variance in mathematics achievement. Model 2 will examine the same research question, but in addition to using *Student Involvement* variables, will also incorporate variables related to *Instructional Methods* into the model. Six theoretical models will be constructed in this way. A seventh model using only school mean SES as a second level explanatory variable will be constructed for comparative purposes.

There is another reason for using HLM to analyze this dataset, however. The data used in the present study are hierarchical in nature. Samples of schools were drawn from each country and samples of students were drawn from each school. Most often, the sample of students in the fourth grade were all drawn from the same classroom within a school, thus leading to the presence of intra-class correlations that were greater than zero. Because this sample of students is more like each other than a random sample of students within a school would be, it is necessary to employ the technique of hierarchical linear modeling to obtain more accurate and efficient estimates. "Conceptually, HLM consists of estimating regressions of regression results, except that the equations at each level are estimated at the same time rather than sequentially, and the variance at one level is taken into account in estimating the next level." (Arnold & Sedlacek, 1995, p. 11).

Using A Two-Level Hierarchical Linear Model[2]

As the name implies, a two-level hierarchical linear model is composed of two steps. Although the two steps occur simultaneously, they will be presented sequentially for explanatory purposes. For this analysis, the first step involves using Ordinary Least Squares (OLS) regression to estimate student achievement as a function of student SES. This results in an equation that yields an intercept value for each school, which may be interpreted as the average mathematics achievement across all schools in the country. The equation also estimates regression coefficients, called Betas, that can be used to describe the association of achievement with student SES level in each school. The student level (Level 1) equation takes the following form:

[2] This overview is based on Arnold, C. & Sedlacek, D. (1995). Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, and 12: Methodology and Results. U.S. Department of Education, NCES 95-697

Level 1: Within-school student-level equation

$$y_{ij} = b_{0j} + b_{1j} X_{1ij} + r_{ij}$$

where:

i represents the ith student
j represents the jth school
y_{ij} represents the achievement score of the ith student in the jth school
b_{0j} is the intercept, or the average achievement in the jth school
b_{1j} is the Beta coefficient for SES in the jth school
X_{1ij} represents the SES score of the ith student in the jth school
r_{ij} is random error in the jth school

The average student achievement will vary from school to school within a country, with some schools having relatively higher mean achievement than others. It is these differences in mean achievement that this study seeks to explain. Consequently, the intercept term will be modeled and will be allowed to vary randomly.

To some extent, the differences in mean achievement between schools may be attributed to the fact that the students in each school come from different background circumstances. Given that student background has been shown to be a strong predictor of student achievement, it is desirable to statistically adjust for differences in student background in order to obtain a more accurate picture of the influence of school factors. As such, the level 1 equation calculates the mean achievement for each school after adjusting for differences in student background within the school. Note that the relationship between student SES and mathematics achievement is assumed to be the same in all schools. In statistical terms, the SES variable is fixed.

In the second step of the HLM analysis, the intercept estimated at level one becomes the outcome measure at level two. In practice, then, the intercept value is used as a dependent variable in a level two equation. The variance of this parameter is then modeled using other explanatory variables. When using the intercept from level one as the dependent variable at level two, the between-school equation produces coefficients called Gammas that estimate the association of each level two explanatory variable with the average mathematics achievement across schools.

In HLM, a model known as the unconditional model is always calculated first. It is called unconditional because it examines the variability in achievement without regard to any explanatory variables. In this way, it is similar to a one-way ANOVA with random effects where achievement is the dependent variable and school membership is the independent variable. The unconditional model is useful for estimating the amount of variance that lies between schools v. the amount of variance that lies within-schools. The unconditional level-two model is presented below:

Level 2: Between-school equations - The Unconditional Model

$b_{0j} = g_{00} + m_{0j}$ (Intercept equation)
$b_{1j} = g_{10}$ (SES equation)

where:

b_{0j} represents the intercept, or the average achievement in the j^th^ school
b_{1j} represents the SES coefficient in the j^th^ school
p is the number of within-school parameter equations
g_{p0} is the intercept, or the average within-school parameter value in the p^th^ equation
m_{pj} is random error in the p^th^ equation

The unconditional model will be used to provide an estimate of the between-school variance in each country. For example, suppose that after calculating the unconditional model it is found that 85% of the variance in student achievement lies within-schools while 15% lies between schools. The task is then to attempt to build an explanatory model that will explain 100% of the 15% of the between-school variance that is potentially explainable. In order to explain the variance, a model may be constructed that includes any number of explanatory variables. This model is known as the conditional model and is presented below:

Level 2: Between-school school equation – The Conditional Model

$$b_{0j} = g_{00} + g_{01}W_{01j} + g_{02}W_{02j} + \ldots + g_{0m}W_{0mj} + m_{0j}$$
$$b_{1j} = g_{10}$$

where:

g_{p1} is the Gamma coefficient for the first school-level variable in the pth equation
g_{p2} is the Gamma coefficient for the second school-level variable in the pth equation
g_{pm} is the Gamma coefficient for the mth school-level variable in the pth equation
W_{p1j} represents the value of the first school-level variable in the jth school in the pth equation
W_{p2j} represents the value of the second school-level variable in the jth school in the pth equation
W_{pmj} represents the value of the mth school-level variable in the jth school in the pth equation
m is the number of school-level parameter variables

The variables that will be used to construct the seven conditional models for explaining the variability in mathematics achievement are listed in Table 3.1. Similarly, seven separate but related conditional models will be constructed to explain the variability in science achievement. The variables that will be used to construct the science model are listed in Table 3.2.

Fixed v. Random Coefficients

In constructing a hierarchical linear model, the researcher must decide which variables should be specified as fixed and which as random at level one. Variables that are thought to vary randomly between schools should include an error term at level one. Variables that are considered to be essentially the same across schools do not need this error term in the level one model and are therefore considered fixed. “The usual purpose of fixing is to allow a more efficient estimate of HLM models if there is in fact no variation around the parameter. . . another purpose of fixing is to add control variables to a within-school equation without losing degrees of freedom in the estimate.” (Arnold & Sedlacek, 1995, p. 21).

Because student level SES is being used as a control variable in this study, the student level SES variable will be fixed and not modeled. The intercept parameter, however, will be allowed to vary randomly because it is the variable of interest in that it represents the average achievement across schools within each country. The variation in the intercept parameter will be modeled as a function of school-level characteristics across the schools.

Variable Centering

An important decision in HLM is whether or not to center the explanatory variables in the model. Centering is used to improve the interpretation of the intercept value. In a normal HLM equation, as a regression, the intercept is the value that occurs when all explanatory variables are set to zero. Variable centering, however, is a technique by which the school mean is subtracted from selected explanatory variables. Although this technique does not impact upon the relationship of the variable with achievement, it does serve to change the interpretation of the intercept. The intercept then becomes the value that occurs when a person scores at the average level on each of the selected predictor variables that has been centered. Because explanatory variables have already been standardized, the centered and uncentered values of the intercept are identical. Consequently, all variables will be entered into the equation uncentered.

SUMMARY

This chapter has outlined the methodology that will be employed in the present study. Details regarding the criteria for variable selection were presented and the explanatory models of school effectiveness in mathematics and science at the fourth grade were developed and discussed. Six separate explanatory models will be constructed using six blocks of variables: (i) student involvement, (ii) instructional methods, (iii) classroom organization, (iv) school climate, and (v) school structure (vi) mean SES. Each of these blocks contains a number of variables that are associated with the blocks and justified in terms of their theoretical importance to this study based upon prior studies of school effectiveness. In addition, a seventh model containing only mean SES will be constructed for comparative purposes.

Based on the criteria for participation in the present study discussed in this chapter, a preliminary analysis determined that fourteen countries met the criteria for inclusion in this investigation in mathematics and thirteen for science (Hong Kong did not meet the criteria in science at the fourth grade). Finally, given the complex nature of the sample design, the analytic technique of Hierarchical Linear Modeling will be used to test the research hypotheses. The important elements of a two-level HLM such as the unconditional model, variable centering, and using fixed v. random coefficients were also presented and explained.

CHAPTER 4 – RESULTS

OVERVIEW

This purpose of this chapter is to present the results of analyses targeted at answering the five major research questions posed in this study. Those research questions are:

1. Is there sufficient variability in achievement between schools in each country to develop a model for explaining that variability?
2. What factors that contribute to schools' effectiveness can be identified and are any of them open to policy manipulation?
3. Are the variables associated with effective schools at the fourth grade stable across different cultural contexts?
4. After correcting for the differences in student intake across schools with regard to SES, how much variance in mathematics and science achievement across schools can be explained by factors associated with: student involvement, instructional methods, classroom organization, school climate, and school structural features?
5. Are the variables associated with effective schools at the fourth grade stable across different subject areas (i.e., are the same variables important in both mathematics and science)?

In order to place the findings in context, this chapter will begin with a presentation of the relevant descriptive statistics for each independent variable used in this study. In order to facilitate interpretation of the data, the results from the United States will be discussed. Next, the mathematics results across all countries will be presented in relation to each of the four research hypotheses. The science results for the same four hypotheses will then be presented. Throughout the presentation of the results, information relevant to the interpretation of important statistics will be presented and discussed using data from the United States for illustrative purposes. Each of the major findings will then be summarized in relation to the five research questions.

DESCRIPTIVE STATISTICS

A series of tables reporting the descriptive statistics on each independent variable used in the models for each country may be found in Appendix A. Table 4.1 presents the descriptive statistics for the United States and will be used to illustrate the information presented in these tables. The information is based upon the unstandardized values for each variable and includes the minimum, maximum, mean, and standard deviation values for each independent variable used in the study. In addition, each variable is accompanied by the number of people responding to the item, the percentage of missing data for that item, and the file from which the item was drawn. Because the same question was sometimes asked of both students and teachers or of both teachers and administrators, it is important to note the file from which the variable was drawn.

All of the data used in this study were drawn from three sources: student, teacher, and school background questionnaires. Consequently, each table presents information on the number of students, teachers, and schools that participated in TIMSS at the upper grade of Population 1 for the country. As Table 4.1 illustrates, in the United States, the number of upper grade students (i.e., level one units) was 7,296; and the number of schools (i.e., level two units) was 189. Next, a listing of each independent variable to be used in the study is presented along with descriptive data based upon the unstandardized values. The table reflects the values after all of the necessary recoding has been implemented, but before the standardization of the variables occurred[3].

In the case of continuous variables, the minimum values represent the lowest numeric value. For example, the smallest school in the United States sample had only 66 students, while the largest school in the United States sample consisted of 1,659 students. Examining the mean value reveals that the mean school size in the United States sample was equal to 417 students. The standard deviation value of 233 indicates that 68 percent of schools in the U.S. sample had between 184 and 650 students.

In cases where responses were categorical, variables were recoded so that the higher values represented the strongest level of agreement. For example, one item asked student how often they had a quiz or a test in their mathematics lesson. The response options were 1 = most lessons, 2 = some lessons, and 3 = never. This variable was reverse coded so that the most frequent occurrence corresponded with the highest score (i.e., -1 would represent the highest value and represent a response of most lessons).

In some cases, variables had only two response categories and were therefore recoded into dummy variables. In interpreting such data, the mean value can be taken to represent the percentage of respondents in the reference group. For example, Table 4.1

---

[3] With the exception of variables that were standardized in order to create indexes (e.g., SES)

shows that the mean value on the urban location variable in the United States is equal to 0.4 indicating that 40% of the schools in the U.S. sample were located in urban areas.

Table 4.1 Descriptive Statistics for the United States at the Upper Grade of Population 1

United_States
N of Students = 7,296
N of Schools = 182

		Min	Max	Mean	(SD)	N	%Missing	File*
LEVEL 1 - STUDENT LEVEL								
Control								
General	Student SES	-9.2	1.3	0.0	(0.68)	7248	1%	S
LEVEL 2 - SCHOOL LEVEL								
Student Involvement								
Mathematics	Locus of control in mathematics	-4.4	1.5	0.0	(0.60)	7187	1%	S
	Time on mathematics homework	1.0	5.0	2.4	(0.82)	7024	4%	S
	Likes mathematics	-4.0	-1.0	-1.7	(0.91)	7143	2%	S
Science	Locus of control in science	-2.7	1.4	0.0	(0.59)	7160	2%	S
	Time on science homework	1.0	5.0	2.1	(0.86)	7004	4%	S
	Likes science	-4.0	-1.0	-1.7	(0.86)	7075	3%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.63)	7054	3%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.57)	7047	3%	S
	Frequency of calculator use	-3.0	-1.0	-2.2	(0.66)	7052	3%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.72)	6987	4%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.7	(0.70)	7020	4%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.60)	7033	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.7	(0.60)	7008	4%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.63)	6969	4%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.2	(0.45)	7123	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.72)	7087	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.58)	7011	4%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.6	(0.67)	7065	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.69)	7038	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.8	(0.62)	6974	4%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.7	(0.45)	7075	3%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.46)	7076	3%	S
General	Percentage of students who begin and finish the year in the same school	45.0	100.0	91.5	(10.54)	135	26%	C
	Index of minor behavior problems	0.0	0.3	0.0	(0.04)	126	31%	C
	Index of major behavior problems	0.0	0.6	0.0	(0.04)	124	32%	C
	Hours per month principal teaches	0.0	60.0	7.6	(17.12)	147	19%	C
	Teachers years of experience	1.0	40.0	14.9	(9.46)	315	17%	T
School Structure								
General	Urban location	0.0	1.0	0.4	(0.49)	149	18%	C
	School size	66.0	1659.0	419.6	(236.54)	130	29%	C
	Average class size	3.0	53.0	23.0	(7.15)	145	20%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

RESEARCH QUESTIONS IN MATHEMATICS

This section will present the results of the HLM analysis targeted at each of the five research questions driving this investigation. Although the research questions were targeted at both mathematics and science achievement, for ease of presentation, this section will address each research question for mathematics and a subsequent section will address the same four research questions in relation to the subject of science. The last section of this chapter will summarize the similarities and differences between the findings in mathematics and science.

Is there sufficient variability in mathematics achievement between schools in each country to develop a model for explaining that variability?

Table 4.2 presents the results of the unconditional model for mathematics achievement. It is important to note that because the HLM4 program does not allow missing data at level two, schools with extensive missing data on any of the explanatory variables of interest were discarded from the analysis. Because the design of this study required seven explanatory models to be constructed and compared, it was critically important that the models be comparing the same sets of schools in each country. Consequently, only the schools that could fulfill the most stringent requirements of the study (i.e., no missing data on any level two explanatory variables to be used in any model) were used for analysis. In practice, as Table 4.2 reveals, this led to a reduction in the school sample across all countries. Whereas the United States had 182 schools in the total sample at the fourth grade, only 92 schools were retained after analysis was restricted to schools with sufficient data on all explanatory variables.

In addition to the sample sizes, Table 4.2 also present a comparison of the variance structures across the two scenarios; (i) the full unconditional model for all schools participating in TIMSS at the fourth grade, and (ii) the unconditional model for the restricted sample of schools meeting the data requirements of the present study. Table 4.2 further illustrates a direct comparison of the impact the sample size reduction has upon the variance structures. Note that in all but three countries the variance structure was the same within +/- 5%. This implies that although the number of schools in the sample are reduced when only those schools with full data are included, the variance structure is in fact similar to the full sample, thereby implying that the remaining schools may have some degree of representativeness in spite of the reduced sample size.

Table 4.2 Comparison of variance components in the full and restricted unconditional models in fourth grade mathematics

Country	Full Unconditional Model			Restricted Unconditional Model		
	N	Betweeen School Variance	Within School Variance	N	Between School Variance	Within School Variance
Australia	178	29%	71%	83	27%	73%
Canada	390	28%	72%	190	26%	74%
Cyprus	146	16%	84%	76	16%	84%
Czech Republic	187	27%	73%	149	24%	76%
Greece	174	41%	59%	96	26%	74%
Hong Kong	124	41%	59%	71	46%	54%
Iran	180	46%	54%	109	44%	56%
Ireland	165	27%	73%	93	25%	75%
Korea	150	21%	79%	132	18%	82%
Latvia (LSS)	125	48%	52%	73	56%	44%
New Zealand	149	47%	53%	89	46%	54%
Portugal	148	31%	69%	31	23%	77%
Slovenia	121	17%	83%	85	17%	83%
United States	182	38%	62%	92	40%	60%

Table 4.3 presents the amount of variance in mathematics achievement that lies between schools in each of the two unconditional models. Between school variance is the parameter of interest in this study, and the table therefore illustrates the comparison of the two models. Again, note that the between school variance in mathematics achievement in the reduced model in Greece is 15% less than that in the full model. While results are comparable in most countries, the table indicates that particular caution should be exercised in interpreting the results of the analysis in Greece, Latvia, and Portugal.

Table 4.3 Comparison of between-school variance in fourth grade mathematics observed in the full and restricted unconditional models

Country	Between School Variance		
	Full Unconditional Model	Restricted Unconditional Model	Difference
Australia	29%	27%	-2%
Canada	28%	26%	-2%
Cyprus	16%	16%	0%
Czech Republic	27%	24%	-3%
Greece	41%	26%	-15%
Hong Kong	41%	46%	5%
Iran	46%	44%	-2%
Ireland	27%	25%	-2%
Korea	21%	18%	-3%
Latvia (LSS)	48%	56%	8%
New Zealand	47%	46%	-1%
Portugal	31%	23%	-8%
Slovenia	17%	17%	1%
United States	38%	40%	2%

Two additional indicators of the similarity of the two unconditional models can be found by examining the amount of parameter variance in each model along with the reliability values.

> Parameter variance, or Tau, is an estimate of the actual non-sampling variation between schools around the parameters of the intercept and SES coefficients in the within-school equations. The parameter variance usually changes between models, where it indicates how much variance there is around each of the two parameters before any between-school variables are taken into account. The purpose of the between-school models is to identify school-level variables that explain, or reduce, this parameter variance, and thus explain school variations in average achievement (Arnold & Sedlacek, 1995, p. 48).

The total variance of each parameter consists of both parameter variance and sampling variance. In HLM, reliability refers to the percentage of the total variance around each parameter that is parameter variance that can be explained, as opposed to sampling variance, which results from error (Arnold & Sedlacek, 1995).

Table 4.4 presents the values for the parameter variance and the reliability in each country under both the full unconditional model and the restricted unconditional model. Again, note that the differences in the reliability are less than .05 in all countries indicating a comparable amount of explainable between-school variation in both models.

Table 4.4 Comparison of parameter variance and reliability across full and restricted unconditional models in fourth grade mathematics

	Full Unconditional Model		Restricted Unconditional Model	
Country	Parameter Variance	Reliability	Parameter Variance	Reliability
Australia	1954.7	0.87	1703.1	0.86
Canada	1675.2	0.70	1557.2	0.70
Cyprus	1114.2	0.77	1154.1	0.77
Czech Republic	1533.0	0.77	1386.6	0.76
Greece	2058.0	0.86	1596.5	0.83
Hong Kong	2143.6	0.94	2605.3	0.95
Iran	1744.1	0.87	1620.0	0.88
Ireland	1513.6	0.77	1331.1	0.76
Korea	653.6	0.71	630.1	0.71
Latvia (LSS)	2611.9	0.90	3366.5	0.92
New Zealand	2763.2	0.88	2881.9	0.88
Portugal	1614.6	0.85	1040.8	0.80
Slovenia	924.4	0.76	932.6	0.76
United States	2193.0	0.92	2526.5	0.93

Finally, Table 4.5 presents a comparison of the mean mathematics achievement for students in the full sample and students in the restricted sample. Despite the reduction in the number of students found in the restricted sample across countries, the mean achievement and standard deviation of the restricted sample are typically very similar to those of the full sample. At the extremes, the mean achievement of the restricted sample in Australia was 12 points higher than in the full sample, however, there was no difference in mean achievement between the full and restricted samples in Slovenia.

Table 4.5 Comparison of means and standard deviations of the full and restricted samples in fourth grade mathematics

Country	Full Sample			Restricted Sample		
	N	Mean	SD	N	Mean	SD
Australia	6,507	546	(92)	3,096	558	(92)
Canada	8,408	532	(84)	3,982	525	(83)
Cyprus	3,376	502	(86)	1,756	509	(86)
Czech Republic	3,268	567	(86)	2,622	566	(85)
Greece	3,053	492	(90)	1,757	495	(87)
Hong Kong	4,411	587	(79)	2,540	593	(82)
Iran	3,385	429	(69)	1,986	435	(69)
Ireland	2,873	550	(85)	1,691	545	(86)
Korea	2,812	611	(74)	2,483	613	(74)
Latvia (LSS)	2,216	525	(85)	1,246	527	(88)
New Zealand	2,421	499	(90)	1,435	498	(87)
Portugal	2,853	475	(80)	607	472	(76)
Slovenia	2,566	552	(82)	1,793	552	(83)
United States	7,296	545	(85)	3,724	534	(89)

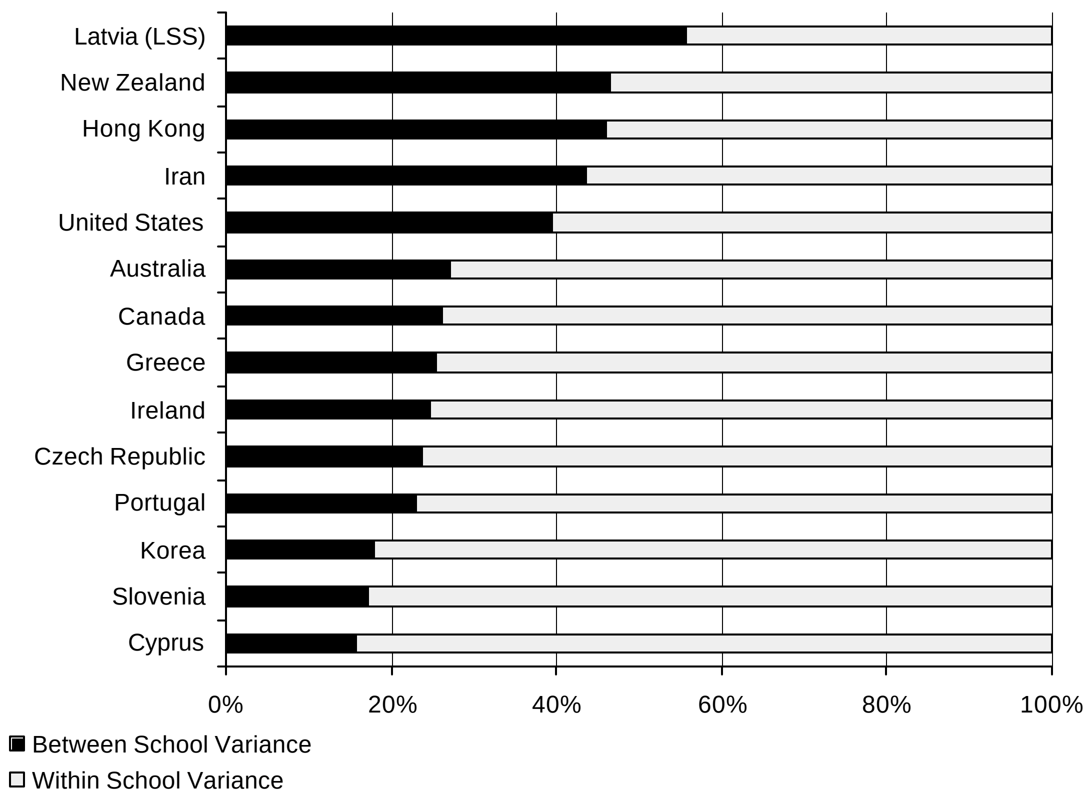
Having established that the restricted model and the full unconditional model are similar, the remainder of this section will be devoted to discussing results based upon the values from the schools that were presented in the restricted unconditional model.

Figure 4.1 graphically displays the range of variability in mathematics achievement that lies between schools and can therefore potentially be explained across the 14 countries eligible for the analysis in mathematics. For the United States, fully 40% of the variance in achievement lies between schools[4]. Recall that in order to be

[4] Note that with the exception of Canada and the United States, the countries in this analysis chose to sample one mathematics classroom in the upper grade in each school. Consequently, the variability in achievement between schools is confounded with the amount of variability that between classrooms.

eligible for analysis schools were required to have at least 10% of the variability in achievement attributable to between-school differences.

Figure 4.1 Graphic display of the between-school variance in the restricted unconditional model in fourth grade mathematics



In sum, each of the 14 countries used in this analysis exhibited sufficient variability in mathematics achievement for use in the present study.

What factors that contribute to schools' effectiveness can be identified and are any of them open to policy manipulation?

Table 4.6 displays the results of the analyses of all seven explanatory models for the United States. This same table was created for each individual country. Those tables may be found in Appendix B.

Table 4.6 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in the United States

UNITED STATES

	Unconditional Model
Between School Variance	40%
Parameter Variance	2526.5
Reliability	0.93

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	48%	53%	63%	69%	73%	78%	61%
		Parameter Variance	1317.7 **	1189.0 **	936.2 **	790.4 **	688.8 **	566.3 **	989.9 **
		Reliability	0.88	0.87	0.85	0.82	0.80	0.77	0.85
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	541.7 **	542.6 **	544.8 **	549.8 **	558.9 **	557.6 **	543.1 **
Control	Student SES	Student socio-economic status (SES)	20.8 **	20.6 **	20.4 **	20.2 **	20.3 **	19.3 **	19.3 **
Block 1	Student Involvement	Time spent studying mathematics	-56.0 **	-48.2 **	-36.5 **	-40.6 **	-21.4	-10.2	
		How much students like mathematics	-19.2	-14.8	-12.2	-2.4	-10.3	0.0	
		Locus of control in mathematics	131.1 **	108.3 **	109.3 **	94.1 **	117.1 **	85.4 **	
Block 2	Instructional Methods	Frequency of worksheets in class		20.2	18.8	20.3 *	12.8	9.3	
		Frequency of testing		-20.7 *	-7.7	-6.4	-5.8	-3.1	
		Frequency of calculator use		-2.9	-2.0	-1.5	1.3	-2.9	
		Frequency of computer use		-13.9 **	-3.4	0.6	1.5	3.9	
Block 3	Classroom Organization	Students take notes in class			-44.7 **	-35.3 **	-29.2 **	-20.5 **	
		Students do problems in class			3.2	2.7	7.1	-1.0	
		Students work in groups in class			0.2	-4.8	-10.0	-13.3	
Block 4	School Climate	Stability of student body				206.6 **	143.4 *	31.7	
		Index of major discipline problems				70.8	50.8	27.7	
		Index of minor discipline problems				-203.3	-155.9	-118.9	
		Principal leadership				114.0	79.9	72.3	
		Teacher's years of experience				-6.6	-5.7	-5.3	
		Students perception of peer attitudes toward mathematics				2.6	-5.2	1.0	
Block 5	School Structure	Urban location					-6.7	-6.2	
		School size					-69.0	-68.8	
		Class size					-289.5 **	-302.0 **	
Block 6	Mean SES	Mean socio-economic status of the school						53.7 **	95.8 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table 4.6 presents the results for each of the seven models tested in the United States. The first thing to note in this table is the between-school variance specified by the unconditional model. In the U.S., this value is equal to 40%. The fact that the reliability of this value is equal to 0.93 means that 93% of the between-school variance in achievement is actual parameter variance that is potentially explainable by a correctly specified model and the remaining 7% is due to sampling variance.

Next, each model is presented along with information about the amount of between-school variance explained by the model, the remaining parameter variance, and the reliability. The amount of explained variance is known as R^{2*}. "To obtain the R^{2*} for a parameter in a between-school model, the difference between the original parameter variance in the unconditional model and the parameter variance left from each conditional between-school mode is divided by the original parameter variance." (Arnold & Sedlacek, 1995, p. 49). The asterisks next to the parameter values indicate that a significant amount of parameter variance still remains to be explained. For example, in the United States, although Model 6 explains 78% of the between-school variance in achievement, a significant amount of parameter variance still remains to be explained by a more complete explanatory model.

Finally, note that the Beta values for each explanatory values are presented for each of the seven explanatory models. Recall that each of the explanatory variables has been standardized and may therefore be directly compared in a meaningful way. The intercept provides an estimate of the average mathematics achievement across schools in the country. Consequently, when evaluating Model 6, a proper interpretation would be that, holding all other things in the model constant, schools who average one standard deviation above the average amount for the variable “students take notes in class”, will tend to have scores that are 20.5 points less than the average mathematics achievement across schools in the United States (i.e., 557.6 – 20.5 = 537.1).

Note that these summary tables present statistically significant values at both the alpha = .10 and alpha = .05 levels. Although many variables may appear to have large effects on the basis of their Beta values, it should be noted that the standard errors have not been provided in the tables. Consequently, if a variable has a large effect and yet is not indicated as statistically significant, it is usually because the variable has a large standard error value.

The summary tables provided here and in Appendix B serve to highlight the individual variables that were significantly related to mathematics achievement across all seven models.

Are the variables associated with effective schools at the fourth grade stable across different cultural contexts?

Table 4.7 provides an overall summary of the number of variables that were found to be statistically significant at the alpha = .10 level across the six explanatory models in the 14 participating countries. The .10 alpha level was chosen in order to provide results that are consistent with and comparable to prior studies of school effectiveness (see Arnold & Sedlacek, 1995; Martin et al., 2000).

Table 4.7 Number of variables found to be significant at the .10 level across models in fourth grade mathematics (14 countries)

	Explanatory Block	Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
			Student Involvement	Model 1 with Instructional Methods	Model 2 with Classroom Organization	Model 3 with School Climate	Model 4 with School Structure	Model 5 with Mean SES
Block 1	Student Involvement	Time spent studying mathematics	4	5	5	3	2	2
		How much students like mathematics	5	3	3	2	1	2
		Locus of control in mathematics	11	8	8	8	9	14
Block 2	Instructional Methods	Frequency of worksheets in class		2	1	2	1	1
		Frequency of testing		2	1	0	0	0
		Frequency of calculator use		3	2	1	1	2
		Frequency of computer use		6	6	4	5	5
Block 3	Classroom Organization	Students take notes in class			2	2	2	2
		Students do problems in class			1	2	2	1
		Students work in groups in class			1	1	2	0
Block 4	School Climate	Stability of student body				3	2	1
		Index of major discipline problems				2	2	0
		Index of minor discipline problems				0	0	1
		Principal leadership				0	1	1
		Teacher's years of experience				2	2	2
		Students perception of peer attitudes toward mathematics				4	3	2
Block 5	School Structure	Urban location					3	0
		School size					1	2
		Class size					3	4
Block 6	Mean SES	Mean socio-economic status of the school						12

* results statistically significant at the .10 alpha level
SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Although a number of variables were listed as statistically significant in Table 4.7, it is important to note, that the direction of significant effects were not always consistent across countries. Tables 4.8 - 4.13 serve to document the direction of the statistically significant effects reported for each one of the six major explanatory models that were developed (the SES Alone model is not included here).

Table 4.7 shows that in explanatory Model 1, the variable related to time spent studying mathematics was statistically significant in four countries. Table 4.8, however, reveals that this relationship was significantly positively associated with school effectiveness in Korea, but the same variable was significantly negatively associated with school effectiveness in Cyprus, Latvia and the United States.

Tables 4.7 – 4.13 also reveal the fact that variables that are significant in one model may no longer be significant in other models. Consequently, these tables serve to document the stability of variable effects across different cultural contexts.

Table 4.8 Indication of the direction of statistically significant results for fourth grade mathematics – Model 1

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math			-						+	-				-	**3**	**1**
		how much students like math				-			-		+		-		-		**4**	**1**
		locus of control in math	+	+	+	+		+	+	+	+		+		+	+	**0**	**11**

* results statistically significant at the .10 alpha level

Table 4.9 Indication of the direction of statistically significant results for fourth grade mathematics – Model 2

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math			-					-	+	-				-	**4**	**1**
		how much students like math				-					+				-		**2**	**1**
		locus of control in math	+			+		+		+	+		+		+	+	**0**	**8**
Block 2	Instructional Methods	worksheets in class	+												+		**0**	**2**
		tests								-						-	**2**	**0**
		calculator use		+				-	-								**2**	**1**
		computer use		-						-		-	-		-	-	**6**	**0**

* results statistically significant at the .10 alpha level

Table 4.10 Indication of the direction of statistically significant results for fourth grade mathematics – Model 3

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math	-		-						+	-				-	**4**	**1**
		how much students like math				-					+				-		**2**	**1**
		locus of control in math		+		+		+		+	+		+		+	+	**0**	**8**
Block 2	Instructional Methods	worksheets in class		+													**0**	**1**
		tests					+										**0**	**1**
		calculator use						-	-								**2**	**0**
		computer use	-	-						-		-	-		-		**6**	**0**
Block 3	Classroom Organization	notes in class								-						-	**2**	**0**
		problems in class	-														**1**	**0**
		works in groups in class												-			**1**	**0**

* results statistically significant at the .10 alpha level

Table 4.11 Indication of the direction of statistically significant results for fourth grade mathematics – Model 4

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math	-		-											-	**3**	**0**
		how much students like math				-									-		**2**	**0**
		locus of control in math	+	+		+		+		+	+				+	+	**0**	**8**
Block 2	Instructional Methods	worksheets in class	+													+	**0**	**2**
		tests															**0**	**0**
		calculator use						-									**1**	**0**
		computer use	-	-								-	-				**4**	**0**
Block 3	Classroom Organization	notes in class								-						-	**2**	**0**
		problems in class	-			-											**2**	**0**
		works in groups in class								-							**1**	**0**
Block 4	School Climate	Stability of student body		+								-				+	**1**	**2**
		Index of major discipline problems					+						-				**1**	**1**
		Index of minor discipline problems															**0**	**0**
		Principal leadership															**0**	**0**
		Teacher's years of experience					+		+								**0**	**2**
		Perception of peer attitudes toward math	-				+				+	-					**2**	**2**

* results statistically significant at the .10 alpha level

Table 4.12 Indication of the direction of statistically significant results for fourth grade mathematics – Model 5

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math	-		-												**2**	**0**
		how much students like math													-		**1**	**0**
		locus of control in math	+	+		+		+		+	+		+		+	+	**0**	**9**
Block 2	Instructional Methods	worksheets in class	+														**0**	**1**
		tests															**0**	**0**
		calculator use						-									**1**	**0**
		computer use	-	-						-		-	-				**5**	**0**
Block 3	Classroom Organization	notes in class										-				-	**2**	**0**
		problems in class				-								-			**2**	**0**
		works in groups in class					-			-							**2**	**0**
Block 4	School Climate	Stability of student body								+						+	**0**	**2**
		Index of major discipline problems					+						-				**1**	**1**
		Index of minor discipline problems															**0**	**0**
		Principal leadership					-										**1**	**0**
		Teacher's years of experience					+		+								**0**	**2**
		Perception of peer attitudes toward math	-				+				+						**1**	**2**
Block 5	School Structure	Urban location			+	+			+								**0**	**3**
		School size								+							**0**	**1**
		Class size		-										-		-	**3**	**0**

* results statistically significant at the .10 alpha level

Table 4.13 Indication of the direction of statistically significant results for fourth grade mathematics – Model 6

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Hong Kong	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying math	-		-												**2**	**0**
		how much students like math												+	-		**1**	**1**
		locus of control in math	+	+	+	+	-	+	+	+	+	-	+	+	+	+	**2**	**12**
Block 2	Instructional Methods	worksheets in class	+														**0**	**1**
		tests															**0**	**0**
		calculator use							-					-			**2**	**0**
		computer use	-	-						-		-	-				**5**	**0**
Block 3	Classroom Organization	notes in class										-				-	**2**	**0**
		problems in class												-			**1**	**0**
		works in groups in class															**0**	**0**
Block 4	School Climate	Stability of student body			-												**1**	**0**
		Index of major discipline problems															**0**	**0**
		Index of minor discipline problems												-			**1**	**0**
		Principal leadership					-										**1**	**0**
		Teacher's years of experience					+							-			**1**	**1**
		Perception of peer attitudes toward math	-								+						**1**	**1**
Block 5	School Structure	Urban location															**0**	**0**
		School size					-			+							**1**	**1**
		Class size		-							-			-		-	**4**	**0**
Block 6	Mean SES	Mean Socio-economic status of the school		+	+	+	+	+	+	+	+		+	+	+	+	**0**	**12**

* results statistically significant at the .10 alpha level

Tables 4.8 – 4.13 show that although the direction of significant effects did vary on some variables, in general, the direction of significant effects were stable across different cultural contexts.

After correcting for the differences in student intake across schools with regard to SES, how much variance in mathematics and science achievement across schools can be explained by each explanatory model?

When evaluating the explanatory blocks as a whole, the best indication of their explanatory power is the R^{2*}, the explained variance. Recall that each model is cumulative and thus Model 2 includes all variables from the *Student Involvement* block as well as the variables from the *Instructional Methods* block. Table 4.14 allows the explanatory power of each model to be compared both across models and within countries. Model 1 ranged widely in explanatory power from a low of 6% in Latvia to a high of 52% in Korea. Model 5 ranged from a low of 12% in the Czech Republic to a high of 73% in the United States. A low amount of explained variance indicates that the explanatory model was, for the most part, incorrectly specified for the data. A high percentage of explained variance means that the model was correctly specified.

Table 4.14 Between-school variance in mathematics achievement explained by each explanatory model in fourth grade mathematics

	Explained Variance						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Country	Student Involvement	Model 1 with Instructional Methods	Model 2 with Classroom Organization	Model 3 with School Climate	Model 4 with School Structure	Model 5 with Mean SES	SES Model
Australia	34%	41%	44%	47%	48%	49%	20%
Canada	36%	39%	39%	40%	41%	45%	27%
Cyprus	20%	23%	19%	19%	23%	31%	28%
Czech Republic	16%	13%	13%	12%	12%	20%	19%
Greece	13%	17%	17%	28%	26%	39%	29%
Hong Kong	35%	44%	43%	42%	41%	55%	26%
Iran	10%	22%	21%	36%	42%	60%	52%
Ireland	25%	37%	46%	51%	56%	60%	24%
Korea	52%	55%	54%	58%	60%	67%	52%
Latvia (LSS)	6%	20%	19%	27%	27%	27%	2%
New Zealand	34%	44%	45%	47%	46%	60%	63%
Portugal	38%	32%	54%	58%	69%	77%	46%
Slovenia	43%	52%	52%	47%	46%	52%	31%
United States	48%	53%	63%	69%	73%	78%	61%

RESEARCH QUESTIONS IN SCIENCE

This section will present the results of the HLM analysis targeted at each of the four research questions articulated in this study. Although the research questions were targeted at both mathematics and science achievement, this section will address each research question for science.

Is there sufficient variability in science achievement between schools in each country to develop a model for explaining that variability?

Table 4.15 presents the results of the unconditional model for science achievement. Again, note that because the HLM4 program does not allow missing data at level two, schools with extensive missing data on any of the explanatory variables of interest were discarded from the analysis. Because the design of this study require seven explanatory models to be constructed and compared, it was critically important that the models be comparing the same sets of schools in each country. Consequently, only the schools that could fulfill the most stringent requirements of the study (i.e., no missing data an any level two explanatory variable to be used in any model) were used for analysis. In practice, as Table 4.15 reveals, this led to a reduction in the school sample across all countries. Whereas the United States had 182 schools in the total sample at the fourth grade, only 92 schools were retained after analysis was restricted to schools satisfying the data requirements of the study,

In addition to the sample sizes, Table 4.15 also present a comparison of the variance structures across the two scenarios, the full unconditional model, and the unconditional model for the restricted sample of schools meeting the requirements of the study. Table 4.15 further illustrates a direct comparison of the impact the sample size reduction has upon the variance structures.

Table 4.15 Comparison of variance components in the full and restricted unconditional models in fourth grade science

Country	Full Unconditional Model			Restricted Unconditional Model		
	N	Betweeen School Variance	Within School Variance	N	Between School Variance	Within School Variance
Australia	178	25%	75%	83	20%	80%
Canada	390	22%	78%	190	20%	80%
Cyprus	146	18%	82%	76	17%	83%
Czech Republic	187	27%	73%	149	23%	77%
Greece	174	39%	61%	96	24%	76%
Iran	180	44%	56%	109	43%	57%
Ireland	165	28%	72%	93	29%	71%
Korea	150	15%	85%	132	12%	88%
Latvia (LSS)	125	53%	47%	73	59%	41%
New Zealand	149	49%	51%	89	52%	48%
Portugal	148	31%	69%	31	22%	78%
Slovenia	121	17%	83%	85	17%	83%
United States	182	40%	60%	92	42%	58%

Table 4.16 presents the amount of variance in science achievement that lies between schools in each of the two unconditional models. Between school variance is the parameter of interest in this study, and Table 4.16 illustrates the comparison of the two models. While results are comparable in most countries, the table indicates that particular caution should be exercised in interpreting the results of the analysis in Greece, Latvia, and Portugal.

Table 4.16 Comparison of between-school variance in fourth grade science observed in the full and restricted unconditional models

Country	Between School Variance		
	Full Unconditional Model	Restricted Unconditional Model	Difference
Australia	25%	20%	-5%
Canada	22%	20%	-2%
Cyprus	18%	17%	0%
Czech Republic	27%	23%	-4%
Greece	39%	24%	-15%
Iran	44%	43%	-1%
Ireland	28%	29%	2%
Korea	15%	12%	-3%
Latvia (LSS)	53%	59%	6%
New Zealand	49%	52%	3%
Portugal	31%	22%	-9%
Slovenia	17%	17%	1%
United States	40%	42%	2%

Two additional indicators of the similarity of the two unconditional models can be found by examining the amount of parameter variance in each model along with the reliability values. Table 4.17 presents the values for the parameter variance and the reliability in each country under both the full unconditional model and the restricted unconditional model. Again, note that the differences in the reliability are less than .06 in all countries indicating a comparable amount of explainable between school variation in both models.

Table 4.17 Comparison of parameter variance and reliability across full and restricted unconditional models in fourth grade science

Country	Full Unconditional Model		Restricted Unconditional Model	
	Parameter Variance	Reliability	Parameter Variance	Reliability
Australia	1707.2	0.84	1234.9	0.81
Canada	1467.0	0.65	1301.9	0.64
Cyprus	955.7	0.79	957.8	0.79
Czech Republic	1370.4	0.77	1182.5	0.75
Greece	2605.2	0.85	1301.1	0.81
Iran	1944.8	0.87	1865.8	0.87
Ireland	1574.2	0.78	1695.6	0.80
Korea	374.1	0.61	342.9	0.59
Latvia (LSS)	2654.5	0.91	3445.2	0.92
New Zealand	3397.7	0.89	4045.8	0.90
Portugal	1787.8	0.85	1138.7	0.79
Slovenia	825.6	0.76	834.6	0.77
United States	2931.9	0.93	3271.9	0.94

Finally, Table 4.18 presents a comparison of the mean science achievement for students in the full sample and students in the restricted sample. Despite the reduction in the number of students found in the restricted sample across countries, the mean achievement and standard deviation of the restricted sample are typically very similar to those of the full sample. At the extremes, the mean achievement of the restricted sample in the United States was 14 points lower than in the full sample, however, there was no difference in ,mean achievement between the full and restricted samples in New Zealand.

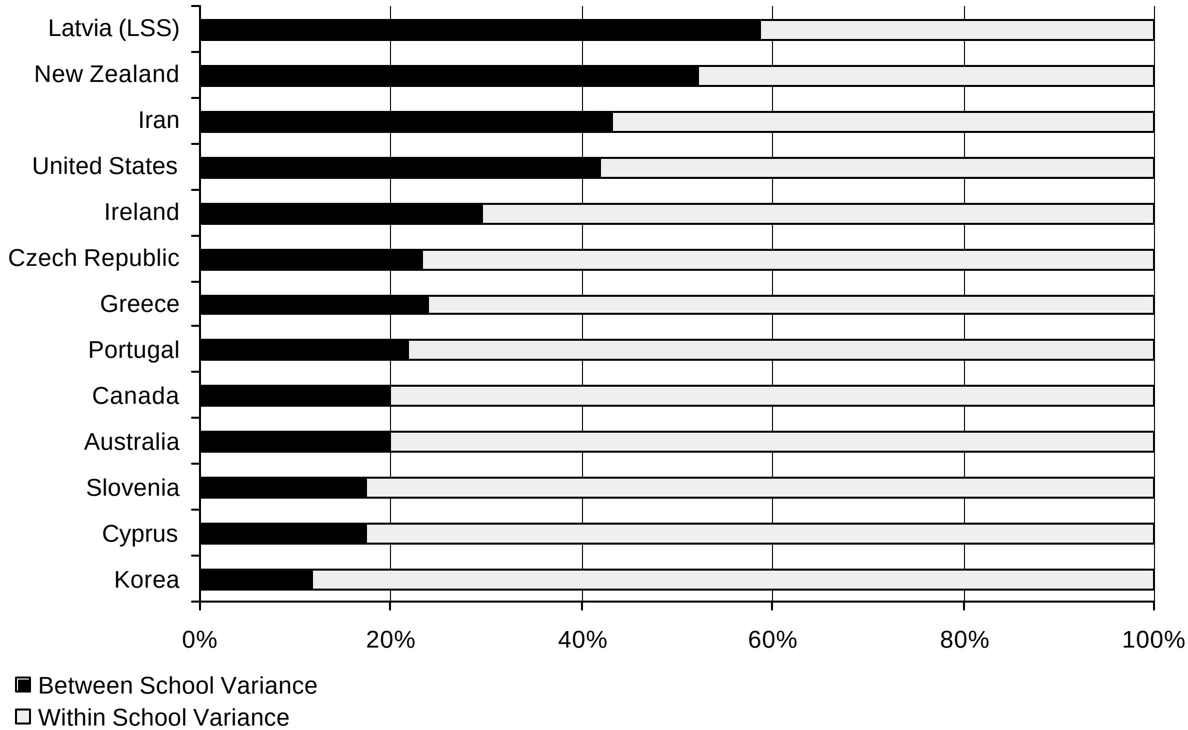
Table 4.18 Comparison of means and standard deviations in the full and restricted samples in science

Country	Full Sample			Restricted Sample		
	N	Mean	SD	N	Mean	SD
Australia	6,507	562	(93)	3,096	574	(91)
Canada	8,408	549	(86)	3,982	540	(90)
Cyprus	3,376	475	(76)	1,756	481	(74)
Czech Republic	3,268	557	(81)	2,622	556	(80)
Greece	3,053	497	(83)	1,757	499	(80)
Iran	3,385	416	(74)	1,986	418	(73)
Ireland	2,873	539	(85)	1,691	536	(87)
Korea	2,812	597	(68)	2,483	599	(68)
Latvia (LSS)	2,216	512	(84)	1,246	514	(87)
New Zealand	2,421	531	(97)	1,435	531	(96)
Portugal	2,853	480	(84)	607	478	(81)
Slovenia	2,566	546	(76)	1,793	545	(75)
United States	7,296	565	(95)	3,724	551	(100)

Having established that the restricted model and the full unconditional model are similar, the remainder of this section will be devoted to discussing results based upon the values from the schools that were presented in the restricted unconditional model.

Figure 4.2 graphically displays the range of variability in science achievement that lies between schools and can therefore potentially be explained across the 13 countries eligible for the analysis in science. For the United States, 42% of the variance in science achievement lies between-schools. Recall that in order to be eligible for analysis schools were required to have at least 10% of the variability in achievement attributable to between-school differences.

Figure 4.2 Graphic display of the between-school variance in the restricted unconditional model in fourth grade science



In sum, each of the 13 countries used in this analysis exhibited sufficient variability in science achievement for use in the present study.

What factors that contribute to schools' effectiveness can be identified and are any of them open to policy manipulation?

Table 4.19 displays the results of the analyses of all seven explanatory models for the United States. This same table was created for each individual country. Those tables may be found in Appendix C.

Table 4.19 Results of seven explanatory models of school effectiveness in science at the fourth grade in the United States

UNITED STATES

	Unconditional Model
Between School Variance	42%
Parameter Variance	3271.9
Reliability	0.94

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	29%	49%	49%	66%	73%	83%	69%
		Parameter Variance	2317.4 **	1670.3 **	1666.0 **	1103.8 **	879.1 **	549.9 **	1018.4 **
		Reliability	0.92	0.89	0.89	0.85	0.82	0.74	0.84
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	556.8 **	562.7 **	562.1 **	566.8 **	579.6 **	576.0 **	562.0 **
Control	Student SES	Student socio-economic status (SES)	26.5 **	26.5 **	26.3 **	26.0 **	26.0 **	24.4 **	24.4 **
Block 1	Student Involvement	Time spent studying science	-24.6 *	0.0	-2.6	-7.6	-13.4	4.3	
		How much students like science	11.9	9.8	9.0	-5.1	-9.4	-0.8	
		Locus of control in science	90.7 **	93.4 **	94.9 **	83.2 **	87.1 **	39.1	
Block 2	Instructional Methods	Frequency of worksheets in class		18.7	25.6	16.8	24.8 *	1.3	
		Frequency of testing		-33.7 **	-45.2 **	-36.1 **	-33.1 **	-19.1 **	
		Frequency of calculator use		-35.6 **	-35.5 **	-30.4 **	-29.0 **	-25.3 **	
		Frequency of computer use		-23.9 **	-24.4 **	-9.5	-11.4	-5.6	
Block 3	Classroom Organization	Students take notes in class			19.9 *	12.5	8.6	4.9	
		Students do problems in class			-3.9	-2.7	0.5	5.2	
		Students works in groups in class			-1.8	-2.5	5.3	-0.9	
Block 4	School Climate	Stability of student body				376.7 **	334.6 **	114.7	
		Index of major discipline problems				73.1	39.1	41.7	
		Index of minor discipline problems				-354.5 **	-267.1 *	-229.4 *	
		Principal leadership				-49.2	-229.0	-72.2	
		Teacher's years of experience				0.5	0.5	1.0	
		Students perception of peer attitudes toward science				16.5	13.0	15.1	
Block 5	School Structure	Urban location					-15.7 **	-8.9	
		School size					-184.0 **	-117.5 **	
		Class size					-87.8	-137.0	
Block 6	Mean SES	Mean socio-economic status of the school						74.5 **	115.6 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table 4.19 presents the results for each of the seven models tested in the United States. The first thing to note in this table is the between-school variance specified by the unconditional model. In the U.S., this value is equal to 42%. The fact that the reliability of this value is equal to 0.94 means that 94% of the between-school variance in achievement is actual parameter variance that is potentially explainable by a correctly specified model and the remaining 6% is due to sampling variance.

Next, each model is presented along with information about the amount of between-school variance explained by the model, the remaining parameter variance, and the reliability. "To obtain the R^{2*} for a parameter in a between-school model, the difference between the original parameter variance in the unconditional model and the parameter variance left from each conditional between-school model is divided by the original parameter variance." (Arnold & Sedlacek, 1995, p. 49). The asterisks next to the parameter values indicate that a significant amount of parameter variance still remains to be explained. For example, in the United States, although Model 6 explains 83% of the between-school variance in achievement, a significant amount of parameter variance still remains to be explained by a more complete explanatory model.

Finally, note that the Beta values for each explanatory values are presented for each of the seven explanatory models. Recall that each of the explanatory variables has been standardized and may therefore be directly compared in a meaningful way. The intercept provides an estimate of the average mathematics achievement across schools in the country. Consequently, when evaluating Model 6, a proper interpretation would be that, holding all other variables in the model constant, schools who average one standard deviation above the average amount for the variable "frequency of testing", will tend to have scores that are 19.1 points less than the average science achievement across schools in the United States (i.e., 576.0 – 19.1 = 556.9).

Note that these summary tables present statistically significant values at both the alpha = .10 and alpha = .05 levels. Although many variables may appear to have large effects on the basis of their Beta values, it should be noted that the standard errors have not been provided in the tables. Consequently, if a variable has a large effect and yet is not indicated as statistically significant, it is usually because the variable has a large standard error value.

The summary tables provided here and in Appendix B serve to highlight the individual variables that were significantly related to mathematics achievement across all seven models.

Are the variables associated with effective schools at the fourth grade stable across different cultural contexts?

Table 4.20 provides an overall summary of the number of variables that were found to be statistically significant at the alpha = .10 level across the six explanatory models in the 13 participating countries. The .10 alpha level was chosen so that the results would be consistent with and comparable to prior studies of school effectiveness (Arnold & Sedlacek, 1995; Martin et al., 2000).

Table 4.20 Number of variables found to be significant at the .10 level across models in fourth grade science (13 countries)

	Explanatory Block	Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
			Student Involvement	Model 1 with Instructional Methods	Model 2 with Classroom Organization	Model 3 with School Climate	Model 4 with School Structure	Model 5 with Mean SES
Block 1	Student Involvement	Time spent studying science	4	2	4	2	2	2
		How much students like science	3	3	3	4	3	3
		Locus of control in science	8	5	6	6	6	1
Block 2	Instructional Methods	Frequency of worksheets in class		1	2	1	2	1
		Frequency of testing		3	2	3	2	2
		Frequency of calculator use		7	7	6	7	4
		Frequency of computer use		4	5	4	4	4
Block 3	Classroom Organization	Students take notes in class			3	2	2	3
		Students do problems in class			3	2	2	2
		Students work in groups in class			0	0	0	0
Block 4	School Climate	Stability of student body				3	3	1
		Index of major discipline problems				1	1	1
		Index of minor discipline problems				2	2	1
		Principal leadership				1	1	2
		Teacher's years of experience				2	2	2
		Students perception of peer attitudes toward science				2	2	2
Block 5	School Structure	Urban location					3	1
		School size					1	2
		Class size					3	2
Block 6	Mean SES	Mean socio-economic status of the school						10

* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Although a number of variables were listed as statistically significant in Table 4.20, it is important to note, that the direction of the significant effects were not always consistent across countries. Tables 4.21- 4.26 serve to document the direction of the statistically significant effects reported for each one of the six major explanatory models that were developed (the SES Alone model is not included here).

Table 4.20 shows that in explanatory Model 6, the variable related to students taking notes in science class was significantly related to school effectiveness in three countries. Table 4.26, however, reveals that this relationship was significantly positively associated with school effectiveness in Australia, but the same variable was significantly negatively associated with school effectiveness in Latvia and New Zealand.

Tables 4.21 – 4.26 also reveal the fact that variables that are significant in one model may no longer be significant in other models. Consequently, these tables serve to document the stability of variable effects across different cultural contexts.

Table 4.21 Indication of the direction of statistically significant results for fourth grade science – Model 1

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science		-	-						-				-	**4**	**0**
		how much students like science				+			+		+					**0**	**3**
		locus of control in science	+	+			+	+	+	+		+			+	**0**	**8**

* results statistically significant at the .10 alpha level

Table 4.22 Indication of the direction of statistically significant results for fourth grade science – Model 2

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science		-	-											**2**	**0**
		how much students like science				+			+		+					**0**	**3**
		locus of control in science	+						+	+		+			+	**0**	**5**
Block 2	Instructional Methods	worksheets in class										-				**1**	**0**
		tests						+	-						-	**2**	**1**
		calculator use		-		-	-	-	-	-					-	**7**	**0**
		computer use		-							-	-			-	**4**	**0**

* results statistically significant at the .10 alpha level

Table 4.23 Indication of the direction of statistically significant results for fourth grade science – Model 3

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science	-	-	-								-			**4**	**0**
		how much students like science				+			+		+					**0**	**3**
		locus of control in science	+	+					+	+		+			+	**0**	**6**
Block 2	Instructional Methods	worksheets in class										-	+			**1**	**1**
		tests							-						-	**2**	**0**
		calculator use		-		-	-	-	-	-					-	**7**	**0**
		computer use		-							-	-		-	-	**5**	**0**
Block 3	Classroom Organization	notes in class					-		-						+	**2**	**1**
		problems in class				+	-							-		**2**	**1**
		works in groups in class														**0**	**0**

* results statistically significant at the .10 alpha level

Table 4.24 Indication of the direction of statistically significant results for fourth grade science – Model 4

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science	-	-												**2**	**0**
		how much students like science				+		-	+		+					**1**	**3**
		locus of control in science	+	+					+	+		+			+	**0**	**6**
Block 2	Instructional Methods	worksheets in class										-				**1**	**0**
		tests						+	-						-	**2**	**1**
		calculator use		-		-	-		-	-					-	**6**	**0**
		computer use		-							-	-		-		**4**	**0**
Block 3	Classroom Organization	notes in class	+				-									**1**	**1**
		problems in class					-							-		**2**	**0**
		works in groups in class														**0**	**0**
Block 4	School Climate	Stability of student body		+			+								+	**0**	**3**
		Index of major discipline problems										-				**1**	**0**
		Index of minor discipline problems		-											-	**2**	**0**
		Principal leadership	+													**0**	**1**
		Teacher's years of experience					+	+								**0**	**2**
		Perception of peer attitudes toward science	-					-								**2**	**0**

* results statistically significant at the .10 alpha level

Table 4.25 Indication of the direction of statistically significant results for fourth grade science – Model 5

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science	-	-												**2**	**0**
		how much students like science				+			+		+					**0**	**3**
		locus of control in science	+	+					+	+		+			+	**0**	**6**
Block 2	Instructional Methods	worksheets in class						-							+	**1**	**1**
		tests							-						-	**2**	**0**
		calculator use	-	-		-	-		-	-					-	**7**	**0**
		computer use		-							-	-		-		**4**	**0**
Block 3	Classroom Organization	notes in class	+								-					**1**	**1**
		problems in class					-							-		**2**	**0**
		works in groups in class														**0**	**0**
Block 4	School Climate	Stability of student body		+			+								+	**0**	**3**
		Index of major discipline problems										-				**1**	**0**
		Index of minor discipline problems		-											-	**2**	**0**
		Principal leadership	+													**0**	**1**
		Teacher's years of experience					+	+								**0**	**2**
		Perception of peer attitudes toward science	-					-								**2**	**0**
Block 5	School Structure	Urban location						+			+				-	**1**	**2**
		School size													-	**1**	**0**
		Class size		-					-					+		**2**	**1**

* results statistically significant at the .10 alpha level

Table 4.26 Indication of the direction of statistically significant results for fourth grade science – Model 6

	Explanatory Block	Independent Variables	Australia	Canada	Cyprus	Czech Rep	Greece	Iran	Ireland	Korea	Latvia (LSS)	New Zealand	Portugal	Slovenia	United States	Total Significant (-)	Total Significant (+)
Block 1	Student Involvement	time spent studying science	-	-												**2**	**0**
		how much students like science				+			+		+					**0**	**3**
		locus of control in science	+													**0**	**1**
Block 2	Instructional Methods	worksheets in class										-				**1**	**0**
		tests							-						-	**2**	**0**
		calculator use					-		-	-					-	**4**	**0**
		computer use		-							-	-		-		**4**	**0**
Block 3	Classroom Organization	notes in class	+								-	-				**2**	**1**
		problems in class					-							-		**2**	**0**
		works in groups in class														**0**	**0**
Block 4	School Climate	Stability of student body					+									**0**	**1**
		Index of major discipline problems										-				**1**	**0**
		Index of minor discipline problems													-	**1**	**0**
		Principal leadership	+				-									**1**	**1**
		Teacher's years of experience					+	+								**0**	**2**
		Perception of peer attitudes toward science	-					-								**2**	**0**
Block 5	School Structure	Urban location									+					**0**	**1**
		School size					-								-	**2**	**0**
		Class size		-					-							**2**	**0**
Block 6	Mean SES	Mean Socio-economic status of the school	+	+	+	+	+	+		+		+		+	+	**0**	**10**

* results statistically significant at the .10 alpha level

Tables 4.21 – 4.26 show that although the direction of significant effects did vary on some variables, in general, the direction of significant effects were stable across different cultural contexts.

After correcting for the differences in student intake across schools with regard to SES, how much variance in science achievement across schools can be explained by each explanatory model?

When evaluating the explanatory blocks as a whole, the best indication of their explanatory power is the R^{2*}, the explained variance. Recall that each model is cumulative and thus Model 2 includes all variables from the *Student Involvement* block as well as the variables from the *Instructional Methods* block. Table 4.27 allows the explanatory power of each model to be compared both across models and within countries. Model 1 ranged widely in explanatory power from a low of 5% in Iran to a high of 43% in Korea. Model 5 ranged from a low of 10% in Portugal to a high of 73% in the United States. A low amount of explained variance indicates that the explanatory model was, for the most part, incorrectly specified for the data. A high percentage of explained variance means that the model was correctly specified.

Table 4.27 Between-school variance in science achievement explained by each explanatory model in fourth grade science

	Explained Variance						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Country	Student Involvement	Model 1 with Instructional Methods	Model 2 with Classroom Organization	Model 3 with School Climate	Model 4 with School Structure	Model 5 with Mean SES	SES Model
Australia	27%	24%	27%	44%	41%	46%	21%
Canada	36%	43%	42%	53%	56%	64%	47%
Cyprus	17%	26%	24%	23%	24%	30%	23%
Czech Republic	15%	19%	23%	20%	19%	23%	22%
Greece	16%	34%	44%	51%	50%	63%	30%
Iran	5%	16%	16%	38%	46%	58%	49%
Ireland	23%	45%	46%	46%	47%	48%	19%
Korea	43%	48%	49%	47%	47%	58%	55%
Latvia (LSS)	20%	35%	34%	34%	36%	35%	1%
New Zealand	30%	50%	52%	59%	57%	69%	58%
**Portugal	27%	27%	38%	23%	10%	0%	35%
Slovenia	20%	25%	30%	24%	27%	44%	39%
United States	29%	49%	49%	66%	73%	83%	69%

SUMMARY OF FINDINGS IN MATHEMATICS

The results presented in response to the first research question reveal that the fourteen countries included in the mathematics portion of this study did in fact demonstrate sufficient between-school variability in mathematics achievement between schools to justify the development of theoretical models for explaining that variability. Although the variability between schools ranged from 16% in Cyprus to 56% in Latvia, the results for most countries showed that about one-quarter of the variability in mathematics achievement could be explained by differences between schools.

The second research question addressed the identification of policy malleable variables associated with school effectiveness after adjusting for differences in student intake. In that regard, student locus of control in mathematics was identified as the most pervasive variable associated with school effectiveness in mathematics. This finding was consistent across all six explanatory models, and the effect was found to have a positive association in most countries. In addition to the importance of students' locus of control, the findings suggest that frequency of computer use in mathematics class is negatively associated with school effectiveness in mathematics in Australia, Canada, Ireland, Latvia, and New Zealand. Additionally, class size was another variable having a significant association with school effectiveness in a number of countries across models. Class size was negatively related to school effectiveness in mathematics in Canada, Korea, Portugal, and the United States. Finally, even after controlling for SES at the student level and adding in five full blocks of explanatory variables, the means socioeconomic status of the school was significantly positively related to school effectiveness in mathematics in 12 of 14 countries.

The third research question asked how stable the statistically significant findings are across various cultural contexts. The finding that student locus of control in mathematics was associated with school effectiveness at fourth grade was found across all 14 countries in the complete model (Model 6). Furthermore, the relationship was positive across countries and even across explanatory models strongly suggests that this finding is truly persistent across cultural contexts[5]. Other variables, such as teachers years of experience did have a positive relationship in most countries, however, in the United States the relationship was negative.

---

[5] The exceptions came in Greece and Latvia in model 6. Recall, however, that the variance structure in the restricted model in Greece and Portugal was quite different than the full unconditional model indicating that care should be exercised in interpreting their results.

The fourth research question addressed by this dissertation study dealt with the amount of between-school variance in mathematics achievement across schools explained by factors related to: student involvement, instructional methods classroom organization, school climate, and school structural features. In attempting to gauge the explanatory power of each of these blocks, one way to make the comparison is to contrast explanatory powers of best of Models 1-5 with the model for SES alone in each country[6]. In doing this, Ireland recorded the biggest jump in explained variance with the SES model accounting for 24% of between-school variance in mathematics achievement, and Model 5 accounting for 56% of the between-school variance, indicating that Model 5 (which included all explanatory blocks except school mean SES) explained 33% more variability in student mathematics achievement than the SES model alone. By contrast, however, none of the explanatory blocks in Cyprus, the Czech Republic, Greece, Iran, or New Zealand were able to explain as much of the variability in mathematics achievement as the SES model alone. In general, the more variables included in the model, the more variance in between-school achievement was explained.

SUMMARY OF FINDINGS IN SCIENCE

The results presented in this chapter reveal that the thirteen countries included in the science portion of this study did in fact demonstrate sufficient variability in between-school science achievement to justify the development of theoretical models for

[6] Model 6 was not included in this comparison because mean school SES is not considered to be a policy relevant variable. It was included at the end of Model 5 in order to explain any additional variance in

explaining that variability. Although the variability between schools ranged from 12% in Korea to 59% in Latvia, in most countries, about one-quarter of the variability in science achievement could be explained by differences between schools.

In response to the second research hypothesis, the most important finding of the study was the persistent association of student locus of control in science with school effectiveness in science, even after adjusting for differences in student backgrounds across schools. Although the results in science were not as pervasive as in mathematics, student locus of control in science was significantly positively related to school effectiveness in several countries across the first five explanatory models. When mean SES of the school was included, however, the impact of this variable dropped. In addition to students' locus of control, the findings suggest that both the frequency of calculator use in science class and the frequency of computer use in science class are negatively associated with school effectiveness in science in a number of countries across models. In particular, the results from Greece, Ireland, Korea, and the United States indicate that calculator use in science class was negatively associated with school effectiveness in science. Interestingly, the results from a different set of countries (Canada, Latvia, New Zealand, and Slovenia) indicated that computer use in science class was negatively associated with school effectiveness. Finally, even after controlling for SES at the student level and adding in five full blocks of explanatory variables, the means socioeconomic status of the school was significantly positively related to school effectiveness in 10 of 13 countries.

---

achievement and to facilitate comparisons with the country.

The third research question asked how stable the statistically significant findings are across various cultural contexts. The finding that student locus of control in science was associated with school effectiveness in science at the fourth grade was found across 6-8 of 13 countries in the Models 1-5. The fact that the relationship was positive across all of these countries and even across explanatory models strongly suggests that this finding is truly prevalent across cultural contexts. Another consistent finding across explanatory models had to do with the frequency with which students were asked to solve problems during science class. The frequency with students are asked to solve problems during class was negatively associated with school effectiveness in science across all models in Greece and Slovenia.

The fourth research question asked in this dissertation study was related to the amount of variance in science achievement explained by factors associated with student involvement, instructional methods classroom organization, school climate, and school structural features. As was done in mathematics, the explanatory power from the best of Models 1-5 was compared to the model for SES alone in each country. In doing this, Latvia recorded the biggest jump in explained variance from 1% in the SES model to 36% in Model 5, indicating that Model explained 35% more variability in science achievement than the SES model alone. By contrast, however, none of the explanatory models developed for Iran, Korea, or Slovenia were able to explain as much as the SES model alone. As in mathematics, the general trend was that the more variables included in the model, the more variance in between-school achievement was explained.

COMPARISON OF FINDINGS IN MATHEMATICS AND SCIENCE

The final research question in this study dealt with the extent to which important explanatory variables were consistent across subject areas. The locus of control in the subject area was an important explanatory variable in both mathematics and science, however, the variable was more consistent across countries and models in mathematics. The instructional methods variables that were important tended to be slightly more consistent across countries in mathematics than in science, however, the variable related to frequency of computer use was negatively related to school effectiveness in about 1/3 of the countries in both mathematics and science. In addition, the variable proved to be important across subjects for the same countries (i.e., it was significantly negatively related to school effectiveness in mathematics and science in Canada and New Zealand). Finally, the class size variable proved to be significant in several countries across models in mathematics, but fewer in science.

SUMMARY

This chapter has provided the results aimed at answering each of four research questions in the subject areas of mathematics and science. The chapter began with a brief presentation and description of the descriptive statistics tables created for each country. The results were then presented in two separate sections, first for mathematics and then for science. The five areas explored by the study dealt with the amount of between-school variance in achievement, the extent to which significant predictors were identified within models, the extent to which those predictors were stable across different cultural contexts, the amount of explained variance contributed by each of six exploratory models, and the stability of explanatory variables across subject areas.

CHAPTER 5– CONCLUSIONS

OVERVIEW

This chapter will summarize and discuss the results of this dissertation study in relation to each of the four research questions. Each of the research questions will be presented and discussed in relation to the specific subject area under investigation (mathematics or science). This chapter will begin with a summary of the findings in mathematics and will then proceed to summarize the findings for science. The discussion will then focus on how the results may be applied in practice. Finally, the limitations of the study will be presented and directions for future research suggested.

SUMMARY AND INTERPRETATION OF THE FINDINGS

This study has revealed a number of important findings, some of which support prior research, and some of which do not. Despite the fact that this study employed the sophisticated analytic technique of Hierarchical Linear Modeling, the reader should bear in mind that the results of this study are still fundamentally associative in nature and therefore, causal inferences cannot be supported by the data. Nevertheless, it is clear that some consistent patterns have emerged that warrant further exploration.

The results presented in response to the first research question reveal that the fourteen countries included in the mathematics portion of this study, and the thirteen countries included in the science portion, did in fact demonstrate sufficient between-school variability in achievement to justify the development of theoretical models for explaining that variability. Although the variability between schools in mathematics ranged from 16% in Cyprus to 56% in Latvia, the results for most countries showed that about one-quarter of the variability in mathematics achievement could be explained by differences between schools. Similarly, the between-school variability in science achievement ranged from 12% in Korea to 56% in Latvia, although in most countries about one-quarter of the variability in science achievement was found to lie between schools. This finding is on par with results reported in many other studies of school effectiveness (Creemers et al., 1994; Sammons, 1999; Tymms, 1993; Willms, 1987b).

It should be noted that for the purposes of school effectiveness research, it is usually desirable to observe high levels of between-school variance because that indicates that there are some systematic differences occurring between schools such that students in some schools are performing better than students in other schools. From a policy standpoint, however, it is often desirable to observe low between-school variance because that may indicate that the school system is structured in a fairly equitable manner. In other words, student differences in achievement can be seen as independent of the school they attend. In the present investigation, lower levels of between-school variance were observed in countries with a more nationalized curriculum such as Korea (21%), while countries without a national curriculum, such as the United States (38%) had more of the variance in mathematics achievement attributable to differences between-schools.

The second research question addressed the identification of policy malleable variables associated with school effectiveness. In response to this question, the major finding of this study is that effective schools tend to have students who feel higher levels of internal locus of control over their environment with regard to the subject matter. In other words, effective schools tend to have a student body who is able to see a connection between hard work, ability, and achievement and are less likely to attribute achievement to factors such as luck. This finding was prevalent across the vast majority of countries in the study and across both subject areas under investigation. When it was statistically significant, high levels of internal locus of control were overwhelmingly found to be positively associated with school effectiveness. That locus of control is an important predictor of achievement is substantiated by prior literature on the topic (Dweck, 1986; Fenn & Iwanicki, 1983; Steele, 1997; Sterbin & Rakow, 1996). It is also important to note that this variable is associated with effectiveness even after taking into account differences in student socioeconomic status so that locus of control is accounting for unique variance beyond that which is accounted for by student SES.

Other findings from the *Student Involvement* block are not strongly supported across countries. The results findings do begin to corroborate findings by Martin et al. (2000), however, in suggesting that where significant relationships exist between the amount of time spent studying the subject matter and school effectiveness, the direction of the relationship is usually negative. Furthermore, the extent to which students have a positive attitude toward the subject matter tends to be positively associated with achievement, when significant relationships do exist.

With regard to the exploration of *Instructional Methods*, the results from this study were quite revealing. Given the increasing emphasis on technology in the classroom, it is interesting and important to note that in this study, the only associations between the use of technology and school effectiveness were negative for both computer use and calculator use in the classroom, particularly in the area of science. This finding, however, is consistent with a study of school effectiveness conducted in the United States that found that the use of calculators in the fourth grade classrooms was associated with lower mathematics achievement (Arnold & Sedlacek, 1995). Given the associative nature of this study, however, it is not possible to make any causal attributions to these results. It may in fact be the case that schools that are failing are targeted as needing special resources and therefore are being exposed to special curricula integrating technology in the classroom. It may in fact be the case that those schools are improving, but are still not up to par with other schools in the country. On the other hand, it may be the case that the use of technology detracts from real instruction and therefore is associated with lower achievement. In either case, it will take further research to understand the mechanism underlying this relationship.

The results from the exploration of *Classroom Organization* variables were generally inconsistent across countries and subject areas. It was the case, however, that where a significant relationship between achievement and the frequency with which students took notes in class existed, the relationship was negative. Despite the emphasis on group work that has emerged in the educational literature, this variable was almost never significantly related to school effectiveness in any model across either subject.

With respect to *School Climate* variables, students' perceptions of peers' attitudes showed a consistent pattern across subject areas in Australia and Korea. The variable was negatively associated with school effectiveness in Australia and positively associated with school effectiveness in Korea, perhaps reflecting the different value systems surrounding educational achievement in these two countries. Finally, the teacher's years of experience variable showed modest patterns of significance across models and across subject areas. This study found that the relationship between school effectiveness (in both mathematics and science) and teachers years of experience tended to be significantly positively related to school effectiveness in Greece and Iran so that the more years of teaching experience the teachers in the school had, the higher the average achievement of the student body in those schools.

The most consistent finding from the *School Structure* set of variables was the relationship between class size and school effectiveness. Class size was significantly negatively related to school effectiveness in mathematics in Canada, Korea, Portugal, and the United States. A negative relationship between achievement and class size has been noted in the United States for some time (Mosteller, 1995; U.S. Department of Education, 1998) and that finding has led to various policy initiatives throughout the country to reduce class size, most notably in Tennessee and California. The results of this analysis corroborate the relationship between class size and achievement in the United States, however, it will take experimental research to determine the underlying relationship between class size and achievement. Yet, as Mayer et al. (2000) aptly note, "Large-scale efforts to reduce class size may result in negative consequences if, as was the case recently in California, large numbers of unqualified teachers are hired because there are not enough qualified teachers available to staff the smaller classes (Bohrnstedt & Stecher, 1999).” (p. v).

Additionally, after controlling for all other variables in each of the previous blocks, the final model included a mean school socioeconomic status variable. Even after all of the shared variance had been attributed to the previous variables, the mean SES variable was statistically significant in 12 countries in mathematics and 10 countries in science.

The third research question asked how stable the statistically significant findings are across various cultural contexts. The finding that student locus of control in mathematics was found in anywhere from 8 to 14 countries across models while student locus of control in science was associated with school effectiveness in science at grade four across 6 to 8 of 13 countries in the Models 1-5. The fact that the relationship was positive across all of these countries and even across explanatory models strongly suggests that this finding is truly prevalent across cultural contexts. Additionally, the results regarding educational technology (i.e., frequency of computer use and frequency of calculator use) were found to be stable across cultural contexts and were negatively associated with school effectiveness in both mathematics and science.

The fourth research question asked in this dissertation study was related to the amount of variance in achievement explained by factors associated with student involvement, instructional methods classroom organization, school climate, and school structural features. To evaluate this question, the best explanatory model in each country was compared to the SES alone model. In mathematics, Ireland recorded the biggest jump in explained variance with the SES model while Latvia recorded the biggest jump in explained variance in science. Each of these findings suggests that the variables included in the explanatory blocks were indeed strongly related to achievement in those countries. By contrast, however, none of the explanatory blocks in Cyprus, the Czech Republic, Greece, Iran, or New Zealand were able to explain as much of the variability in mathematics achievement as the SES model alone. In science, none of the explanatory models developed for Iran, Korea, or Slovenia were able to explain as much as the SES model alone. In fact, in some models, the error variance that was introduced through an expanded model actually served to decrease the amount of explained variance for models with more variables. This finding suggests that the explanatory models that were constructed in this study did not do a very good job of explaining variability between-schools in those countries. Yet, it is not entirely surprising given that most of the literature on school effectiveness was drawn from research conducted in western Europe and the United States.

Finally, the fifth research question asked how stable the findings were across subject areas. The results of this study show that although the findings were often more pervasive in mathematics, the relationships that were found between the explanatory variables and school effectiveness were most often stable across subject areas. For example, frequency of calculator use and the frequency of computer use were consistently negatively associated with school effectiveness in both mathematics and science while students locus of control in the subject area was consistently positively associated with school effectiveness in both mathematics and science.

PRACTICAL APPLICATION OF RESULTS

The most suggestive and pervasive findings relate to students' perceived locus of control in the subject area. Although this study is associative in nature, the absolute dominance of this finding strongly suggests that this is a policy variable that should be studied further, perhaps even in an experimental setting. In some ways, this finding corroborates the work done by Steele (1997) on stereotype threat. Steele's research looked at gender differences in achievement and asked females to participate in a study with three conditions. In the first condition, females were given a mathematics test and not told anything about the test (control group). In the second condition, females were given a mathematics test and told that although males tended to do better on most mathematics tests, that this particular test had been piloted and did not exhibit gender differences. In the third conditions, females were given a mathematics test and reminded that males typically outperform females in mathematics. The findings revealed that females in the experimental group that was told the particular test was gender neutral outperformed all others. Although Steele does not frame it in this way, the findings are suggestive of a positive relationship between students' internal locus of control and their subsequent academic achievement. This is so because females in the gender neutral experimental group came to believe that differences in achievement were not due to differences in natural ability or luck, but that achievement was in fact attainable.

Similarly, the call for high expectations in schools also speaks directly to the concept of students' locus of control. The movement toward high standards and expectations is predicated on the belief that all students can achieve, a belief that, when internalized, gives students a high degree of locus of control. The present study suggests that perhaps schools may become more effective by clearly articulating their belief in a causal association between hard work and achievement. Furthermore, schools support the connection between students beliefs in their own ability, hard work, and achievement will articulate a message that endows students with higher levels of internal locus of control. In this way, books such as “The Bell Curve” (Herrnstein & Murray, 1994) that promote the idea that differences in achievement are mostly attributable to native ability by race, when internalized by students, could in fact have a deleterious impact upon minority student achievement.

Benham (1995) has explored various strategies that may be used to foster self-motivated behavior, personal responsibility, and internal locus of control in the school setting. In particular, “…school activities in which children are encouraged to exert personal control and self responsibility for learning tasks with clearly identified goals, and which provide ample opportunity to infer ability from success, are beneficial for helping to foster internally locused causal attributions.” (Benham, 1995, p. 31). Yet Benham also notes that self-responsibility for success and failure may be learned separately, and that elementary school children may assume more responsibility for one than the other (1995). Consequently, teachers should provide instructional exercises that demonstrate the relationship between behaviors and both good and bad outcomes.

In addition, Miller (1980) found that students who are praised for specific accomplishments are likely to attribute the praise to their own ability and effort. By contrast, however, making ambivalent distinctions between correct and incorrect answers or positively reinforcing students for incorrect answers encouraged students to attribute such praise to forces outside of their control (Miller, 1980). Thus, in order to foster higher levels of internal locus of control, teachers should praise specific accomplishments while making clear distinctions between acceptable and unacceptable outcomes.

Finally, Dweck (1975) showed that students who had been identified as helpless could be taught to attribute their failure to lack of effort, and subsequently improve their performance through instructional strategies fostering the development of persistence. “He suggested that an instructional program for children who are at risk for failure should not gloss over students’ errors, but instead, should include procedures for training the students to deal with their errors directly, and view errors as vehicles for teaching children increased responsibility for their own behavior.” (Benham, 1995, p. 18).

The other major finding of this study relates to the use of technology and its association with school effectiveness. The results of this study suggest that despite the push in educational policy to integrate technology into the classroom, the relationships that appear (especially in science) tend to be negatively associated with achievement such that the more frequently students use computers or calculators in the classroom, the less effective the school is in that particular subject area.

Some caution should be exercised in interpreting these results, however. Wenglinsky (1998) has pointed out that the association of computer use with academic achievement depends more heavily upon how computers are being used in the classroom than how frequently they are being used. In particular, his findings from NAEP reveal that students who used computers in school for drill and skill exercises exhibited test scores that were negatively association with achievement, while students using computers for higher-order operations had test scores that were positively associated with achievement (Wenglinsky, 1998). Consequently, the results of this study suggest that schools in Australia, Canada, Latvia, Iran, Slovenia, and New Zealand should investigate exactly how computers are being used in their most and least effective schools.

Furthermore, given that the nature of this study is associative, we cannot be sure which way the causal arrow (if there is one) between frequency of computer use and school effectiveness points. It could be the case that because schools have low achievement, policies have been put into place that attempt to integrate technology to raise achievement. When viewed in light of the small, but persistent finding that more effective schools tend to have more students who spend time working on worksheets in class, the advisability of integrating technology into the fourth grade classroom clearly warrants further study.

DIRECTIONS FOR FUTURE RESEARCH

The first suggestion for future research is to continue exploration of the relationship between student locus of control and school effectiveness. The next step in this exploration is to set up some experimental studies aimed at empirically testing the relationship between locus of control and achievement. Some research suggests that internal locus of control takes several years to develop or to alter (Auer, 1992), while recent findings by Steele (1997) intimate that perhaps the construct is less stable than previously thought. The state or trait nature of this variable also should be explored by future research. At a minimum, more research should be done on the variables associated with locus of control so that, if necessary, more information about this construct can be integrated into future IEA studies of student achievement (such as TIMSS 2003).

In addition, with the release of the results for TIMSS 1999, it would be interesting to explore the models of school effectiveness constructed in this study and compare them across time, particularly in light of the fact that the population of students in the eighth grade in TIMSS 1999 are the same cohort of students who were in the fourth grade in TIMSS and represent the population of students investigated by this dissertation study.

A third suggestion for future research is to attempt to explain the variance in achievement not only at the school level, but at the classroom level as well. With the advent of the pseudo-classroom procedure (O'Dwyer, 2000), it would be possible to empirically test a three-level hierarchical model where students represent the first level, classrooms represent the second level, and schools represent the third level. Given that school effects are of interest, it would be interesting to compare the results of a three-level model using the pseudo-classroom procedure to the results of a study that simply uses grade level as a control variable when analyzing an entire population of students (i.e., both upper and lower grades).

Finally, several practical suggestions for the development of the TIMSS 2003 questionnaires follow. With regard to the area of educational technology, future studies should attempt to gather information about the level of cognitive demand required by students when using computers in the classroom (Mayer et al., 2000). Next, while it is difficult to adequately capture some dimensions of school climate (such as culture) on questionnaires, future studies should consider incorporating questions relating to areas such as: teacher expectations of student achievement; the use of punishments and rewards within the school; and the extent to which students, teachers, and administrators are clear about the goals of the school (Anderson, 1982). Finally, in the area of student discipline, Mayer et al. (2000) state that student achievement is linked to student's perceptions of the fairness and effectiveness of various discipline policies. Consequently, it is important to gather information that relates not only to the occurrence of particular discipline problems, but also to collect information on the way students perceive the resolution of those problems within their school.

LIMITATIONS OF THE STUDY

All research invariably involves some benefits and some limitations, and this study is no exception. Perhaps most importantly, this study is limited by its associative nature. Even after using the sophisticated analytic technique of HLM, the results of this study are fundamentally associative in nature and therefore, cannot support causal inferences.

Next, although this study is exploring school effectiveness, it should be kept in mind that this study defines effectiveness in terms of test scores on TIMSS. There are arguably many more dimensions to school effectiveness than test scores can possibly reflect. "The American public often assumes that countries whose students score high on international comparisons of educational achievement are pleased with their existing programs." (Atkin & Black, 1997, p. 23). Just because certain schools have higher average achievement does not necessarily imply that they are content with their educational programs.

A third limitation of this study is that the between-school variance in achievement is necessarily confounded with the between-classroom variance in achievement due to the fact that TIMSS only sampled one classroom per school. Given that the topic under investigation is school effectiveness, this limitation is really more of a technicality. Yet, in light of prior research that shows that school effects are in fact more often really classroom effects, this question cannot truly be explored without the use of special procedures (see O'Dwyer, 2000) and we cannot, from the results of the present study, truly know if the results are more important at the classroom level or the school level.

Finally, it turned out to be the case that across both subjects and across all explanatory models, a significant amount of parameter variance remained to be explained in every country. This implies that although some of the explanatory models of school effectiveness did account for a great deal of between-school variance in achievement, none of the models were perfectly specified. Part of the reason for this is due to the fact that this study is a secondary analysis. TIMSS was designed to serve many purposes and was not specifically tailored toward school effectiveness research. As a consequence, many of the background variables that might have been used to build a more accurately specified model were simply not included in the original study.

SUMMARY

The results of this study of school effectiveness in mathematics and science at the fourth grade have shown that fourteen countries possessed sufficient variability between schools in mathematics achievement to justify the creation of explanatory models of school effectiveness. Furthermore, thirteen countries possessed sufficient variability between schools in science achievement to justify the pursuit of explanatory models for school effectiveness in science.

The research findings presented in this dissertation study revealed that after adjusting for differences in student backgrounds between schools, schools with students who reported seeing a connection between hard work, belief in their own ability, and academic achievement tended to be the most effective across countries at the fourth grade on TIMSS. In addition, a negative association was found between school effectiveness and the use of computers and calculators in the classroom. Each of these findings was relatively stable across explanatory models, cultural contexts, and subject areas.

This study has contributed a unique element to the literature by examining school effectiveness at the fourth grade across two domains and across fourteen different countries. Continued work in the area of school effectiveness is necessary, perhaps using trend data and looking at different grade levels. In addition, it is recommended that the impact of the locus of control variable upon achievement receive serious consideration in an experimental setting.

REFERENCES

Adams, R.J., & Gonzalez, E.G. (1996). The TIMSS Test Design. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume II: Implementation and Analysis at the Primary and Middle School Years (pp. 3-1–3-36). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Aiken, L.R. (1970). Attitudes toward mathematics. Review of Educational Research, 40, pp. 551-596.

Aitkin, M. A., & Longford, N. T. (1986). Statistical modeling issues in school effectiveness studies. Journal of the Royal Statistical Society A, 149, pp. 1-43. (with commentary).

Alexander, K.L., & Pallas, A.M. (1983). Private Schools and Public Policy: New Evidence on Cognitive Achievement in Public and Private Schools. Sociology of Education, 56, pp. 170-182.

Anderson, C.S. (1982). The Search for School Climate: A review of the research. Review of Educational Research, 52, 3, pp. 368-420.

Arnold, C.L., Kaufman, & Sedlacek, D. (1992). School Effects on Educational Achievement in Mathematics and Science: 1985-1986. Washington, D.C.: U.S. Department of Education, National Center for Educational Statistics.

Arnold, C.L. & Sedlacek, D.A. (1995). Using HLM and NAEP Data to Explore School Correlates of 1990 Mathematics and Geometry Achievement in Grades 4, 8, and 12: Methodology and Results. Washington, D.C.: National Center for Educational Statistics.

Atkin, J. M., & Black, P. (1997). Policy perils of international comparisons. Phi Delta Kappan, 79, pp. 22-28.

Auer, C.J. (1992). A comparison of the locus of control of first- and second-grade students in whole language, basal reader, and eclectic instructional approach classrooms (Doctoral dissertation, Northern Illinois University, 1992). Dissertation Abstracts International, 53, 11, 3856.

Barton, P.E., Colye, R.J., & Wenglinsky, H. (1998). Order in the Classroom: Violence, Discipline, and Student Achievement. Policy Information Center. Princeton, NJ: Educational Testing Service.

Beaton, A., & Gonzalez, E. (1997). TIMSS Test-Curriculum Matching Analysis. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume II: Implementation and Analysis at the Primary and Middle School Years (pp. 187–194). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Beaton, A., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.J., & Smith, T.A. (1996). Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Benham, J.M. (1995). Fostering Self-Motivated Behavior, Personal Responsibility, and Internal Locus of Control in the School Setting. ERIC ED 386 621: University of Southern Maine.

Bennett, S.N. (1976). Teaching Styles and Pupil Progress. London: Open Books.

Bidwell, C.E., & Kasarda, J.D. (1980). Conceptualizing and Measuring the Effects of School and Schooling, American Journal of Education , 88, 4, pp. 401-430.

Bierhoff, H. (1996). Laying the Foundation of Numeracy: A Comparison of Primary School Textbooks in Britain, Germany and Switzerland. London: National Institute for Economic and Social Research.

Bohrnstedt, G.W., & Stecher, B.M. (1999). Class Size Reduction in California: Early Evaluation Findings, 1996–1998 (CSR Research Consortium, Year 1 Evaluation Report). Palo Alto, CA: American Institute for Research.

Bondi, L. (1991). Attainment in primary schools. British Educational Research Journal, 17, 3, pp. 203-217.

Boss, K., Kuiper, W., & Plomp, T. (1999). Student performance and curricular appropriateness in the Netherlands. Studies in Educational Evaluation, 25, 3, pp. 269-276.

Bryk, A. S., & Raudenbush, S. W. (1989). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R. D. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 159–204). San Diego: Academic Press.

Brookover, W.B., Schweitzer, J.G., Schneider, J.M., Beady, C.H., Flood, P.K., & Wisenbaker, J.M. (1978). Elementary School Social Climate and School Achievement. American Educational Research Journal, 15, pp. 301-318.

Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage Publications.

Bryk, A.S., Raudenbush, S.W., & Congdon, R.T. (1996). HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs. Chicago, IL: Scientific Software International.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. Review of Research in Education, 8, pp. 158-233.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R. (1966). Equality of Educational Opportunity. Washington, DC: U.S. Department of Health, Education, and Welfare, U.S. Government Printing Office.

Coleman, J.S., Hoffer, T., & Kilgore, S.B. (1982). High School Achievement: Public, Catholic, and Other Private Schools Compared. New York: Basic.

Cook, T.D., & Campbell, D.T. (1979). Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston, MA: Houghton Mifflin Corp.

Creemers, B.P.M., & Osinga, N. (Eds.). (1995). ICSEI Country Reports. Leeuwarden, Netherlands: ICSEI Secretariat.

Creemers, B. P. M., Reynolds, D., & Swint, F. E. (1994). The International School Effectiveness Research Project Quantitative and Class Study Data.

Creemers, B.P.M., & Scheerens, J. (Eds.). (1989). Developments in School Effectiveness Research, A special issue of International Journal of Educational Research, 13, 7, pp. 685-825.

Cronbach, L.J. (1976). Research on classrooms and schools: Formulation of questions, design, and analysis. Occasional paper: Stanford Evaluation Consortium.

Cronbach, L.J., & Webb, N. (1975). Between and Within-Class Effects in a Reported Aptitude-by-Treatment Interaction: Reanalysis of a Study by G.L. Anderson. Journal of Educational Psychology, 6, pp. 717-724.

Cuttance, P. (1986). Effective Schooling: Report to the Scottish Education Department. Edinburgh: Centre for Educational Sociology, University of Edinburgh.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. Education Policy Analysis Archives, 8, 1. Available: http://olam.ed.asu.edu/epaa/v8n1/.

Davie, R., Butler, N. & Goldstein, H. (1972). From Birth to Seven. London: Longman.

Douglas, J.W.B. (1964). The Home and the School: The Study of Ability and Attainment in the Primary School. London: MacGibbon & Kee.

Dweck, C.S. (1975). The role of expectations and atttributions in the alleviation of learned helplessness. Journal of Personality and Social Psychology, 31, pp. 647-685.

Dweck, C.S. (1986). Motivational Processes Affecting Learning. American Psychologist, 41, pp. 1040-1048.

Edmonds, R.R. (1979a). Effective Schools for the Urban Poor, Educational Leadership, 37, 10, pp. 15-24.

Edmonds, R.R. (1979b). Some schools work and more can, Social Policy, 9, 2, pp. 28-32.

Essen, J. & Wedge, P. (1982). Continuities in Childhood Disadvantage. London: Heinemann.

Fenn, L.M., & Iwanicki, E.F. (1983, April). An Investigation of the Relationship between Student Affective Characteristics and Student Achievement within More and Less Effective School Settings. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Quebec.

Fitz-Gibbon, C., & Kochan, S. (2000). School Effectiveness and Education Indicators. In C. Teddlie & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 257–282). London: Falmer Press.

Foy, P., Rust, K., & Schleicher, A. (1996). Sample Design. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume I: Design and Development (pp. 4-1–4-17). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Foy, P. (1997). Calculation of Sampling Weights. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume II: Implementation and Analysis at the Primary and Middle School Years (pp. 71–80). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Frase, M. (1997). Pursuing Excellence: A Study of U.S. Fourth-Grade Mathematics and Science Achievement in International Context. Initial Findings from the Third International Mathematics and Science Study. Office of Educational Research and Improvement (ED), Washington, D.C. NCES 97255.

Fuller, B. (1987). School effects in the Third World. Review of Educational Research, 57, 3, pp. 255-92.

Goldstein, H. (1995). Multilevel Statistical Models (2nd Edition) London: Edward Arnold & New York: Halsted Press.

Goldstein, H., Rabsbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., & Thomas, S. (1992). Multlilevel models for comparing schools. Multilevel Modeling Newsletter 4, 2, pp. 5-6. London: Department of Mathematics, Statistics & Computing, Institute of Education, University of London.

Gonzalez, E.J. (1997). Reporting Student Achievement in Mathematics and Science. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume II: Implementation and Analysis in the Primary and Middle School Years (pp. 147–174). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Gonzalez, E.J., & Foy, P. (1997). Estimation of Sampling Variability, Design Effects, and Effective Sample Sizes. In M.O. Martin & D.L. Kelly (Eds.), Third International Mathematics and Science Study Technical Report Volume II: Implementation and Analysis in the Primary and Middle School Years (pp. 81–100). Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Gonzalez, E.J., & Smith, T.A. (1997). User Guide for the TIMSS International Database: Primary and Middle School Years. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Gray, J. (1989). Multilevel Models: Issues and Problems Emerging from their Recent Application in British Studies of School Effectiveness. In D. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 127–157). San Diego, CA: Academic Press Inc.

Gray, J., Jesson, D., & Jones, B. (1986). The search for a fairer way of comparing schools' examination results. Research Papers in Education, 1, 2, pp. 91-122.

Haney, W. (1980). Units and Levels of Analysis in Large-scale Evaluation. New Directions for Methodology of Social and Behavioral Science, 6, pp. 1-15.

Henderson, V., Mieszkowski, P., & Sauvageau, Y. (1978). Peer Group effects and Educational Production Functions. Journal of Public Economics, 10, pp. 97-106.

Herrnstein, R.J., & Murray, C. (1994). The Bell Curve: Intelligence and Class Structure in American Life. New York, NY: Free Press.

Hill, P.W., & Rowe, K.J. (1996). Multilevel modeling in school effectiveness research. School Effectiveness and School Improvement, 7, 1, pp. 1-34.

Hoyle, C. (2001). Effective Schools in Grade 8 Mathematics: A Secondary Analysis of the Third International Mathematics and Science Study. Unpublished Ph.D. dissertation, Boston College: Chestnut Hill, MA.

Husen, T. (1996). Lessons from the IEA studies. International Journal of Educational Research, pp. 207-218.

Jencks, C., Smith, M., Acland, H., Bane, J.J., Cohen, D. Gintis, H., Heyns, B. & Michelson, S. (1972) Inequality: A Reassessment of the Effects of Family and Schooling in America. New York: Basic Books.

Jesson, D., & Gray, J. (1991). Slants on slops: Using multi-level models to investigate differential school effectiveness and its impact on pupils' examination results. School Effectiveness and School Improvement, 2, 3, pp. 230-247.

Jones, R. (1998). Solving Problems in Math and Science Education. American School Board Journal, 185, 7, pp. 16-20.

Kilgore, S., & Pendleton, W.W. (1986). Discussion of the paper by Aitkin & Longford. Journal of the Royal Statistical Society, Series A, 149, 1, pp. 36-37.

Kish, L. (1965). Survey Sampling. New York: John Wiley & Sons.

Kreft, I., & De Leew, J. (1998). Introducing Multilevel Modeling. Thousand Oaks, CA: Sage Publications.

Lee, J. (1998, April). Missing Links in International Education Studies: Comparing the U.S. with East Asian Countries in TIMSS. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Lockheed, M.E., & Komenan, A. (1989). Teaching quality and student achievement in Africa: The case of Nigeria and Swaziland. Teaching and Teacher Education, 5, 2, pp. 93-113.

Ma, X., & Kishor, N. (1997). Assessing the Relationship Between Attitude Toward Mathematics and Achievement in Mathematics: A Meta-Analysis. Journal for Research on Mathematics Education, 28, pp. 27-47.

Madaus, G.F., Haney, W., & Kreitzer, A. (1992). Testing and Evaluation: Learning From the Projects We Fund. Council for Aid to Education.

Madaus, G.F., Kellaghan, T., Rakow, E.A., & King, D. (1979). The sensitivity of measures of school effectiveness. Harvard Educational Review, 49, pp. 207-230.

Madaus, G.F., Scriven, M.S., & Stufflebeam, D.L (Eds.). (1983). Evaluation Models: Viewpoints on educational and human services evaluation. Boston, MA: Kluwer-Nijhoff.

Maker, C.J. (1982). Teaching models in education of the gifted. Rockville, MD: Aspen Systems Corp.

Martin, M.O., & Kelly, D.L. (Eds.). (1996). Third International Mathematics and Science Study Technical Report Volume I: Design and Development. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Martin, M.O., Mullis, I.V.S., Beaton, A.E., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1997). Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Martin, M.O., Mullis, I.V.S., Gregory, K., Hoyle, C., & Shen, C. (2000). Effective Schools in Science and Mathematics. Chestnut Hill, MA: The International Study Center at Boston College.

Mayer, D.P., Mullens, J.E., & Moore, M.T. (2000). Monitoring School Quality: An Indicators Report. National Center for Educational Statistics (NCES 2001-030). Washington, D.C.: U.S. Department of Education.

McPartland, J.M., & McDill, E.L. (1982). Control and Differentiation in the Structure of American Education. Sociology of Education, 55, pp. 77-88.

McPherson, A. (1992). Measuring added value in schools, National Commission on Education Briefing, No 1. February 1992, London.

Miller, S.K. (1980). Raising low socio-economic/minority school achievement: Overcoming the effects of student sense of academic futility. Boston, MA: American Educational Research Association. (ERIC Document Reproduction Service No. ED 186 575).

Mortimore, P. (1993). School effectiveness and the management of effective learning and teaching, School Effectiveness and School Improvement, 4, 4, pp. 290-310.

Mortimore, P., & Blackstone, T. (1982). Disadvantage in Education. London: Heinemann.

Mortimore, P. Sammons, P. Stoll, L., Lewis, D. & Ecob, R. (1988a). School Matters: The Junior Years. Wells: Open Books.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988b). The effects of school membership on pupils educational outcomes. Research Papers in Education 3, 1, pp. 3-26.

Mosteller, F. (1995). The Tennessee Study of Class Size in the Early School Grades. Future of Children, 5, 2, pp.113-127.

Mullis, I. V. S., Jenkins, F., & Johnson, E. G. (1994). Effective Schools in Mathematics: Perspectives from the NAEP 1992 Assessment. Washington DC: OERI.

Mullis, I. V. S., Martin, M.O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. S. (1997). Science Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy.

Murnane, R.J., & Phillips, B.R. (1981). Learning by Doing, Vintage, and Selection: Three Pieces of the Puzzle Relating Teaching Experience and Teaching Performance. Economics of Education Review, 1, 4, pp. 453-465.

O'Dwyer, L. (2000). Extending the Application of Multilevel Modeling to Data from the Third International Mathematics and Science Study. Ph.D. dissertation, Boston College: Chestnut Hill, MA.

Page, E., & Keith, T. (1981). Effects of U.S. Private Schools: A Technical Analysis of Two Recent Claims. Educational Researcher, 10, pp. 7-17.

Pedhazur, E. J. (1997). Multiple Regression in Behavioral Research: Explanation and Prediction (3rd Ed.). Orlando, FL: Harcourt Brace & Company

Plewis, I.F. (1986). Discussion of the paper by Aitkin & Longford. Journal of the Royal Statistical Society, Series A, 149, 1, pp. 26-27.

Plowden Committee (1967). Children and Their Primary Schools. London: HMSO.

Ranson, S. (1997). For citizenship and the remaking of civil society. In R. Pring & G. Walford (Eds.), Affirming the Comprehensive Ideal (pp. 183 – 196). London: Falmer Press.

Raudenbush, S. W. & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, pp. 1-17.

Raudenbush, S.W., & Bryk, A. S. (1989). Quantitative Models for Estimating Teacher and School Effectiveness. In R.D. Bock (Ed.), Multilevel Analysis of Educational Data (pp. 205–234). San Diego, CA: Academic Press, Inc.

Raudenbush, S.W., Bryk, A., Cheong, Y.F., & Congdon, R. (2000). HLM5: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International.

Reynolds, D. (2000). School Effectiveness: The International Dimension. In C. Teddlie & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 232–256). London: Falmer Press.

Reynolds, D., Teddlie, C, Creemers, B., Scheerens, J., & Townsend, T. (2000). An Introduction to School Effectiveness Research. In C. Teddlie & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 3–25). London: Falmer Press.

Rotter, J.B. (1966). Generalized Expectancies for Internal Versus External Control of Reinforcement. Psychological Monographs, 80 (1, Whole No. 609).

Russell, M. (2000). It's Time to Upgrade: Tests and Administration Procedures for the New Millennium. Paper presented at the Secretary's Conference on Educational Technology. Available: http://www.ed.gov/Technology/techconf/2000/rullell_paper.html

Rutter, M. & Madge, N. (1976). Cycles of Disadvantage. London: Heinemann.

Rutter, M., Maughan, B., Mortimore, P. & Ouston, J. (1979). Fifteen Thousand Hours: Secondary Schools and Their Effects on Children. London: Open Books.

Sammons, P. (1999). School Effectiveness: Coming of Age in the Twenty-first Century. Swets & Zeitlinger B.V.: Lisse, The Netherlands.

Sammons, P., Nuttall, D. & Cuttance, P. (1993). Differential School Effectiveness: Results from a reanalysis of the inner London education authority's junior school project data, British Educational Research Journal 19, 4, pp. 381-405.

Sammons, P., Mortimore, P. & Thomas, S. (1993). First weigh your ingredients, The Independent. 20 November 1993.

Sammons, P., Thomas, S., & Mortimore, P. (1997). Forging Links: Effective Schools and Effective Departments. London: Paul Chapman.

Scheerens, J. (1992). Effective Schooling: Research, Theory, and Practice. London: Cassell.

Scheerens, J., Vermeulen, C. J., & Pelgrum, W. J. (1989). Generalizability of school and instructional effectiveness indicators across nations. In B.P.M. Creemers & J. Scheerens (Eds.), Developments in School Effectiveness Research, special issue of International Journal of Educational Research, 13, 7, pp. 789-99.

Slee, R., & Weiner, G. (1998). Introduction: School Effectiveness for Whom? In R. Slee & G. Weiner (Eds.), School Effectiveness for Whom?: Challenges to the School Effectiveness and School Improvement Movements (pp. 1–9). Portland, OR: Taylor & Francis, Inc.

Smith, D., Tomlinson, S., Tomes, H. & Bannerjea, L. (1984). Report to the D.E.S. on the project "Factors Associated with Success in Multi-Ethnic Secondary School."

Steedman, J. (1983). Examination Results in Selective and Non-Selective Schools. London: National Children's Bureau.

Steele, C. (1997). A Threat in the Air: How stereotypes shape intellectual identity and performance. American Psychologist, 52, pp. 613-629.

Sterbin, A., & Rakow, E. (1996, November). Self-Esteem, Locus of Control, and Student Achievement. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Tuscaloosa, Al.

Stufflebeam, D. (1983). The CIPP Model for Program Evaluation. In G.F. Madaus, M.S. Scriven, & D.L. Stufflebeam (Eds.), Evaluation Models: Viewpoints on educational and human services evaluation (pp. 117–142). Boston, MA: Kluwer-Nijhoff.

Tarr, J., Mittag, K., Uekawa, K., Lennex, L. (1999, April). A Comparison of Calculator Use in Eighth-Grade Mathematics Classrooms in the United States, Japan, and Portugal: Results from the Third International Mathematics and Science Study. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Tamir, P., & Zuzovsky, R. (1999). Performance of Israeli 4th and 8th grade students on science and combined science and math performance assessment tasks. Studies in Educational Evaluation, 25, 3, pp. 283-286.

Teddlie, C. & Reynolds, D. (2000). The International Handbook of School Effectiveness Research. New York, NY: Falmer Press.

Teddlie, C., Reynolds, D., & Pol, S. (2000). Current Topics and Approaches in School Effectiveness Research: The Contemporary Field. In C. Teddlie & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 26–52). London: Falmer Press.

Teddlie, C., Reynolds, D., & Sammons, P. (2000). The Methodology and Scientific Properties of School Effectiveness Research. In C. Teddlie & D. Reynolds (Eds.), The International Handbook of School Effectiveness Research (pp. 55–133). London: Falmer Press.

Tymms, P. (1993). Accountability: Can it be fair? Oxford Review of Education, 19, 3, pp. 291-299.

U.S. Department of Education. (1998). Reducing Class Size: What Do We Know? Washington, D.C.

Walberg, H.J. (1991). Improving School Science in Advanced and Developing Countries. Review of Educational Research, 61, 1, pp. 25-69.

Weiner, B. (1972). Theories of Motivation. Chicago, IL: Markham Publishing Co.

Welch, W. W., Walberg, H. J., & Fraser, B. J. (1986). Predicting Elementary Science Learning Using National Assessment Data. Journal of Research in Science Teaching, 23, 8, pp. 699-706.

Wenglinsky, H. (1998). Does It Compute? The Relationship Between Educational Technology and Student Achievement in Mathematics. Princeton, NJ: Educational Testing Service.

West, M., & Hopkins, D. (1996, April). Reconceptualising school effectiveness and school improvement. Paper presented at the School Effectiveness and Improvement Symposium of the Annual Conference of the American Educational Research Association, New York, NY.

White, J. & Barber, M. (Eds.). (1997). Perspectives on School Effectiveness and Improvement. London: Bedford Way Paper, Institute of Education, University of London.

Willms, J.D. (1986). Social class segregation and its relationship to pupils' examination results in Scotland. American Sociological Review, 51, pp. 224-241.

Willms, J.D. (1987a). Differences between Scottish education authorities in their examination attainment. Oxford Review of Education, 13, 2, pp. 211-232.

Willms, J. D. (1987b). Comparing schools on performance indicators: Two questions about school effects. Paper presented to the Council of Europe Educational Research Workshop, Honefoss, Edinburgh, Centre for Educational Sociology, University of Edinburgh.

Wilson, L., & Blank, R. (1999). Improving Mathematics Education Using Results from NAEP and TIMSS. Council of Chief State School Officers, Washington, D.C. Available: http://www.ccsso.org.

APPENDIX A - DESCRIPTIVE STATISTICS

Table A.1 Descriptive Statistics for Australia at the Upper Grade of Population 1

Australia	
N of Students =	6,507
N of Schools =	178

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-3.0	1.2	0.0	(0.62)	6326	3%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.1	1.4	0.0	(0.62)	6124	6%	S
	Time on mathematics homework	1.0	5.0	2.2	(0.76)	6243	4%	S
	Likes mathematics	-4.0	-1.0	-1.8	(0.88)	6138	6%	S
Science	Locus of control in science	-2.8	1.7	0.0	(0.61)	6103	6%	S
	Time on science homework	1.0	5.0	1.6	(0.80)	6185	5%	S
	Likes science	-4.0	-1.0	-1.9	(0.88)	6085	6%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.6	(0.61)	6080	7%	S
	Frequency of testing	-3.0	-1.0	-1.8	(0.53)	6076	7%	S
	Frequency of calculator use	-3.0	-1.0	-2.2	(0.55)	6062	7%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.63)	5989	8%	S
Science	Frequency of workbook homework	-3.0	-1.0	-2.0	(0.72)	5952	9%	S
	Frequency of testing	-3.0	-1.0	-2.2	(0.66)	5962	8%	S
	Frequency of calculator use	-3.0	-1.0	-2.6	(0.61)	5946	9%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.61)	5912	9%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.51)	6115	6%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.8	(0.61)	6099	6%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.49)	6037	7%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.7	(0.69)	6003	8%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.65)	5978	8%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.9	(0.60)	5923	9%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.43)	6050	7%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.47)	6021	7%	S
General	Percentage of students who begin and finish the year in the same school	60.0	100.0	91.9	(8.03)	148	17%	C
	Index of minor behavior problems	-0.1	0.5	0.0	(0.08)	113	37%	C
	Index of major behavior problems	-0.1	0.8	0.0	(0.10)	112	37%	C
	Hours per month principal teaches	0.0	97.0	26.3	(33.03)	147	17%	C
	Teachers years of experience	1.0	37.0	15.0	(8.45)	266	19%	T
School Structure								
General	Urban location	0.0	1.0	0.3	(0.44)	156	12%	C
	School size	58.0	1967.0	291.2	(215.34)	140	21%	C
	Average class size	2.0	35.0	22.9	(7.79)	148	17%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.2 Descriptive Statistics for Canada at the Upper Grade of Population 1

Canada	
N of Students =	8,408
N of Schools =	390

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-2.2	1.5	0.0	(0.56)	8342	1%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.4	1.3	0.0	(0.58)	8242	2%	S
	Time on mathematics homework	1.0	5.0	2.2	(0.78)	7977	5%	S
	Likes mathematics	-4.0	-1.0	-1.7	(0.81)	8120	3%	S
Science	Locus of control in science	-3.2	1.4	0.0	(0.59)	8240	2%	S
	Time on science homework	1.0	5.0	1.8	(0.83)	7989	5%	S
	Likes science	-4.0	-1.0	-1.9	(0.87)	8073	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.4	(0.55)	8044	4%	S
	Frequency of testing	-3.0	-1.0	-1.7	(0.51)	8074	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.4	(0.60)	8038	4%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.64)	8017	5%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.8	(0.73)	7947	5%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.60)	7986	5%	S
	Frequency of calculator use	-3.0	-1.0	-2.8	(0.51)	7952	5%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.62)	7869	6%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.52)	8141	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.66)	8095	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.56)	8083	4%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.7	(0.68)	8026	5%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.65)	8000	5%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.57)	7945	6%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.39)	8091	4%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.45)	8086	4%	S
General	Percentage of students who begin and finish the year in the same school	45.0	100.0	93.8	(7.05)	338	13%	C
	Index of minor behavior problems	-0.1	0.8	0.0	(0.13)	259	34%	C
	Index of major behavior problems	-0.1	0.8	0.0	(0.13)	248	36%	C
	Hours per month principal teaches	0.0	99.0	15.8	(25.32)	341	13%	C
	Teachers years of experience	1.0	40.0	18.2	(9.12)	434	9%	T
School Structure								
General	Urban location	0.0	1.0	0.4	(0.50)	355	9%	C
	School size	64.0	980.0	308.1	(154.34)	345	12%	C
	Average class size	6.0	57.0	24.2	(7.79)	334	14%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.3 Descriptive Statistics for Cyprus at the Upper Grade of Population 1

Cyprus
N of Students = 3,376
N of Schools = 146

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-4.6	1.4	0.0	(0.55)	3322	2%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.2	1.2	0.0	(0.53)	3261	3%	S
	Time on mathematics homework	1.0	5.0	2.4	(0.87)	3245	4%	S
	Likes mathematics	-4.0	-1.0	-1.3	(0.62)	3255	4%	S
Science	Locus of control in science	-4.8	1.5	0.0	(0.53)	3243	4%	S
	Time on science homework	1.0	5.0	2.2	(0.86)	3232	4%	S
	Likes science	-4.0	-1.0	-1.6	(0.76)	3240	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.57)	3248	4%	S
	Frequency of testing	-3.0	-1.0	-1.5	(0.53)	3211	5%	S
	Frequency of calculator use	-3.0	-1.0	-2.8	(0.49)	3205	5%	S
	Frequency of computer use	-3.0	-1.0	-2.8	(0.49)	3132	7%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.61)	3218	5%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.60)	3208	5%	S
	Frequency of calculator use	-3.0	-1.0	-2.8	(0.51)	3192	5%	S
	Frequency of computer use	-3.0	-1.0	-2.8	(0.50)	3177	6%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.56)	3273	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-2.0	(0.66)	3260	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.9	(0.64)	3215	5%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.5	(0.62)	3247	4%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.66)	3240	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.6	(0.64)	3203	5%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.36)	3236	4%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.8	(0.43)	3202	5%	S
General	Percentage of students who begin and finish the year in the same school	95.0	100.0	99.3	(0.79)	121	17%	C
	Index of minor behavior problems	-0.6	1.8	0.0	(0.59)	122	16%	C
	Index of major behavior problems	-0.5	2.8	0.0	(0.70)	121	17%	C
	Hours per month principal teaches	0.0	80.0	34.2	(20.40)	126	14%	C
	Teachers years of experience	1.0	39.0	13.6	(12.29)	136	36%	T
School Structure								
General	Urban location	0.0	1.0	0.4	(0.48)	127	13%	C
	School size	72.0	1998.0	268.7	(170.34)	122	16%	C
	Average class size	7.0	60.0	27.6	(7.61)	119	18%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.4 Descriptive Statistics for the Czech Republic at the Upper Grade of Population 1

Czech_Rep
N of Students = 3,268
N of Schools = 187

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-2.6	1.4	0.0	(0.53)	3261	0%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.1	2.2	0.0	(0.52)	3239	1%	S
	Time on mathematics homework	1.0	5.0	2.2	(0.62)	3136	4%	S
	Likes mathematics	-4.0	-1.0	-1.8	(0.79)	3222	1%	S
Science	Locus of control in science	-3.1	2.1	0.0	(0.51)	3221	1%	S
	Time on science homework	1.0	5.0	2.0	(0.63)	3133	4%	S
	Likes science	-4.0	-1.0	-1.8	(0.81)	3226	1%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.55)	3217	2%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.50)	3198	2%	S
	Frequency of calculator use	-3.0	-1.0	-2.6	(0.57)	3215	2%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.31)	3177	3%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.8	(0.62)	3221	1%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.50)	3221	1%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.37)	3223	1%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.27)	3183	3%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.53)	3227	1%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.8	(0.66)	3219	1%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.55)	3191	2%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.2	(0.40)	3242	1%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.4	(0.61)	3236	1%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.57)	3191	2%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.33)	3223	1%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.9	(0.36)	3214	2%	S
General	Percentage of students who begin and finish the year in the same school	80.0	100.0	98.5	(2.20)	168	10%	C
	Index of minor behavior problems	-0.1	0.4	0.0	(0.12)	176	6%	C
	Index of major behavior problems	-0.1	0.9	0.0	(0.13)	175	6%	C
	Hours per month principal teaches	1.0	172.0	51.4	(35.02)	179	4%	C
	Teachers years of experience	1.0	45.0	21.4	(12.92)	184	17%	T
School Structure								
General	Urban location	0.0	1.0	0.2	(0.36)	184	2%	C
	School size	17.0	1088.0	238.6	(211.66)	181	3%	C
	Average class size	2.0	46.0	17.4	(8.20)	181	3%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.5 Descriptive Statistics for Greece at the Upper Grade of Population 1

Greece	
N of Students =	3,053
N of Schools =	174

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-4.7	1.5	0.0	(0.57)	2998	2%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-2.0	1.2	0.0	(0.52)	2934	4%	S
	Time on mathematics homework	1.0	5.0	2.8	(1.03)	2881	6%	S
	Likes mathematics	-4.0	-1.0	-1.4	(0.66)	2947	3%	S
Science	Locus of control in science	-2.4	1.3	0.0	(0.51)	2922	4%	S
	Time on science homework	1.0	5.0	2.6	(0.98)	2884	6%	S
	Likes science	-4.0	-1.0	-1.4	(0.64)	2929	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.7	(0.75)	2873	6%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.59)	2878	6%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.44)	2850	7%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.44)	2834	7%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.7	(0.75)	2832	7%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.61)	2807	8%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.42)	2835	7%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.44)	2817	8%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.56)	2923	4%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.71)	2899	5%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.2	(0.71)	2858	6%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.6	(0.69)	2865	6%	S
	Frequency of notetaking from the board	-3.0	-1.0	-2.1	(0.77)	2859	6%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.74)	2819	8%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.26)	2916	4%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.9	(0.31)	2880	6%	S
General	Percentage of students who begin and finish the year in the same school	90.0	100.0	98.6	(2.28)	156	10%	C
	Index of minor behavior problems	-0.1	0.4	0.0	(0.09)	142	18%	C
	Index of major behavior problems	-0.1	0.5	0.0	(0.10)	139	20%	C
	Hours per month principal teaches	0.0	96.0	29.2	(30.04)	169	3%	C
	Teachers years of experience	2.0	38.0	16.1	(8.52)	156	3%	T
School Structure								
General	Urban location	0.0	1.0	0.3	(0.44)	169	3%	C
	School size	5.0	1183.0	118.4	(107.35)	157	10%	C
	Average class size	1.0	60.0	14.2	(10.34)	143	18%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.6 Descriptive Statistics for Hong Kong at the Upper Grade of Population 1

Hong_Kong
N of Students = 4,411
N of Schools = 124

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-3.9	1.5	0.0	(0.56)	4394	0%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.0	1.1	0.0	(0.62)	4373	1%	S
	Time on mathematics homework	1.0	5.0	2.6	(0.87)	4331	2%	S
	Likes mathematics	-4.0	-1.0	-1.9	(0.84)	4344	2%	S
Science	Locus of control in science	-3.0	1.1	0.0	(0.59)	4377	1%	S
	Time on science homework	1.0	5.0	2.3	(0.73)	4318	2%	S
	Likes science	-4.0	-1.0	-1.7	(0.75)	4359	1%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.9	(0.71)	4343	2%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.55)	4311	2%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.32)	4338	2%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.34)	4288	3%	S
Science	Frequency of workbook homework	.	.	.	.	0	100%	S
	Frequency of testing	.	.	.	.	0	100%	S
	Frequency of calculator use	.	.	.	.	0	100%	S
	Frequency of computer use	.	.	.	.	0	100%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.53)	4355	1%	S
	Frequency of notetaking from the board	-3.0	-1.0	-2.4	(0.71)	4317	2%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.6	(0.58)	4328	2%	S
Science	Frequency of problem solving in class	.	.	.	.	0	100%	S
	Frequency of notetaking from the board	.	.	.	.	0	100%	S
	Frequency of small group work sessions	.	.	.	.	0	100%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.36)	4343	2%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.45)	4348	1%	S
General	Percentage of students who begin and finish the year in the same school	50.0	100.0	96.9	(4.64)	100	19%	C
	Index of minor behavior problems	-0.2	1.2	0.0	(0.25)	98	21%	C
	Index of major behavior problems	-0.1	1.8	0.0	(0.24)	92	26%	C
	Hours per month principal teaches	0.0	120.0	9.6	(26.96)	116	6%	C
	Teachers years of experience	1.0	36.0	15.0	(10.67)	125	46%	T
School Structure								
General	Urban location	0.0	1.0	0.7	(0.48)	119	4%	C
	School size	53.0	1998.0	567.9	(330.21)	115	7%	C
	Average class size	9.0	45.0	35.0	(5.79)	108	13%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.7 Descriptive Statistics for Iran at the Upper Grade of Population 1

Iran	
N of Students =	3,385
N of Schools =	180

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-2.6	2.5	0.0	(0.63)	3157	7%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-4.8	2.0	0.0	(0.65)	2987	12%	S
	Time on mathematics homework	1.0	5.0	3.3	(1.09)	2345	31%	S
	Likes mathematics	-4.0	-1.0	-1.3	(0.55)	2925	14%	S
Science	Locus of control in science	-4.7	2.0	0.0	(0.64)	2935	13%	S
	Time on science homework	1.0	5.0	3.2	(1.06)	2301	32%	S
	Likes science	-4.0	-1.0	-1.3	(0.55)	2813	17%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.3	(0.50)	2668	21%	S
	Frequency of testing	-3.0	-1.0	-1.3	(0.53)	2711	20%	S
	Frequency of calculator use	-3.0	-1.0	-2.4	(0.82)	2580	24%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.72)	2459	27%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.4	(0.62)	2640	22%	S
	Frequency of testing	-3.0	-1.0	-1.3	(0.54)	2673	21%	S
	Frequency of calculator use	-3.0	-1.0	-2.5	(0.79)	2580	24%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.74)	2444	28%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.2	(0.43)	2649	22%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.4	(0.55)	2733	19%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.7	(0.77)	2514	26%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.2	(0.47)	2715	20%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.4	(0.55)	2693	20%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.7	(0.74)	2522	25%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.34)	2840	16%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.9	(0.33)	2839	16%	S
General	Percentage of students who begin and finish the year in the same school	70.0	100.0	97.2	(4.74)	169	6%	C
	Index of minor behavior problems	0.0	0.4	0.0	(0.05)	147	18%	C
	Index of major behavior problems	0.0	0.5	0.0	(0.06)	140	22%	C
	Hours per month principal teaches	0.0	96.0	10.1	(19.45)	171	5%	C
	Teachers years of experience	1.0	27.0	10.9	(7.40)	159	12%	T
School Structure								
General	Urban location	0.0	1.0	0.2	(0.43)	174	3%	C
	School size	11.0	1590.0	253.7	(243.87)	173	4%	C
	Average class size	1.0	60.0	26.3	(10.89)	166	8%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.8 Descriptive Statistics for Ireland at the Upper Grade of Population 1

Ireland	
N of Students =	2,873
N of Schools =	165

LEVEL 1 - STUDENT LEVEL		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-3.2	1.2	0.0	(0.57)	2838	1%	S

LEVEL 2 - SCHOOL LEVEL		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-3.3	1.2	0.0	(0.59)	2823	2%	S
	Time on mathematics homework	1.0	5.0	2.2	(0.68)	2725	5%	S
	Likes mathematics	-4.0	-1.0	-1.8	(0.92)	2770	4%	S
Science	Locus of control in science	-3.1	1.2	0.0	(0.58)	2798	3%	S
	Time on science homework	1.0	5.0	1.7	(0.72)	2695	6%	S
	Likes science	-4.0	-1.0	-2.0	(0.92)	2746	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.8	(0.73)	2770	4%	S
	Frequency of testing	-3.0	-1.0	-1.8	(0.63)	2766	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.40)	2773	3%	S
	Frequency of computer use	-3.0	-1.0	-2.8	(0.51)	2758	4%	S
Science	Frequency of workbook homework	-3.0	-1.0	-2.0	(0.79)	2736	5%	S
	Frequency of testing	-3.0	-1.0	-2.0	(0.71)	2748	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.37)	2757	4%	S
	Frequency of computer use	-3.0	-1.0	-2.8	(0.47)	2730	5%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.52)	2814	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.66)	2788	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.4	(0.70)	2751	4%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.7	(0.73)	2777	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.72)	2767	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.4	(0.70)	2736	5%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.41)	2723	5%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.47)	2701	6%	S
General	Percentage of students who begin and finish the year in the same school	90.0	100.0	99.2	(1.34)	155	6%	C
	Index of minor behavior problems	-0.1	0.7	0.0	(0.16)	120	27%	C
	Index of major behavior problems	-0.1	0.8	0.0	(0.16)	119	28%	C
	Hours per month principal teaches	0.0	160.0	70.0	(51.65)	140	15%	C
	Teachers years of experience	1.0	41.0	19.3	(10.65)	157	5%	T
School Structure								
General	Urban location	0.0	1.0	0.2	(0.36)	158	4%	C
	School size	33.0	886.0	159.1	(139.97)	152	8%	C
	Average class size	3.0	39.0	17.1	(10.11)	149	10%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.9 Descriptive Statistics for Korea at the Upper Grade of Population 1

Korea	
N of Students =	2,812
N of Schools =	150

LEVEL 1 - STUDENT LEVEL

			Min	Max	Mean	(SD)	N	%Missing	File*
Control									
	General	Student SES	-2.7	1.4	0.0	(0.56)	2806	0%	S

LEVEL 2 - SCHOOL LEVEL

			Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement									
	Mathematics	Locus of control in mathematics	-2.9	1.8	0.0	(0.60)	2790	1%	S
		Time on mathematics homework	1.0	5.0	2.4	(0.84)	2745	2%	S
		Likes mathematics	-4.0	-1.0	-2.0	(0.88)	2756	2%	S
	Science	Locus of control in science	-4.0	1.1	0.0	(0.59)	2785	1%	S
		Time on science homework	1.0	5.0	2.1	(0.84)	2735	3%	S
		Likes science	-4.0	-1.0	-1.7	(0.72)	2752	2%	S
Instructional Methods									
	Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.9	(0.66)	2770	1%	S
		Frequency of testing	-3.0	-1.0	-1.8	(0.62)	2763	2%	S
		Frequency of calculator use	-3.0	-1.0	-2.9	(0.34)	2758	2%	S
		Frequency of computer use	-3.0	-1.0	-2.9	(0.37)	2742	2%	S
	Science	Frequency of workbook homework	-3.0	-1.0	-2.1	(0.71)	2743	2%	S
		Frequency of testing	-3.0	-1.0	-2.0	(0.67)	2748	2%	S
		Frequency of calculator use	-3.0	-1.0	-2.9	(0.28)	2744	2%	S
		Frequency of computer use	-3.0	-1.0	-2.9	(0.36)	2724	3%	S
Classroom Organization									
	Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.63)	2777	1%	S
		Frequency of notetaking from the board	-3.0	-1.0	-1.6	(0.64)	2769	2%	S
		Frequency of small group work sessions	-3.0	-1.0	-2.5	(0.69)	2741	3%	S
	Science	Frequency of problem solving in class	-3.0	-1.0	-1.8	(0.73)	2757	2%	S
		Frequency of notetaking from the board	-3.0	-1.0	-1.6	(0.68)	2759	2%	S
		Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.70)	2709	4%	S
School Climate									
	Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.6	(0.49)	2683	5%	S
	Science	Perception of peer attitudes toward science	0.0	1.0	0.6	(0.49)	2674	5%	S
	General	Percentage of students who begin and finish the year in the same school	77.0	100.0	96.7	(5.35)	143	5%	C
		Index of minor behavior problems	-0.1	0.9	0.0	(0.13)	141	6%	C
		Index of major behavior problems	-0.1	1.5	0.0	(0.17)	140	7%	C
		Hours per month principal teaches	0.0	97.0	6.0	(16.71)	148	1%	C
		Teachers years of experience	1.0	42.0	17.3	(10.27)	150	0%	T
School Structure									
	General	Urban location	0.0	1.0	0.3	(0.46)	149	1%	C
		School size	49.0	4053.0	870.6	(867.31)	148	1%	C
		Average class size	8.0	60.0	34.4	(12.84)	147	2%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.10 Descriptive Statistics for Latvia at the Upper Grade of Population 1

Latvia_(LSS)	
N of Students =	2,216
N of Schools =	125

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-3.2	1.2	0.0	(0.52)	2191	1%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-2.3	1.8	0.0	(0.51)	2164	2%	S
	Time on mathematics homework	1.0	5.0	2.3	(0.73)	2015	9%	S
	Likes mathematics	-4.0	-1.0	-1.9	(0.86)	2142	3%	S
Science	Locus of control in science	-2.3	1.7	0.0	(0.49)	2160	3%	S
	Time on science homework	1.0	5.0	2.1	(0.67)	1970	11%	S
	Likes science	-4.0	-1.0	-2.0	(0.84)	2127	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.6	(0.63)	2100	5%	S
	Frequency of testing	-3.0	-1.0	-2.3	(0.63)	2098	5%	S
	Frequency of calculator use	-3.0	-1.0	-2.8	(0.51)	2124	4%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.39)	2090	6%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.6	(0.68)	2084	6%	S
	Frequency of testing	-3.0	-1.0	-2.3	(0.66)	2065	7%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.43)	2096	5%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.39)	2068	7%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.5	(0.55)	2161	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.56)	2154	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.66)	2113	5%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.6	(0.61)	2135	4%	S
	Frequency of notetaking from the board	-3.0	-1.0	-2.0	(0.63)	2119	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.69)	2069	7%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.32)	2126	4%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.8	(0.41)	2103	5%	S
General	Percentage of students who begin and finish the year in the same school	80.0	100.0	97.4	(3.32)	92	26%	C
	Index of minor behavior problems	-0.2	0.9	0.0	(0.26)	112	10%	C
	Index of major behavior problems	-0.2	0.9	0.0	(0.24)	112	10%	C
	Hours per month principal teaches	0.0	84.0	37.2	(21.25)	108	14%	C
	Teachers years of experience	1.0	49.0	18.2	(11.65)	152	7%	T
School Structure								
General	Urban location	0.0	1.0	0.1	(0.35)	111	11%	C
	School size	38.0	1300.0	263.7	(259.67)	104	17%	C
	Average class size	4.0	45.0	15.5	(6.82)	107	14%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.11 Descriptive Statistics for New Zealand at the Upper Grade of Population 1

New_Zealand	
N of Students =	2,421
N of Schools =	149

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-2.3	1.1	0.0	(0.56)	2404	1%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-2.5	1.3	0.0	(0.55)	2383	2%	S
	Time on mathematics homework	1.0	5.0	2.1	(0.89)	2257	7%	S
	Likes mathematics	-4.0	-1.0	-1.8	(0.95)	2358	3%	S
Science	Locus of control in science	-2.7	1.1	0.0	(0.55)	2364	2%	S
	Time on science homework	1.0	5.0	1.6	(0.88)	2219	8%	S
	Likes science	-4.0	-1.0	-1.8	(0.92)	2353	3%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.6	(0.60)	2355	3%	S
	Frequency of testing	-3.0	-1.0	-1.8	(0.57)	2333	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.0	(0.62)	2320	4%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.68)	2312	5%	S
Science	Frequency of workbook homework	-3.0	-1.0	-2.1	(0.74)	2321	4%	S
	Frequency of testing	-3.0	-1.0	-2.0	(0.71)	2325	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.4	(0.71)	2308	5%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.70)	2302	5%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.5	(0.56)	2377	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.67)	2354	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.8	(0.55)	2330	4%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.8	(0.72)	2351	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.71)	2344	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.8	(0.64)	2304	5%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.8	(0.43)	2292	5%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.47)	2269	6%	S
General	Percentage of students who begin and finish the year in the same school	50.0	100.0	87.0	(10.24)	131	12%	C
	Index of minor behavior problems	-0.1	0.5	0.0	(0.18)	107	28%	C
	Index of major behavior problems	-0.1	0.7	0.0	(0.17)	107	28%	C
	Hours per month principal teaches	0.0	160.0	54.1	(52.51)	135	9%	C
	Teachers years of experience	1.0	45.0	14.8	(9.72)	170	9%	T
School Structure								
General	Urban location	0.0	1.0	0.3	(0.46)	147	1%	C
	School size	27.0	1996.0	187.7	(189.77)	149	0%	C
	Average class size	4.0	38.0	24.8	(7.59)	141	5%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.12 Descriptive Statistics for Portugal at the Upper Grade of Population 1

Portugal	
N of Students =	2,853
N of Schools =	148

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-2.9	1.4	0.0	(0.63)	2833	1%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-4.4	1.2	0.0	(0.56)	2797	2%	S
	Time on mathematics homework	1.0	5.0	2.6	(0.90)	2720	5%	S
	Likes mathematics	-4.0	-1.0	-1.6	(0.71)	2743	4%	S
Science	Locus of control in science	-4.5	1.2	0.0	(0.54)	2786	2%	S
	Time on science homework	1.0	5.0	2.6	(0.90)	2719	5%	S
	Likes science	-4.0	-1.0	-1.5	(0.61)	2747	4%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.57)	2743	4%	S
	Frequency of testing	-3.0	-1.0	-1.5	(0.56)	2732	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.6	(0.62)	2714	5%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.39)	2670	6%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.59)	2744	4%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.55)	2740	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.7	(0.58)	2726	4%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.39)	2701	5%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.4	(0.54)	2755	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.6	(0.62)	2732	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.9	(0.62)	2723	5%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.5	(0.59)	2762	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.6	(0.62)	2758	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.9	(0.58)	2729	4%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.29)	2759	3%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.9	(0.29)	2758	3%	S
General	Percentage of students who begin and finish the year in the same school	70.0	101.0	97.1	(4.92)	126	15%	C
	Index of minor behavior problems	-0.1	0.8	0.0	(0.14)	92	38%	C
	Index of major behavior problems	-0.1	0.7	0.0	(0.15)	90	39%	C
	Hours per month principal teaches	0.0	92.0	18.7	(20.63)	68	54%	C
	Teachers years of experience	1.0	38.0	21.6	(8.34)	148	1%	T
School Structure								
General	Urban location	0.0	1.0	0.5	(0.50)	141	5%	C
	School size	22.0	840.0	155.4	(127.85)	127	14%	C
	Average class size	2.0	43.0	20.3	(5.57)	129	13%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.13 Descriptive Statistics for Slovenia at the Upper Grade of Population 1

Slovenia	
N of Students =	2,566
N of Schools =	121

LEVEL 1 - STUDENT LEVEL		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-3.2	1.2	0.0	(0.56)	2536	1%	S

LEVEL 2 - SCHOOL LEVEL		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-2.4	1.1	0.0	(0.52)	2513	2%	S
	Time on mathematics homework	1.0	5.0	2.4	(0.70)	2469	4%	S
	Likes mathematics	-4.0	-1.0	-1.7	(0.76)	2515	2%	S
Science	Locus of control in science	-2.5	1.1	0.0	(0.53)	2504	2%	S
	Time on science homework	1.0	5.0	2.4	(0.72)	2464	4%	S
	Likes science	-4.0	-1.0	-1.7	(0.76)	2512	2%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.56)	2510	2%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.39)	2504	2%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.36)	2497	3%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.35)	2497	3%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.8	(0.56)	2502	2%	S
	Frequency of testing	-3.0	-1.0	-1.9	(0.41)	2494	3%	S
	Frequency of calculator use	-3.0	-1.0	-2.9	(0.39)	2504	2%	S
	Frequency of computer use	-3.0	-1.0	-2.9	(0.35)	2490	3%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.5	(0.58)	2512	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.63)	2512	2%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.1	(0.53)	2492	3%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.6	(0.58)	2511	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.60)	2509	2%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.48)	2492	3%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.9	(0.31)	2510	2%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.9	(0.35)	2502	2%	S
General	Percentage of students who begin and finish the year in the same school	95.0	100.0	98.0	(1.26)	96	21%	C
	Index of minor behavior problems	-0.3	1.3	0.0	(0.32)	96	21%	C
	Index of major behavior problems	-0.3	1.9	0.0	(0.36)	95	21%	C
	Hours per month principal teaches	0.0	90.0	15.1	(17.04)	102	16%	C
	Teachers years of experience	1.0	36.0	19.1	(10.24)	110	9%	T
School Structure								
General	Urban location	0.0	1.0	0.3	(0.46)	102	16%	C
	School size	122.0	1407.0	485.2	(254.46)	101	17%	C
	Average class size	11.0	33.0	20.9	(4.19)	101	17%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table A.14 Descriptive Statistics for the United States at the Upper Grade of Population 1

United_States	
N of Students =	7,296
N of Schools =	182

LEVEL 1 - STUDENT LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Control								
General	Student SES	-9.2	1.3	0.0	(0.68)	7248	1%	S

LEVEL 2 - SCHOOL LEVEL

		Min	Max	Mean	(SD)	N	%Missing	File*
Student Involvement								
Mathematics	Locus of control in mathematics	-4.4	1.5	0.0	(0.60)	7187	1%	S
	Time on mathematics homework	1.0	5.0	2.4	(0.82)	7024	4%	S
	Likes mathematics	-4.0	-1.0	-1.7	(0.91)	7143	2%	S
Science	Locus of control in science	-2.7	1.4	0.0	(0.59)	7160	2%	S
	Time on science homework	1.0	5.0	2.1	(0.86)	7004	4%	S
	Likes science	-4.0	-1.0	-1.7	(0.86)	7075	3%	S
Instructional Methods								
Mathematics	Frequency of workbook homework	-3.0	-1.0	-1.5	(0.63)	7054	3%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.57)	7047	3%	S
	Frequency of calculator use	-3.0	-1.0	-2.2	(0.66)	7052	3%	S
	Frequency of computer use	-3.0	-1.0	-2.5	(0.72)	6987	4%	S
Science	Frequency of workbook homework	-3.0	-1.0	-1.7	(0.70)	7020	4%	S
	Frequency of testing	-3.0	-1.0	-1.6	(0.60)	7033	4%	S
	Frequency of calculator use	-3.0	-1.0	-2.7	(0.60)	7008	4%	S
	Frequency of computer use	-3.0	-1.0	-2.6	(0.63)	6969	4%	S
Classroom Organization								
Mathematics	Frequency of problem solving in class	-3.0	-1.0	-1.2	(0.45)	7123	2%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.9	(0.72)	7087	3%	S
	Frequency of small group work sessions	-3.0	-1.0	-2.0	(0.58)	7011	4%	S
Science	Frequency of problem solving in class	-3.0	-1.0	-1.6	(0.67)	7065	3%	S
	Frequency of notetaking from the board	-3.0	-1.0	-1.7	(0.69)	7038	4%	S
	Frequency of small group work sessions	-3.0	-1.0	-1.8	(0.62)	6974	4%	S
School Climate								
Mathematics	Perception of peer attitudes toward mathematics	0.0	1.0	0.7	(0.45)	7075	3%	S
Science	Perception of peer attitudes toward science	0.0	1.0	0.7	(0.46)	7076	3%	S
General	Percentage of students who begin and finish the year in the same school	45.0	100.0	91.5	(10.54)	135	26%	C
	Index of minor behavior problems	0.0	0.3	0.0	(0.04)	126	31%	C
	Index of major behavior problems	0.0	0.6	0.0	(0.04)	124	32%	C
	Hours per month principal teaches	0.0	60.0	7.6	(17.12)	147	19%	C
	Teachers years of experience	1.0	40.0	14.9	(9.46)	315	17%	T
School Structure								
General	Urban location	0.0	1.0	0.4	(0.49)	149	18%	C
	School size	66.0	1659.0	419.6	(236.54)	130	29%	C
	Average class size	3.0	53.0	23.0	(7.15)	145	20%	C

*S = Student Questionnaire
*T = Teacher Questionnaire
*C = School Administrator Questionnaire

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

APPENDIX B - COUNTRY LEVEL RESULTS FOR EACH MODEL TESTED IN MATHEMATICS

Table B.1 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Australia

AUSTRALIA

	Unconditional Model
Between School Variance	27%
Parameter Variance	1703.1
Reliability	0.86

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	34%	41%	44%	47%	48%	49%	20%
		Parameter Variance	1119.6 **	1006.6 **	946.5 **	909.2 **	889.3 **	861.6 **	1354.8 **
		Reliability	0.82	0.80	0.79	0.79	0.78	0.78	0.84
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	549.4 **	548.8 **	548.5 **	548.3 **	547.5 **	548.1 **	552.3 **
Control	Student SES	Student socio-economic status (SES)	29.9 **	30.0 **	30.0 **	29.9 **	29.8 **	29.5 **	29.5 **
Block 1	Student Involvement	Time spent studying mathematics	-20.5	-31.3 **	-36.2 **	-42.6 **	-44.0 **	-36.3 **	
		How much students like mathematics	-9.3	-3.6	-1.9	17.2	20.2	17.6	
		Locus of control in mathematics	133.5 **	122.2 **	129.0 **	120.8 **	116.4 **	111.2 **	
Block 2	Instructional Methods	Frequency of worksheets in class		32.5 **	33.8 **	33.4 **	34.5 **	30.8 **	
		Frequency of testing		8.1	9.5	7.9	8.5	13.4	
		Frequency of calculator use		-0.7	1.6	5.0	7.4	6.4	
		Frequency of computer use		-24.5 **	-24.6 **	-21.6 **	-20.3 **	-22.3 **	
Block 3	Classroom Organization	Students take notes in class			10.6	13.6	11.0	16.4	
		Students do problems in class			-26.9 **	-23.9 *	-21.4	-20.7	
		Students work in groups in class			-9.5	-12.4	-12.6	-11.4	
Block 4	School Climate	Stability of student body				25.4	30.9	13.8	
		Index of major discipline problems				-51.7	-32.5	-22.1	
		Index of minor discipline problems				34.4	37.3	25.2	
		Principal leadership				6.3	26.6	30.5	
		Teacher's years of experience				-0.6	-1.1	0.5	
		Students perception of peer attitudes toward mathematics				-31.4 **	-34.3 **	-33.9 **	
Block 5	School Structure	Urban location					-6.8	-4.2	
		School size					32.3	27.5	
		Class size					13.6	-8.5	
Block 6	Mean SES	Mean socio-economic status of the school						41.5	72.2 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.2 Results of seven explanatory models of school effectiveness in grade four mathematics in Canada

CANADA

	Unconditional Model
Between School Variance	26%
Parameter Variance	1557.2
Reliability	0.70

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	36%	39%	39%	40%	41%	45%	27%
		Parameter Variance	996.0 **	949.3 **	951.4 **	929.3 **	917.8 **	862.0 **	1129.8 **
		Reliability	0.63	0.63	0.63	0.62	0.62	0.61	0.65
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	532.7 **	532.4 **	532.4 **	533.2 **	533.3 **	532.2 **	526.8 **
Control	Student SES	Student socio-economic status (SES)	29.2 **	29.0 **	29.0 **	28.9 **	28.8 **	28.0 **	28.0 **
Block 1	Student Involvement	Time spent studying mathematics	-9.1	-7.2	-7.1	-6.0	-6.3	-6.6	
		How much students like mathematics	-7.3	-3.0	-1.7	-0.6	-0.1	-0.4	
		Locus of control in mathematics	104.4 **	90.2 **	88.5 **	87.5 **	87.0 **	77.2 **	
Block 2	Instructional Methods	Frequency of worksheets in class		14.8 **	15.4 **	12.0	11.0	8.8	
		Frequency of testing		-8.0	-3.8	-7.0	-9.2	-6.6	
		Frequency of calculator use		7.6 *	7.5	6.2	6.9	7.0	
		Frequency of computer use		-11.0 *	-11.8 *	-12.3 *	-11.8 *	-10.9 *	
Block 3	Classroom Organization	Students take notes in class			-5.8	-5.1	-3.1	-1.3	
		Students do problems in class			-12.1	-11.5	-8.1	-8.2	
		Students work in groups in class			-0.2	0.7	0.3	-0.3	
Block 4	School Climate	Stability of student body				20.9 *	19.5	16.3	
		Index of major discipline problems				-38.1	-39.5	-39.6	
		Index of minor discipline problems				5.2	3.7	23.5	
		Principal leadership				6.1	-6.1	4.2	
		Teacher's years of experience				-1.8	-0.7	-0.1	
		Students perception of peer attitudes toward mathematics				-1.3	-3.0	-0.5	
Block 5	School Structure	Urban location					-1.6	0.0	
		School size					5.5	-0.6	
		Class size					-40.8 *	-37.0 *	
Block 6	Mean SES	Mean socio-economic status of the school						40.8 **	73.8 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.3 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Cyprus

CYPRUS

	Unconditional Model
Between School Variance	16%
Parameter Variance	1154.1
Reliability	0.77

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	20%	23%	19%	19%	23%	31%	28%
Parameter Variance	920.8 **	887.6 **	930.7 **	934.0 **	893.2 **	797.6 **	833.0 **
Reliability	0.73	0.73	0.74	0.74	0.73	0.71	0.72

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	507.5 **	507.1 **	507.7 **	506.9 **	499.5 **	501.4 **	506.4 **
Control	Student SES	Student socio-economic status (SES)	15.9 **	15.8 **	15.6 **	15.6 **	15.1 **	14.0 **	14.0 **
Block 1	Student Involvement	Time spent studying mathematics	-23.2 *	-25.8 *	-25.6 *	-30.7 *	-31.1 *	-28.1 *	
		How much students like mathematics	-12.1	-12.2	-13.2	-8.2	-9.9	-6.9	
		Locus of control in mathematics	34.3 *	32.1	28.5	12.4	19.2	11.0 *	
Block 2	Instructional Methods	Frequency of worksheets in class		14.4	11.7	11.0	17.1	11.0	
		Frequency of testing		5.2	5.0	6.4	-1.1	2.3	
		Frequency of calculator use		-11.0	-8.9	-9.0	-6.8	-9.3	
		Frequency of computer use		-9.3	-12.1	-11.4	-8.9	-7.5	
Block 3	Classroom Organization	Students take notes in class			-5.3	-2.7	-8.6	-6.8	
		Students do problems in class			0.7	-1.0	4.6	6.3	
		Students work in groups in class			2.4	-0.6	-0.5	-3.6	
Block 4	School Climate	Stability of student body				-7.9	-6.7	-8.9 *	
		Index of major discipline problems				-10.0	-10.6	-2.6	
		Index of minor discipline problems				6.2	5.3	-2.6	
		Principal leadership				1.0	2.4	3.3	
		Teacher's years of experience				5.4	6.1	5.1	
		Students perception of peer attitudes toward mathematics				-7.4	-4.1	-2.1	
Block 5	School Structure	Urban location					15.6 *	9.2	
		School size					-3.8	-2.5	
		Class size					6.1	1.1	
Block 6	Mean SES	Mean socio-economic status of the school						56.6 **	75.0 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.4 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in the Czech Republic

CZECH REPUBLIC

	Unconditional Model
Between School Variance	24%
Parameter Variance	1386.6
Reliability	0.76

Explanatory Block		Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	16%	13%	13%	12%	12%	20%	19%
		Parameter Variance	1161.3 **	1204.5 **	1213.0 **	1226.3 **	1214.3 **	1109.2 **	1127.2 **
		Reliability	0.73	0.74	0.74	0.74	0.74	0.72	0.72
			Betas	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	564.8 **	564.9 **	565.1 **	565.1 **	560.6 **	563.5 **	565.4 **
Control	Student SES	Student socio-economic status (SES)	27.8 **	27.8 **	27.6 **	27.7 **	27.5 **	26.1 **	26.1 **
Block 1	Student Involvement	Time spent studying mathematics	-7.6	-7.8	-4.7	-6.4	-8.9	-13.2	
		How much students like mathematics	-17.4 *	-17.8 *	-17.0 *	-17.0 *	-14.4	-9.5	
		Locus of control in mathematics	49.8 **	50.3 **	48.9 **	41.7 **	43.1 **	49.5 **	
Block 2	Instructional Methods	Frequency of worksheets in class		0.4	1.3	1.6	1.6	1.0	
		Frequency of testing		2.3	2.6	1.4	-1.2	-1.5	
		Frequency of calculator use		1.2	2.2	1.9	2.5	6.0	
		Frequency of computer use		1.0	0.3	2.0	0.1	2.3	
Block 3	Classroom Organization	Students take notes in class			0.5	2.1	1.5	0.3	
		Students do problems in class			-11.3	-13.0 *	-12.5 *	-10.3	
		Students work in groups in class			-3.2	-4.7	-5.6	-3.3	
Block 4	School Climate	Stability of student body				0.1	-1.4	3.4	
		Index of major discipline problems				-1.8	5.2	-3.7	
		Index of minor discipline problems				22.5	13.3	16.8	
		Principal leadership				15.6	25.8	31.9	
		Teacher's years of experience				5.0	4.8	5.3	
		Students perception of peer attitudes toward mathematics				10.8	10.6	8.0	
Block 5	School Structure	Urban location					14.2 *	12.6	
		School size					4.6	-6.0	
		Class size					7.3	4.4	
Block 6	Mean SES	Mean socio-economic status of the school						51.1 **	41.8 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.5 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Greece

GREECE

	Unconditional Model
Between School Variance	26%
Parameter Variance	1596.5
Reliability	0.83

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	13%	17%	17%	28%	26%	39%	29%
Parameter Variance	1388.4 **	1323.1 **	1323.2 **	1141.8 **	1188.4 **	976.5 **	1139.1 **
Reliability	0.81	0.80	0.80	0.78	0.79	0.75	0.78

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	491.0 **	489.9 **	489.7 **	494.7 **	496.1 **	501.7 **	489.7 **
Control	Student SES	Student socio-economic status (SES)	27.0 **	26.9 **	26.6 **	26.6 **	26.7 **	25.0 **	25.0 **
Block 1	Student Involvement	Time spent studying mathematics	-10.0	-10.4	-10.9	-5.5	-6.0	-2.4	
		How much students like mathematics	-16.9	-19.6	-17.5	-19.0	-18.7	-12.4	
		Locus of control in mathematics	19.4	10.1	8.8	-0.7	-0.5	-10.3 *	
Block 2	Instructional Methods	Frequency of worksheets in class		20.5	23.5	21.2	22.0	12.0	
		Frequency of testing		18.7	22.8 *	16.3	14.6	9.2	
		Frequency of calculator use		-18.0	-12.2	-2.6	-5.3	-21.2	
		Frequency of computer use		-10.0	-11.3	-9.1	-4.7	0.5	
Block 3	Classroom Organization	Students take notes in class			1.5	0.2	1.4	0.6	
		Students do problems in class			-8.0	-15.5	-18.3	-7.4	
		Students work in groups in class			-17.8	-22.2	-24.8 *	-20.6	
Block 4	School Climate	Stability of student body				22.3	23.5	27.1	
		Index of major discipline problems				120.3 *	123.0 *	56.3	
		Index of minor discipline problems				-89.2	-102.4	-48.2	
		Principal leadership				-40.9	-45.8 *	-62.2 **	
		Teacher's years of experience				8.0 *	7.9 *	8.1 *	
		Students perception of peer attitudes toward mathematics				31.8 **	33.1 **	21.6	
Block 5	School Structure	Urban location					5.2	3.2	
		School size					-23.1	-59.4 *	
		Class size					-1.9	-1.3	
Block 6	Mean SES	Mean socio-economic status of the school						82.2 **	80.2 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.6 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Hong Kong

HONG KONG

	Unconditional Model
Between School Variance	46%
Parameter Variance	2605.3
Reliability	0.95

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	35%	44%	43%	42%	41%	55%	26%
		Parameter Variance	1684.8 **	1446.4 **	1491.0 **	1520.1 **	1538.6 **	1182.6 **	1922.5 **
		Reliability	0.93	0.92	0.92	0.92	0.92	0.90	0.94
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	590.9 **	590.6 **	591.2 **	593.4 **	587.9 **	590.8 **	588.7 **
Control	Student SES	Student socio-economic status (SES)	-3.2	-3.0	-3.1	-3.2	-3.3	-4.4	-4.4
Block 1	Student Involvement	Time spent studying mathematics	0.2	4.8	0.9	-2.1	-8.4	-15.4	
		How much students like mathematics	4.2	2.1	4.1	-0.1	8.1	15.3	
		Locus of control in mathematics	209.7 **	147.9 **	148.0 **	125.7 **	133.0 **	110.5 **	
Block 2	Instructional Methods	Frequency of worksheets in class		-0.9	-5.5	-5.8	-6.2	-11.0	
		Frequency of testing		-23.4	-25.4	-14.0	-12.3	-21.8	
		Frequency of calculator use		-58.2 **	-54.8 **	-54.2 **	-48.9 **	-33.3	
		Frequency of computer use		-3.3	-0.3	-1.4	-1.1	-16.0	
Block 3	Classroom Organization	Students take notes in class			9.2	4.2	6.6	1.0	
		Students do problems in class			4.8	14.3	3.5	-7.2	
		Students work in groups in class			-9.3	-9.1	-11.8	-14.7	
Block 4	School Climate	Stability of student body				-3.0	-3.2	-0.4	
		Index of major discipline problems				-24.3	-22.7	-18.0	
		Index of minor discipline problems				-3.7	-4.6	-3.6	
		Principal leadership				26.7	36.0	22.6	
		Teacher's years of experience				-7.9	-4.1	-4.4	
		Students perception of peer attitudes toward mathematics				17.5	21.3	23.3	
Block 5	School Structure	Urban location					2.1	3.5	
		School size					15.9	-8.6	
		Class size					20.0	18.6	
Block 6	Mean SES	Mean socio-economic status of the school						80.7 **	104.6 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.7 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Iran

IRAN

	Unconditional Model
Between School Variance	44%
Parameter Variance	1620.0
Reliability	0.88

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	10%	22%	21%	36%	42%	60%	52%
		Parameter Variance	1460.4 **	1267.8 **	1273.5 **	1029.0 **	936.6 **	647.7 **	774.0 **
		Reliability	0.86	0.85	0.85	0.82	0.81	0.75	0.78
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	434.2 **	437.0 **	437.2 **	434.8 **	426.0 **	434.7 **	434.1 **
Control	Student SES	Student socio-economic status (SES)	2.7	2.8	2.8	2.1	1.7	0.1	0.1
Block 1	Student Involvement	Time spent studying mathematics	16.0	14.0	12.0	10.3	10.4	2.6	
		How much students like mathematics	-28.1 **	-15.6	-12.5	-6.8	-8.0	-4.9	
		Locus of control in mathematics	34.4 **	12.9	14.2	3.4	0.4	1.1 *	
Block 2	Instructional Methods	Frequency of worksheets in class		-17.9	-16.1	-18.9	-17.8	-12.4	
		Frequency of testing		9.0	9.5	15.0	13.1	8.6	
		Frequency of calculator use		-24.5 **	-24.0 **	-14.9	-15.4	-16.8 *	
		Frequency of computer use		-6.2	-4.4	-6.9	-6.1	-7.3	
Block 3	Classroom Organization	Students take notes in class			5.5	-0.2	3.0	7.2	
		Students do problems in class			-0.9	-0.9	-2.6	1.3	
		Students work in groups in class			-14.2	-7.9	-5.7	-4.5	
Block 4	School Climate	Stability of student body				36.6	52.4	69.1	
		Index of major discipline problems				-156.4	-119.0	35.4	
		Index of minor discipline problems				-180.0	-227.9	-168.3	
		Principal leadership				4.6	17.1	25.0	
		Teacher's years of experience				18.3 **	14.7 **	5.2	
		Students perception of peer attitudes toward mathematics				-4.7	0.5	3.3	
Block 5	School Structure	Urban location					21.5 **	7.0	
		School size					37.6	14.5	
		Class size					4.9	-43.1	
Block 6	Mean SES	Mean socio-economic status of the school						75.6 **	95.0 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.8 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Ireland

IRELAND

	Unconditional Model
Between School Variance	25%
Parameter Variance	1331.1
Reliability	0.76

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	25%	37%	46%	51%	56%	60%	24%
		Parameter Variance	1004.3 **	837.5 **	719.1 **	652.0 **	581.1 **	538.3 **	1014.1 **
		Reliability	0.72	0.69	0.66	0.64	0.61	0.59	0.73
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	547.3 **	548.2 **	548.1 **	547.9 **	548.4 **	551.7 **	548.8 **
Control	Student SES	Student socio-economic status (SES)	40.3 **	40.2 **	40.1 **	40.5 **	40.5 **	39.1 **	39.1 **
Block 1	Student Involvement	Time spent studying mathematics	-20.7	-21.7 *	-15.5	-17.5	-18.1	-11.6	
		How much students like mathematics	2.7	6.3	9.1	3.1	1.5	3.8	
		Locus of control in mathematics	38.4 *	48.0 **	37.4 *	37.7 *	39.3 *	39.1 *	
Block 2	Instructional Methods	Frequency of worksheets in class		-0.3	4.8	5.6	6.6	7.0	
		Frequency of testing		-17.8 *	-12.5	-12.3	-15.0	-12.8	
		Frequency of calculator use		-11.1	-7.6	-0.5	0.1	0.3	
		Frequency of computer use		-15.5 **	-13.0 **	-9.6	-12.0 *	-11.3 *	
Block 3	Classroom Organization	Students take notes in class			-23.9 **	-16.3 *	-13.4	-13.9	
		Students do problems in class			4.4	3.1	-0.8	-4.6	
		Students work in groups in class			-10.5	-13.6 *	-12.2 *	-9.7	
Block 4	School Climate	Stability of student body				24.2	30.3 *	29.2	
		Index of major discipline problems				4.4	9.6	8.0	
		Index of minor discipline problems				-58.8	-53.1	-41.9	
		Principal leadership				6.4	16.3	13.5	
		Teacher's years of experience				5.5	5.0	4.3	
		Students perception of peer attitudes toward mathematics				4.9	3.9	8.2	
Block 5	School Structure	Urban location					-9.5	-11.9	
		School size					37.4 **	32.2 *	
		Class size					-25.9	-32.4	
Block 6	Mean SES	Mean socio-economic status of the school						31.9 *	29.1 *

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.9 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Korea

KOREA

	Unconditional Model
Between School Variance	18%
Parameter Variance	630.1
Reliability	0.71

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	52%	55%	54%	58%	60%	67%	52%
Parameter Variance	302.8 **	285.4 **	289.1 **	265.9 **	249.3 **	210.8 **	301.7 **
Reliability	0.55	0.53	0.54	0.51	0.50	0.46	0.55

Explanatory Block		Independent Variables	Betas (Model 1)	Betas (Model 2)	Betas (Model 3)	Betas (Model 4)	Betas (Model 5)	Betas (Model 6)	Betas (SES Model)
	Intercept	Average Achievement	611.6 **	611.8 **	611.7 **	611.6 **	606.1 **	610.3 **	610.6 **
Control	Student SES	Student socio-economic status (SES)	28.0 **	27.9 **	27.7 **	27.4 **	27.2 **	25.2 **	25.2 **
Block 1	Student Involvement	Time spent studying mathematics	16.9 **	16.3 **	16.7 **	12.7	9.2	3.8	
		How much students like mathematics	18.5 **	16.7 **	17.1 **	4.0	2.6	1.5	
		Locus of control in mathematics	53.3 **	54.7 **	53.0 **	46.7 **	42.2 **	34.7 **	
Block 2	Instructional Methods	Frequency of worksheets in class		-5.3	-4.1	-3.2	-3.5	-6.9	
		Frequency of testing		-4.3	-3.3	-3.4	-1.7	-0.7	
		Frequency of calculator use		-14.5	-13.8	-10.5	-10.7	-7.5	
		Frequency of computer use		-4.8	-4.0	-3.9	-3.8	-6.2	
Block 3	Classroom Organization	Students take notes in class			-4.4	-3.3	-4.5	-3.0	
		Students do problems in class			1.0	-3.2	-3.4	-6.4	
		Students work in groups in class			-5.1	-6.1	-6.1	-5.1	
Block 4	School Climate	Stability of student body				-8.3	-5.0	-1.7	
		Index of major discipline problems				-4.1	-9.5	-8.4	
		Index of minor discipline problems				17.1	23.8	20.1	
		Principal leadership				-7.6	-5.3	-5.3	
		Teacher's years of experience				-1.6	-1.2	-0.4	
		Students perception of peer attitudes toward mathematics				27.6 **	26.6 **	23.9 **	
Block 5	School Structure	Urban location					5.7	5.1	
		School size					29.0	25.2	
		Class size					-23.4	-53.2 *	
Block 6	Mean SES	Mean socio-economic status of the school						38.5 **	56.8 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.10 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Latvia

LATVIA

	Unconditional Model
Between School Variance	56%
Parameter Variance	3366.5
Reliability	0.92

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	6%	20%	19%	27%	27%	27%	2%
Parameter Variance	3166.8 **	2709.3 **	2728.2 **	2461.3 **	2467.3 **	2457.5 **	3295.5 **
Reliability	0.91	0.90	0.90	0.89	0.89	0.89	0.91

	Explanatory Block	Independent Variables	Model 1 Betas	Model 2 Betas	Model 3 Betas	Model 4 Betas	Model 5 Betas	Model 6 Betas	SES Model Betas
	Intercept	Average Achievement	525.0 **	527.7 **	527.5 **	527.6 **	521.1 **	523.8 **	524.9 **
Control	Student SES	Student socio-economic status (SES)	11.3 **	11.4 **	11.3 **	11.2 **	11.0 **	10.6 **	10.6 **
Block 1	Student Involvement	Time spent studying mathematics	-38.6 **	-39.5 **	-38.3 **	-26.3	-21.2	-27.4	
		How much students like mathematics	29.0	25.5	27.0	36.2	36.4	33.5	
		Locus of control in mathematics	27.4	11.6	11.1	-5.3	-13.6	-16.7 *	
Block 2	Instructional Methods	Frequency of worksheets in class		1.8	7.5	4.3	2.1	5.2	
		Frequency of testing		0.2	0.1	3.0	1.3	1.5	
		Frequency of calculator use		-21.3	-19.4	-21.5	-18.4	-14.3	
		Frequency of computer use		-49.1 **	-44.6 **	-36.5 *	-39.7 *	-44.3 **	
Block 3	Classroom Organization	Students take notes in class			-13.7	-28.6	-37.3 *	-37.0 *	
		Students do problems in class			-24.5	-16.7	-13.1	-11.0	
		Students work in groups in class			10.8	16.2	14.5	17.5	
Block 4	School Climate	Stability of student body				-35.9 **	-29.4	-29.6	
		Index of major discipline problems				-15.6	-7.2	-2.1	
		Index of minor discipline problems				38.9	37.0	35.0	
		Principal leadership				-5.5	-2.9	-4.7	
		Teacher's years of experience				8.8	10.8	12.2	
		Students perception of peer attitudes toward mathematics				-51.7 **	-41.2	-38.0	
Block 5	School Structure	Urban location					19.2	16.0	
		School size					20.0	16.6	
		Class size					-21.0	-23.8	
Block 6	Mean SES	Mean socio-economic status of the school						42.4	43.8

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.11 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in New Zealand

NEW ZEALAND

	Unconditional Model
Between School Variance	46%
Parameter Variance	2881.9
Reliability	0.88

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	34%	44%	45%	47%	46%	60%	63%
Parameter Variance	1890.6 **	1617.7 **	1595.5 **	1533.5 **	1559.1 **	1142.4 **	1071.8 **
Reliability	0.84	0.82	0.81	0.81	0.81	0.76	0.75

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	492.6 **	494.4 **	494.6 **	493.2 **	496.6 **	503.8 **	494.9 **
Control	Student SES	Student socio-economic status (SES)	35.3 **	34.6 **	34.3 **	33.4 **	33.4 **	30.8 **	30.8 **
Block 1	Student Involvement	Time spent studying mathematics	11.5	6.5	6.3	0.6	1.1	1.4	
		How much students like mathematics	-23.5 *	-17.3	-18.4	-13.0	-11.8	5.2	
		Locus of control in mathematics	57.8 **	61.6 **	55.6 **	43.0	44.7 *	16.6 *	
Block 2	Instructional Methods	Frequency of worksheets in class		4.5	10.6	0.3	-3.9	-10.1	
		Frequency of testing		-17.3	-13.0	-10.9	-14.1	3.7	
		Frequency of calculator use		-0.6	1.6	1.0	1.7	-4.8	
		Frequency of computer use		-34.4 **	-32.3 **	-30.8 **	-31.8 **	-22.2 **	
Block 3	Classroom Organization	Students take notes in class			-11.3	-13.7	-11.9	-11.2	
		Students do problems in class			10.7	11.7	10.1	-7.8	
		Students work in groups in class			-15.4	-13.4	-7.4	-0.3	
Block 4	School Climate	Stability of student body				9.5	6.5	-14.9	
		Index of major discipline problems				-124.8 *	-127.3 *	-85.6	
		Index of minor discipline problems				22.8	12.3	11.3	
		Principal leadership				4.7	-10.8	-16.7	
		Teacher's years of experience				5.5	5.4	1.5	
		Students perception of peer attitudes toward mathematics				-4.0	-4.8	0.7	
Block 5	School Structure	Urban location					-10.8	-11.9	
		School size					-27.2	-34.2	
		Class size					18.5	-5.7	
Block 6	Mean SES	Mean socio-economic status of the school						90.4 **	101.5 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.12 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Portugal

PORTUGAL

	Unconditional Model
Between School Variance	23%
Parameter Variance	1040.8
Reliability	0.80

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	38%	32%	54%	58%	69%	77%	46%
Parameter Variance	645.9 **	708.2 **	482.5 **	436.5 **	324.0 **	235.6 **	560.5 **
Reliability	0.72	0.74	0.66	0.63	0.56	0.48	0.69

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	469.8 **	469.3 **	474.0 **	475.7 **	481.0 **	488.5 **	469.6 **
Control	Student SES	Student socio-economic status (SES)	34.2 **	33.3 **	33.5 **	33.0 **	32.8 **	31.2 **	31.2 **
Block 1	Student Involvement	Time spent studying mathematics	-14.2	-11.6	-5.7	-7.7	-6.2	8.0	
		How much students like mathematics	-2.8	-10.1	5.3	14.2	32.8	44.8 **	
		Locus of control in mathematics	41.0	44.3	14.7	40.1	74.7	65.7 *	
Block 2	Instructional Methods	Frequency of worksheets in class		0.8	8.6	11.5	7.5	-14.5	
		Frequency of testing		11.5	-18.6	-18.0	-16.2	-36.8	
		Frequency of calculator use		-11.2	-5.7	-9.8	-17.7	-19.5 *	
		Frequency of computer use		9.7	19.4	10.0	7.8	0.4	
Block 3	Classroom Organization	Students take notes in class			5.0	9.0	1.5	-4.1	
		Students do problems in class			-27.8	-26.5	-32.3 *	-48.6 **	
		Students work in groups in class			-36.2 **	-20.9	-14.5	7.9	
Block 4	School Climate	Stability of student body				18.7	45.5	24.3	
		Index of major discipline problems				-14.9	54.7	106.2	
		Index of minor discipline problems				-40.8	-97.9	-148.4 **	
		Principal leadership				-24.9	-31.4	-28.6	
		Teacher's years of experience				-4.7	-11.3	-15.4 *	
		Students perception of peer attitudes toward mathematics				7.5	-18.8	-18.1	
Block 5	School Structure	Urban location					-7.5	-6.5	
		School size					34.8	18.5	
		Class size					-99.2 *	-90.5 *	
Block 6	Mean SES	Mean socio-economic status of the school						50.3 *	35.6 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.13 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in Slovenia

SLOVENIA

	Unconditional Model
Between School Variance	17%
Parameter Variance	932.6
Reliability	0.76

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	43%	52%	52%	47%	46%	52%	31%
		Parameter Variance	531.6 **	444.9 **	449.5 **	497.3 **	501.5 **	448.3 **	646.4 **
		Reliability	0.65	0.61	0.62	0.64	0.64	0.61	0.70
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	552.1 **	552.2 **	551.9 **	552.2 **	553.6 **	555.1 **	551.2 **
Control	Student SES	Student socio-economic status (SES)	24.7 **	24.3 **	24.0 **	24.0 **	23.7 **	22.1 **	22.1 **
Block 1	Student Involvement	Time spent studying mathematics	11.8	5.4	6.8	6.6	7.9	2.9	
		How much students like mathematics	-45.3 **	-52.4 **	-51.2 **	-49.9 **	-52.1 **	-47.6 **	
		Locus of control in mathematics	51.5 **	49.1 **	53.3 **	50.5 **	53.7 **	42.2 *	
Block 2	Instructional Methods	Frequency of worksheets in class		13.6 *	12.2	12.7	11.6	5.9	
		Frequency of testing		5.7	8.2	8.1	9.9	4.9	
		Frequency of calculator use		4.0	7.2	4.0	6.2	2.3	
		Frequency of computer use		-34.7 **	-34.8 **	-26.9	-28.1	-21.4	
Block 3	Classroom Organization	Students take notes in class			-1.7	-0.8	-1.8	-3.7	
		Students do problems in class			-11.3	-11.9	-12.5	-7.9	
		Students work in groups in class			-2.8	-3.6	-2.7	-2.4	
Block 4	School Climate	Stability of student body				-1.7	-3.5	-4.5	
		Index of major discipline problems				9.3	13.0	9.4	
		Index of minor discipline problems				-8.3	-9.6	-9.4	
		Principal leadership				7.4	5.1	5.7	
		Teacher's years of experience				-0.1	0.1	-0.6	
		Students perception of peer attitudes toward mathematics				-0.7	-0.3	-2.6	
Block 5	School Structure	Urban location					-0.3	1.1	
		School size					-10.1	-10.4	
		Class size					6.9	-0.1	
Block 6	Mean SES	Mean socio-economic status of the school						37.9 **	50.0 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table B.14 Results of seven explanatory models of school effectiveness in mathematics at the fourth grade in the United States

UNITED STATES

	Unconditional Model
Between School Variance	40%
Parameter Variance	2526.5
Reliability	0.93

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	48%	53%	63%	69%	73%	78%	61%
		Parameter Variance	1317.7 **	1189.0 **	936.2 **	790.4 **	688.8 **	566.3 **	989.9 **
		Reliability	0.88	0.87	0.85	0.82	0.80	0.77	0.85
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	541.7 **	542.6 **	544.8 **	549.8 **	558.9 **	557.6 **	543.1 **
Control	Student SES	Student socio-economic status (SES)	20.8 **	20.6 **	20.4 **	20.2 **	20.3 **	19.3 **	19.3 **
Block 1	Student Involvement	Time spent studying mathematics	-56.0 **	-48.2 **	-36.5 **	-40.6 **	-21.4	-10.2	
		How much students like mathematics	-19.2	-14.8	-12.2	-2.4	-10.3	0.0	
		Locus of control in mathematics	131.1 **	108.3 **	109.3 **	94.1 **	117.1 **	85.4 **	
Block 2	Instructional Methods	Frequency of worksheets in class		20.2	18.8	20.3 *	12.8	9.3	
		Frequency of testing		-20.7 *	-7.7	-6.4	-5.8	-3.1	
		Frequency of calculator use		-2.9	-2.0	-1.5	1.3	-2.9	
		Frequency of computer use		-13.9 **	-3.4	0.6	1.5	3.9	
Block 3	Classroom Organization	Students take notes in class			-44.7 **	-35.3 **	-29.2 **	-20.5 **	
		Students do problems in class			3.2	2.7	7.1	-1.0	
		Students work in groups in class			0.2	-4.8	-10.0	-13.3	
Block 4	School Climate	Stability of student body				206.6 **	143.4 *	31.7	
		Index of major discipline problems				70.8	50.8	27.7	
		Index of minor discipline problems				-203.3	-155.9	-118.9	
		Principal leadership				114.0	79.9	72.3	
		Teacher's years of experience				-6.6	-5.7	-5.3	
		Students perception of peer attitudes toward mathematics				2.6	-5.2	1.0	
Block 5	School Structure	Urban location					-6.7	-6.2	
		School size					-69.0	-68.8	
		Class size					-289.5 **	-302.0 **	
Block 6	Mean SES	Mean socio-economic status of the school						53.7 **	95.8 **

** results statistically significant at the .05 alpha level
* results statistically significant at the .10 alpha level

SOUCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

APPENDIX C - COUNTRY LEVEL RESULTS FOR EACH MODEL TESTED IN SCIENCE

Table C.1 Results of seven explanatory models of school effectiveness in science at the fourth grade in Australia

AUSTRALIA

	Unconditional Model
Between School Variance	20%
Parameter Variance	1234.9
Reliability	0.81

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	27%	24%	27%	44%	41%	46%	21%
		Parameter Variance	904.4 **	936.9 **	898.3 **	686.5 **	729.5 **	662.9 **	969.6 **
		Reliability	0.78	0.78	0.78	0.73	0.74	0.73	0.79
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	570.8 **	570.2 **	569.1 **	572.8 **	575.3 **	574.0 **	570.8 **
Control	Student SES	Student socio-economic status (SES)	36.5 **	36.4 **	36.5 **	36.5 **	36.5 **	35.9 **	35.9 **
Block 1	Student Involvement	Time spent studying science	-28.1 **	-32.6 **	-33.5 **	-28.7 **	-31.4 **	-24.1 *	
		How much students like science	-12.1	-9.0	-14.8	4.2	2.7	3.0	
		Locus of control in science	63.2 **	59.0 **	57.2 **	44.4 **	46.8 *	41.9 *	
Block 2	Instructional Methods	Frequency of worksheets in class		6.7	0.1	3.5	5.8	7.7	
		Frequency of testing		7.4	2.4	3.3	2.8	2.3	
		Frequency of calculator use		-15.8	-19.7	-23.0	-25.6 *	-17.7	
		Frequency of computer use		4.7	9.7	13.5	12.4	5.7	
Block 3	Classroom Organization	Students take notes in class			10.7	19.1 *	19.6 *	22.2 **	
		Students do problems in class			21.7	19.9	20.1	19.7	
		Students works in groups in class			1.7	0.1	-0.1	-4.0	
Block 4	School Climate	Stability of student body				34.4	28.0	12.6	
		Index of major discipline problems				-63.4	-58.5	-43.2	
		Index of minor discipline problems				40.2	29.4	12.8	
		Principal leadership				80.9 **	71.5 *	75.5 **	
		Teacher's years of experience				0.9	0.0	2.0	
		Students perception of peer attitudes toward science				-34.1 **	-32.9 **	-29.8 *	
Block 5	School Structure	Urban location					-4.8	0.1	
		School size					0.7	-1.9	
		Class size					-30.3	-51.2	
Block 6	Mean SES	Mean socio-economic status of the school						51.8 **	55.1 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.2 Results of seven explanatory models of school effectiveness in science at the fourth grade in Canada

CANADA

	Unconditional Model
Between School Variance	20%
Parameter Variance	1301.9
Reliability	0.64

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	36%	43%	42%	53%	56%	64%	47%
		Parameter Variance	832.0 **	739.7 **	750.2 **	614.0 **	578.9 **	469.9 **	685.1 **
		Reliability	0.58	0.56	0.56	0.53	0.52	0.48	0.55
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	550.1 **	549.9 **	549.4 **	550.7 **	551.9 **	550.4 **	546.6 **
Control	Student SES	Student socio-economic status (SES)	41.3 **	41.4 **	41.3 **	41.1 **	41.0 **	39.3 **	39.3 **
Block 1	Student Involvement	Time spent studying science	-16.7 **	-16.0 *	-17.8 **	-18.7 **	-17.8 **	-17.5 **	
		How much students like science	-8.1	0.0	2.2	-1.4	5.3	7.1	
		Locus of control in science	65.2 **	38.7 **	42.4 **	37.5 **	38.1 **	15.9	
Block 2	Instructional Methods	Frequency of worksheets in class		3.6	6.3	0.4	2.3	8.2	
		Frequency of testing		-3.4	-4.1	-6.2	-9.0	-2.0	
		Frequency of calculator use		-16.5 **	-16.6 **	-16.1 **	-14.0 *	-9.0	
		Frequency of computer use		-23.5 **	-23.6 **	-23.5 **	-23.0 **	-25.2 **	
Block 3	Classroom Organization	Students take notes in class			5.7	4.1	6.8	4.4	
		Students do problems in class			-6.9	-0.1	-3.6	-3.4	
		Students works in groups in class			6.8	4.5	3.8	3.3	
Block 4	School Climate	Stability of student body				23.5 **	20.5 *	10.5	
		Index of major discipline problems				-3.4	-4.3	-5.1	
		Index of minor discipline problems				-67.7 **	-69.6 **	-42.7	
		Principal leadership				25.7	12.7	26.3	
		Teacher's years of experience				-0.1	0.9	1.5	
		Students perception of peer attitudes toward science				6.5	2.5	3.7	
Block 5	School Structure	Urban location					-4.2	-1.4	
		School size					7.7	-4.5	
		Class size					-49.9 **	-40.8 **	
Block 6	Mean SES	Mean socio-economic status of the school						61.7 **	78.2 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.3 Results of seven explanatory models of school effectiveness in science at the fourth grade in Cyprus

CYPRUS

	Unconditional Model
Between School Variance	17%
Parameter Variance	957.8
Reliability	0.79

Explanatory Block		Independent Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	17%	26%	24%	23%	24%	30%	23%
		Parameter Variance	797.3 **	709.3 **	730.4 **	732.8 **	729.0 **	672.6 **	737.0 **
		Reliability	0.76	0.74	0.75	0.75	0.75	0.73	0.75
			Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	480.1 **	480.0 **	480.6 **	480.4 **	475.7 **	477.0 **	479.3 **
Control	Student SES	Student socio-economic status (SES)	13.1 **	13.1 **	13.0 **	12.7 **	12.5 **	11.7 **	11.7 **
Block 1	Student Involvement	Time spent studying science	-23.1 **	-20.5 *	-20.3 *	-17.9	-19.4	-18.5	
		How much students like science	4.5	2.5	1.3	5.0	4.1	6.8	
		Locus of control in science	32.2	21.3	17.0	7.1	1.5	-3.9	
Block 2	Instructional Methods	Frequency of worksheets in class		-9.2	-7.7	-4.1	-4.1	-4.7	
		Frequency of testing		1.9	1.3	3.0	0.6	-0.3	
		Frequency of calculator use		-15.3	-10.4	-9.1	-7.8	-10.7	
		Frequency of computer use		-22.0	-25.8	-25.8	-27.2	-24.9	
Block 3	Classroom Organization	Students take notes in class			-7.1	-4.5	-8.7	-3.4	
		Students do problems in class			0.3	1.9	5.5	3.1	
		Students works in groups in class			-4.5	-4.1	-6.6	-2.4	
Block 4	School Climate	Stability of student body				-2.7	-2.3	-4.1	
		Index of major discipline problems				-13.1	-11.3	-4.3	
		Index of minor discipline problems				12.3	10.8	3.3	
		Principal leadership				-1.3	-1.7	-0.4	
		Teacher's years of experience				2.9	3.0	2.8	
		Students perception of peer attitudes toward science				-12.7	-14.6	-10.2	
Block 5	School Structure	Urban location					11.8	6.0	
		School size					-2.7	-1.3	
		Class size					-2.6	-5.5	
Block 6	Mean SES	Mean socio-economic status of the school						47.1 *	61.5 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.4 Results of seven explanatory models of school effectiveness in science at the fourth grade in the Czech Republic

CZECH REPUBLIC

	Unconditional Model
Between School Variance	23%
Parameter Variance	1182.5
Reliability	0.75

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	15%	19%	23%	20%	19%	23%	22%
Parameter Variance	1010.0 **	954.3 **	911.7 **	942.0 **	956.6 **	916.3 **	918.2 **
Reliability	0.73	0.72	0.71	0.72	0.72	0.71	0.71

Explanatory Block		Independent Variables	Model 1 Betas	Model 2 Betas	Model 3 Betas	Model 4 Betas	Model 5 Betas	Model 6 Betas	SES Model Betas
	Intercept	Average Achievement	554.2 **	554.5 **	554.8 **	554.8 **	552.1 **	554.0 **	554.8 **
Control	Student SES	Student socio-economic status (SES)	32.3 **	31.8 **	31.6 **	31.6 **	31.5 **	30.4 **	30.4 **
Block 1	Student Involvement	Time spent studying science	-2.3	-3.8	1.4	0.0	0.7	-0.3	
		How much students like science	16.2 *	16.8 *	18.4 **	18.0 *	18.8 **	19.3 **	
		Locus of control in science	12.8	15.3	4.4	-1.1	1.8	0.5	
Block 2	Instructional Methods	Frequency of worksheets in class		-8.4	-6.2	-6.3	-5.3	-5.3	
		Frequency of testing		1.0	1.4	2.7	1.9	1.3	
		Frequency of calculator use		-14.4 **	-12.2 *	-12.9 **	-11.8 *	-9.1	
		Frequency of computer use		-2.5	-1.3	-0.2	-1.2	-1.4	
Block 3	Classroom Organization	Students take notes in class			2.6	2.2	0.7	1.0	
		Students do problems in class			10.5 *	10.1	10.6	10.4	
		Students works in groups in class			-9.0	-9.1	-9.3	-7.3	
Block 4	School Climate	Stability of student body				-3.9	-4.6	-1.6	
		Index of major discipline problems				10.9	14.2	8.7	
		Index of minor discipline problems				10.1	3.3	2.9	
		Principal leadership				10.1	20.4	25.3	
		Teacher's years of experience				3.3	3.1	3.5	
		Students perception of peer attitudes toward science				0.8	-1.6	-1.8	
Block 5	School Structure	Urban location					5.0	4.3	
		School size					4.9	-2.6	
		Class size					13.3	11.3	
Block 6	Mean SES	Mean socio-economic status of the school						32.7 **	43.4 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.5 Results of seven explanatory models of school effectiveness in science at the fourth grade in the Greece

GREECE

	Unconditional Model
Between School Variance	24%
Parameter Variance	1301.1
Reliability	0.81

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	16%	34%	44%	51%	50%	63%	30%
Parameter Variance	1098.4 **	861.2 **	729.9 **	636.9 **	652.7 **	477.7 **	905.8 **
Reliability	0.79	0.75	0.72	0.69	0.69	0.63	0.76

Explanatory Block		Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	497.1 **	496.5 **	496.1 **	499.4 **	499.8 **	505.1 **	494.8 **
Control	Student SES	Student socio-economic status (SES)	22.0 **	22.0 **	21.8 **	22.1 **	22.2 **	20.0 **	20.0 **
Block 1	Student Involvement	Time spent studying science	-17.6	-1.6	0.3	-0.7	-0.9	5.7	
		How much students like science	7.9	0.1	1.9	-3.6	-4.8	0.3	
		Locus of control in science	44.1 *	33.8	22.4	17.6	15.2	3.0	
Block 2	Instructional Methods	Frequency of worksheets in class		-4.6	-1.9	-8.4	-7.4	-9.1	
		Frequency of testing		-6.2	3.9	4.2	1.2	-4.8	
		Frequency of calculator use		-49.1 **	-49.6 **	-43.6 **	-45.0 **	-42.5 **	
		Frequency of computer use		-24.7	-18.5	-17.3	-18.3	-15.3	
Block 3	Classroom Organization	Students take notes in class			-15.2 **	-14.6 **	-11.4	-7.5	
		Students do problems in class			-30.3 **	-30.5 **	-33.5 **	-23.9 **	
		Students works in groups in class			12.6	11.0	9.2	6.3	
Block 4	School Climate	Stability of student body				29.3 *	33.6 *	37.4 **	
		Index of major discipline problems				49.8	47.3	7.1	
		Index of minor discipline problems				-22.8	-25.4	8.3	
		Principal leadership				-20.3	-26.4	-40.8 **	
		Teacher's years of experience				7.7 *	7.4 *	6.9 *	
		Students perception of peer attitudes toward science				9.2	10.5	8.1	
Block 5	School Structure	Urban location					4.9	4.1	
		School size					-30.8	-64.0 **	
		Class size					17.7	16.2	
Block 6	Mean SES	Mean socio-economic status of the school						73.9 **	77.1 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.6 Results of seven explanatory models of school effectiveness in science at the fourth grade in Iran

IRAN

	Unconditional Model
Between School Variance	43%
Parameter Variance	1865.8
Reliability	0.87

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	5%	16%	16%	38%	46%	58%	49%
Parameter Variance	1765.6 **	1562.4 **	1572.2 **	1160.6 **	1011.7 **	777.9 **	943.9 **
Reliability	0.87	0.85	0.85	0.82	0.80	0.75	0.78

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
	Intercept	Average Achievement	418.1 **	421.7 **	421.2 **	416.7 **	406.5 **	414.3 **	417.8 **
Control	Student SES	Student socio-economic status (SES)	3.9	3.9	3.8	3.0	2.7	1.1	1.1
Block 1	Student Involvement	Time spent studying science	1.2	-0.3	1.0	-0.3	-1.6	2.5	
		How much students like science	-18.5	-21.0	-21.8	-23.8 *	-19.5	-15.2	
		Locus of control in science	34.6 *	15.0	19.6	-4.4	-3.0	-1.9	
Block 2	Instructional Methods	Frequency of worksheets in class		-21.2	-20.6	-22.1	-24.2 *	-17.3	
		Frequency of testing		23.7 *	20.8	22.6 *	16.3	15.0	
		Frequency of calculator use		-23.7 *	-21.9 *	-13.5	-7.7	-8.2	
		Frequency of computer use		-3.7	-1.8	-2.9	-6.6	-8.2	
Block 3	Classroom Organization	Students take notes in class			-11.4	2.5	2.5	-2.9	
		Students do problems in class			18.6	18.3	18.4	14.4	
		Students works in groups in class			-4.3	-3.8	3.1	9.1	
Block 4	School Climate	Stability of student body				0.3	4.1	34.0	
		Index of major discipline problems				-353.2	-274.6	-97.4	
		Index of minor discipline problems				-283.2	-303.8	-241.1	
		Principal leadership				-7.5	24.7	38.3	
		Teacher's years of experience				18.8 **	15.7 **	8.0 *	
		Students perception of peer attitudes toward science				-24.7 **	-21.3 **	-17.0 *	
Block 5	School Structure	Urban location					25.6 **	11.3	
		School size					-26.6	-48.9	
		Class size					98.8	81.5	
Block 6	Mean SES	Mean socio-economic status of the school						68.0 **	98.7 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.7 Results of seven explanatory models of school effectiveness in science at the fourth grade in Ireland

IRELAND

	Unconditional Model
Between School Variance	29%
Parameter Variance	1695.6
Reliability	0.80

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	23%	45%	46%	46%	47%	48%	19%
Parameter Variance	1300.9 **	931.8 **	911.3 **	918.9 **	891.9 **	882.5 **	1370.2 **
Reliability	0.77	0.72	0.71	0.71	0.71	0.70	0.78

Explanatory Block		Independent Variables	Model 1 Betas	Model 2 Betas	Model 3 Betas	Model 4 Betas	Model 5 Betas	Model 6 Betas	SES Model Betas
	Intercept	Average Achievement	538.3 **	539.6 **	539.5 **	538.6 **	542.2 **	544.0 **	541.3 **
Control	Student SES	Student socio-economic status (SES)	45.8 **	45.1 **	45.0 **	44.9 **	45.4 **	44.7 **	44.7 **
Block 1	Student Involvement	Time spent studying science	-17.7	-13.3	-8.1	-8.2	-6.7	-5.2	
		How much students like science	20.5 *	29.7 **	25.9 **	23.5 *	27.8 **	27.7 **	
		Locus of control in science	49.1 **	52.8 **	39.0 *	43.3 *	39.7 *	38.1	
Block 2	Instructional Methods	Frequency of worksheets in class		-7.2	-3.2	-7.0	-8.5	-8.1	
		Frequency of testing		-25.5 **	-27.2 **	-27.1 **	-31.4 **	-29.2 **	
		Frequency of calculator use		-38.8 **	-39.9 **	-30.8 **	-33.1 **	-33.5 **	
		Frequency of computer use		-6.8	-3.8	-6.9	-4.9	-3.4	
Block 3	Classroom Organization	Students take notes in class			-15.2 *	-14.8	-12.2	-11.6	
		Students do problems in class			9.0	9.0	10.0	7.8	
		Students works in groups in class			0.1	3.8	2.1	2.4	
Block 4	School Climate	Stability of student body				7.1	-0.6	0.0	
		Index of major discipline problems				25.8	40.8	33.0	
		Index of minor discipline problems				-70.7	-75.9	-64.6	
		Principal leadership				3.3	-16.6	-16.3	
		Teacher's years of experience				-5.5	-6.2	-5.7	
		Students perception of peer attitudes toward science				6.8	3.0	3.7	
Block 5	School Structure	Urban location					-7.6	-7.8	
		School size					13.5	11.1	
		Class size					-46.7 *	-51.4 *	
Block 6	Mean SES	Mean socio-economic status of the school						16.0	25.5

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.8 Results of seven explanatory models of school effectiveness in science at the fourth grade in Korea

KOREA

	Unconditional Model
Between School Variance	12%
Parameter Variance	342.9
Reliability	0.59

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	43%	48%	49%	47%	47%	58%	55%
Parameter Variance	195.1 **	179.5 **	174.6 **	180.2 **	182.6 **	144.3 **	154.4 **
Reliability	0.46	0.44	0.43	0.44	0.44	0.39	0.40

Explanatory Block		Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	597.4 **	597.4 **	597.4 **	597.3 **	592.9 **	598.0 **	596.6 **
Control	Student SES	Student socio-economic status (SES)	26.5 **	26.5 **	26.4 **	26.2 **	25.7 **	23.3 **	23.3 **
Block 1	Student Involvement	Time spent studying science	-1.6	-2.2	0.1	0.3	-1.8	-0.7	
		How much students like science	4.8	6.6	4.7	4.7	6.3	2.2	
		Locus of control in science	37.0 **	34.9 **	33.8 **	34.1 **	28.9 **	21.1	
Block 2	Instructional Methods	Frequency of worksheets in class		-4.2	-2.6	0.1	-2.0	-6.5	
		Frequency of testing		0.5	3.2	3.0	2.9	4.2	
		Frequency of calculator use		-19.3 **	-20.3 **	-19.6 **	-20.3 **	-19.1 **	
		Frequency of computer use		0.0	0.5	0.3	0.9	2.2	
Block 3	Classroom Organization	Students take notes in class			-6.7	-6.4	-5.5	-3.6	
		Students do problems in class			-8.0	-7.9	-7.2	-7.3	
		Students works in groups in class			0.1	-1.9	-0.8	-1.4	
Block 4	School Climate	Stability of student body				-12.8	-12.5	-9.6	
		Index of major discipline problems				-7.3	-7.5	-3.0	
		Index of minor discipline problems				9.8	8.9	2.6	
		Principal leadership				2.6	3.0	4.5	
		Teacher's years of experience				-2.2	-1.9	-1.5	
		Students perception of peer attitudes toward science				1.4	1.1	3.1	
Block 5	School Structure	Urban location					1.8	-0.3	
		School size					9.5	5.8	
		Class size					13.1	-18.5	
Block 6	Mean SES	Mean socio-economic status of the school						35.6 **	39.1 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.9 Results of seven explanatory models of school effectiveness in science at the fourth grade in Latvia

LATVIA

	Unconditional Model
Between School Variance	59%
Parameter Variance	3445.2
Reliability	0.92

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	20%	35%	34%	34%	36%	35%	1%
Parameter Variance	2746.3 **	2234.6 **	2262.1 **	2265.8 **	2210.9 **	2255.4 **	3403.4 **
Reliability	0.91	0.89	0.89	0.89	0.89	0.89	0.93

Explanatory Block		Independent Variables	Model 1 Betas	Model 2 Betas	Model 3 Betas	Model 4 Betas	Model 5 Betas	Model 6 Betas	SES Model Betas
	Intercept	Average Achievement	510.9 **	515.1 **	516.4 **	516.6 **	509.2 **	509.6 **	510.2 **
Control	Student SES	Student socio-economic status (SES)	20.9 **	20.9 **	20.9 **	20.9 **	20.7 **	20.6 **	20.6 **
Block 1	Student Involvement	Time spent studying science	-41.6 *	-22.5	-26.9	-17.1	-8.7	-8.5	
		How much students like science	63.9 **	46.5 **	48.9 **	50.0 **	53.7 **	53.5 **	
		Locus of control in science	-1.1	-22.4	-28.4	-16.4	-26.8	-26.7	
Block 2	Instructional Methods	Frequency of worksheets in class		-10.9	-13.6	-8.9	-12.1	-11.8	
		Frequency of testing		3.1	4.5	11.2	9.0	8.2	
		Frequency of calculator use		-2.9	6.0	-8.1	-2.6	-1.5	
		Frequency of computer use		-79.7 **	-90.6 **	-67.7 **	-75.9 **	-77.4 **	
Block 3	Classroom Organization	Students take notes in class			-27.0	-22.8	-36.5 *	-36.7 *	
		Students do problems in class			2.8	6.1	5.7	6.1	
		Students works in groups in class			4.2	0.4	-1.2	0.0	
Block 4	School Climate	Stability of student body				-17.5	-14.1	-14.1	
		Index of major discipline problems				17.3	27.4	28.1	
		Index of minor discipline problems				-2.3	-5.0	-5.1	
		Principal leadership				3.8	4.9	5.0	
		Teacher's years of experience				6.1	6.4	6.6	
		Students perception of peer attitudes toward science				-29.5	-17.2	-16.5	
Block 5	School Structure	Urban location					31.0 *	30.4 *	
		School size					7.7	7.4	
		Class size					-13.0	-14.0	
Block 6	Mean SES	Mean socio-economic status of the school						8.7	28.0

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.10 Results of seven explanatory models of school effectiveness in science at the fourth grade in New Zealand

NEW ZEALAND

	Unconditional Model
Between School Variance	52%
Parameter Variance	4045.8
Reliability	0.90

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	30%	50%	52%	59%	57%	69%	58%
		Parameter Variance	2830.2 **	2022.0 **	1939.1 **	1650.3 **	1720.4 **	1259.0 **	1707.8 **
		Reliability	0.87	0.83	0.83	0.80	0.81	0.76	0.81
	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
	Intercept	Average Achievement	525.2 **	529.4 **	529.2 **	523.6 **	525.2 **	531.6 **	526.5 **
Control	Student SES	Student socio-economic status (SES)	38.3 **	38.1 **	38.0 **	37.1 **	37.0 **	34.1 **	34.1 **
Block 1	Student Involvement	Time spent studying science	2.5	21.8	19.6	7.7	7.9	2.2	
		How much students like science	-6.5	5.4	7.8	-2.0	-0.8	-4.2	
		Locus of control in science	86.6 **	97.3 **	98.5 **	69.3 **	68.2 **	23.7	
Block 2	Instructional Methods	Frequency of worksheets in class		-41.3 **	-45.5 **	-33.6 *	-29.3	-34.5 *	
		Frequency of testing		-6.7	-4.7	-10.5	-14.5	16.6	
		Frequency of calculator use		-9.2	-4.6	-11.1	-12.8	-6.3	
		Frequency of computer use		-40.3 **	-45.4 **	-40.0 **	-37.5 **	-37.7 **	
Block 3	Classroom Organization	Students take notes in class			-22.6	-25.4	-25.4	-25.1 *	
		Students do problems in class			15.5	19.9	19.1	3.9	
		Students works in groups in class			25.0	15.1	14.3	9.1	
Block 4	School Climate	Stability of student body				2.6	0.4	-11.4	
		Index of major discipline problems				-193.5 **	-198.6 **	-146.1 **	
		Index of minor discipline problems				15.5	17.1	44.4	
		Principal leadership				3.7	-1.5	-13.9	
		Teacher's years of experience				2.7	2.4	0.1	
		Students perception of peer attitudes toward science				-2.3	-2.4	0.9	
Block 5	School Structure	Urban location					-7.3	-4.8	
		School size					-8.0	-10.4	
		Class size					12.9	-12.7	
Block 6	Mean SES	Mean socio-economic status of the school						92.6 **	118.9 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.11 Results of seven explanatory models of school effectiveness in science at the fourth grade in Portugal

PORTUGAL

	Unconditional Model
Between School Variance	22%
Parameter Variance	1138.7
Reliability	0.79

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	27%	27%	38%	23%	10%	0%	35%
Parameter Variance	827.0 **	833.3 **	705.6 **	880.6 **	1020.0 **	1133.6 **	742.3 **
Reliability	0.74	0.74	0.71	0.75	0.77	0.79	0.72

	Explanatory Block	Independent Variables	Betas	Betas	Betas	Betas	Betas	Betas	Betas
			Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**SES Model**
	Intercept	Average Achievement	476.8 **	475.2 **	475.6 **	478.3 **	469.1 **	472.9 **	475.8 **
Control	Student SES	Student socio-economic status (SES)	30.5 **	29.8 **	29.4 **	28.9 **	28.0 **	27.6 **	27.6 **
Block 1	Student Involvement	Time spent studying science	-26.0	-31.5	-44.0 *	-33.4	-32.4	-26.2	
		How much students like science	3.2	5.4	21.9	-10.9	-10.2	-10.3	
		Locus of control in science	20.1	-4.3	-28.7	26.6	17.9	23.5	
Block 2	Instructional Methods	Frequency of worksheets in class		25.5	45.9 *	43.2	44.3	39.6	
		Frequency of testing		5.6	2.7	-0.7	7.5	1.2	
		Frequency of calculator use		-21.3	-4.1	-8.4	-13.1	-12.7	
		Frequency of computer use		-0.9	-2.0	-2.3	4.0	2.6	
Block 3	Classroom Organization	Students take notes in class			27.6	18.4	1.9	0.4	
		Students do problems in class			-36.0	-37.2	-33.4	-33.0	
		Students works in groups in class			-30.7	-15.4	-11.5	-9.2	
Block 4	School Climate	Stability of student body				4.2	23.2	20.3	
		Index of major discipline problems				-12.6	11.3	23.6	
		Index of minor discipline problems				-57.4	-66.1	-74.4	
		Principal leadership				-0.1	7.9	3.8	
		Teacher's years of experience				-5.1	-8.9	-10.0	
		Students perception of peer attitudes toward science				9.2	10.7	6.1	
Block 5	School Structure	Urban location					11.1	8.4	
		School size					32.9	23.5	
		Class size					-75.5	-71.1	
Block 6	Mean SES	Mean socio-economic status of the school						14.0	34.3 *

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.12 Results of seven explanatory models of school effectiveness in science at the fourth grade in Slovenia

SLOVENIA

	Unconditional Model
Between School Variance	17%
Parameter Variance	834.6
Reliability	0.77

			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
		Explained Variance	20%	25%	30%	24%	27%	44%	39%
		Parameter Variance	669.9 **	629.0 **	583.8 **	633.7 **	611.0 **	465.5 **	505.7 **
		Reliability	0.73	0.72	0.70	0.72	0.71	0.65	0.67
	Explanatory Block	**Independent Variables**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**	**Betas**
	Intercept	Average Achievement	543.3 **	543.8 **	544.3 **	544.2 **	540.8 **	543.8 **	544.3 **
Control	Student SES	Student socio-economic status (SES)	22.5 **	22.6 **	22.6 **	22.4 **	21.9 **	19.6 **	19.6 **
Block 1	Student Involvement	Time spent studying science	2.3	6.4	1.6	-0.4	-7.7	-10.5	
		How much students like science	-13.7	-13.5	-9.9	-5.0	-5.2	-2.3	
		Locus of control in science	15.7	7.6	10.3	5.7	16.5	9.5	
Block 2	Instructional Methods	Frequency of worksheets in class		1.3	4.2	3.5	5.2	-2.6	
		Frequency of testing		-11.1	-10.0	-8.8	-13.7	-8.6	
		Frequency of calculator use		4.4	9.9	14.0	25.8	27.6	
		Frequency of computer use		-28.3	-37.5 *	-45.0 *	-53.1 **	-54.3 **	
Block 3	Classroom Organization	Students take notes in class			-0.1	-0.6	1.5	-2.8	
		Students do problems in class			-25.1 **	-25.8 **	-27.6 **	-24.0 **	
		Students works in groups in class			8.1	9.6	9.7	10.4	
Block 4	School Climate	Stability of student body				-2.3	1.7	0.4	
		Index of major discipline problems				-1.3	1.3	-3.1	
		Index of minor discipline problems				2.3	2.9	4.4	
		Principal leadership				-2.7	-0.5	0.0	
		Teacher's years of experience				0.1	-0.9	-1.7	
		Students perception of peer attitudes toward science				-15.4	-10.2	-9.5	
Block 5	School Structure	Urban location					5.1	4.1	
		School size					-1.1	-2.4	
		Class size					13.6 *	5.1	
Block 6	Mean SES	Mean socio-economic status of the school						55.6 **	57.3 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.

Table C.13 Results of seven explanatory models of school effectiveness in science at the fourth grade in the United States

UNITED STATES

	Unconditional Model
Between School Variance	42%
Parameter Variance	3271.9
Reliability	0.94

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	SES Model
Explained Variance	29%	49%	49%	66%	73%	83%	69%
Parameter Variance	2317.4 **	1670.3 **	1666.0 **	1103.8 **	879.1 **	549.9 **	1018.4 **
Reliability	0.92	0.89	0.89	0.85	0.82	0.74	0.84

Explanatory Block		Independent Variables	Model 1 Betas	Model 2 Betas	Model 3 Betas	Model 4 Betas	Model 5 Betas	Model 6 Betas	SES Model Betas
	Intercept	Average Achievement	556.8 **	562.7 **	562.1 **	566.8 **	579.6 **	576.0 **	562.0 **
Control	Student SES	Student socio-economic status (SES)	26.5 **	26.5 **	26.3 **	26.0 **	26.0 **	24.4 **	24.4 **
Block 1	Student Involvement	Time spent studying science	-24.6 *	0.0	-2.6	-7.6	-13.4	4.3	
		How much students like science	11.9	9.8	9.0	-5.1	-9.4	-0.8	
		Locus of control in science	90.7 **	93.4 **	94.9 **	83.2 **	87.1 **	39.1	
Block 2	Instructional Methods	Frequency of worksheets in class		18.7	25.6	16.8	24.8 *	1.3	
		Frequency of testing		-33.7 **	-45.2 **	-36.1 **	-33.1 **	-19.1 **	
		Frequency of calculator use		-35.6 **	-35.5 **	-30.4 **	-29.0 **	-25.3 **	
		Frequency of computer use		-23.9 **	-24.4 **	-9.5	-11.4	-5.6	
Block 3	Classroom Organization	Students take notes in class			19.9 *	12.5	8.6	4.9	
		Students do problems in class			-3.9	-2.7	0.5	5.2	
		Students works in groups in class			-1.8	-2.5	5.3	-0.9	
Block 4	School Climate	Stability of student body				376.7 **	334.6 **	114.7	
		Index of major discipline problems				73.1	39.1	41.7	
		Index of minor discipline problems				-354.5 **	-267.1 *	-229.4 *	
		Principal leadership				-49.2	-229.0	-72.2	
		Teacher's years of experience				0.5	0.5	1.0	
		Students perception of peer attitudes toward science				16.5	13.0	15.1	
Block 5	School Structure	Urban location					-15.7 **	-8.9	
		School size					-184.0 **	-117.5 **	
		Class size					-87.8	-137.0	
Block 6	Mean SES	Mean socio-economic status of the school						74.5 **	115.6 **

** results statistically significant at the .05 alpha level
* results statistically signficant at the .10 alpha level

SOURCE: IEA Third International Mathematics and Science Study (TIMSS), 1994-95.