2008

The Effects of Student-Level and Classroom-Level Factors on Elementary Students' Science Achievement in Five Countries

Sibel Kaya
THE EFFECTS OF STUDENT-LEVEL AND CLASSROOM-LEVEL FACTORS
ON ELEMENTARY STUDENTS’ SCIENCE ACHIEVEMENT
IN FIVE COUNTRIES

By

SIBEL KAYA

A Dissertation submitted to the
School of Teacher Education
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Fall Semester, 2008

Copyright © 2008
Sibel Kaya
All Rights Reserved
The members of the Committee approve the Dissertation of Sibel Kaya defended on October 21, 2008.

Diana C. Rice  
Professor Directing Dissertation

Robert Schwartz  
Outside Committee Member

Ithel Jones  
Committee Member

Carol Connor  
Committee Member

Approved:

Walt Wager, Chair, School of Teacher Education

The Office of Graduate Studies has verified and approved the above named committee members.
AKNOWLEDGEMENTS

I would like to thank my dissertation director Dr. Diana C. Rice. I appreciate the time and patience she devoted toward my study and the feedbacks that she gave on the numerous drafts of this dissertation. She has been a wonderful mentor and I am very grateful for everything she has done for me. Most of all, thank you Dr. Rice for your friendship and humor 😊

Next, I would like to thank Dr. Carol Connor for her help and guidance in every step of my statistical analysis and interpretations. Her expertise and feedbacks made the process easier. Also, my sincere thanks to my other committee members, Dr. Ithel Jones and Dr. Robert Schwartz, for their invaluable insights and criticisms throughout my study.

I would like to thank Dr. Salih Binici for his help and insights in statistical analysis of my data. Many thanks to my friend and co-worker Meghann Montgomery for editing my manuscript. I am also thankful to my friends and co-workers at Florida Center for Reading Research for believing in me and supporting me.

Most importantly, I would like to send my deepest appreciation to my family, my dad, my mom, my brother and other family members for their unconditional love and support for seven years form thousands of miles away.

Finally, my sincere appreciation to the Turkish Ministry of Education for providing me the scholarship which made my graduate studies possible. Studying in the United States has been a wonderful experience for me both academically and culturally.
# TABLE OF CONTENTS

List of Tables ........................................................................................................ vi
List of Figures .......................................................................................................... viii
Abstract ................................................................................................................... ix

1. INTRODUCTION ................................................................................................. 1
   Purpose of the Study ........................................................................................... 4
   Research Questions ............................................................................................ 4
   Statement of the Problem .................................................................................. 4
   Significance of the Study ................................................................................... 6
   Conceptual Model ............................................................................................... 7
   Definitions ........................................................................................................... 9

2. LITERATURE REVIEW ......................................................................................... 11
   School Learning Models ..................................................................................... 11
   Student-Level Factors and Achievement ........................................................ 14
      Gender ............................................................................................................ 14
      Self-Efficacy .................................................................................................. 16
      Home Resources ............................................................................................. 18
   Classroom-Level Factors and Achievement ....................................................... 20
      Teacher Characteristics .................................................................................. 21
      Teacher’s Experience ..................................................................................... 21
      Teacher’s Education and Subject Major ......................................................... 23
      Teacher Support ............................................................................................. 25
   Instructional Variables ....................................................................................... 27
      Science Inquiry ............................................................................................... 27
      Amount of Science Instruction ..................................................................... 31
      Class Size ....................................................................................................... 33
   The Trends in International Mathematics and Science Study (TIMSS) 2003 .. 36
   Instrument Development and Test Administration ............................................ 38
   Background Questionnaires ............................................................................. 40
   Sampling Procedure ......................................................................................... 40
   Scaling Procedure ............................................................................................. 41

3. METHODOLOGY .................................................................................................. 43
   Data Source ....................................................................................................... 44
   Sample ............................................................................................................... 44
   Sampling Weights ............................................................................................. 45
   Plausible Values ............................................................................................... 45
   Variables ............................................................................................................ 46
   Data Analysis ..................................................................................................... 52
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Education Statistics for Five Countries</td>
<td>2</td>
</tr>
<tr>
<td>2.1</td>
<td>TIMSS 2003 Fourth Grade Science Results</td>
<td>37</td>
</tr>
<tr>
<td>3.1</td>
<td>List of Variables</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>Factor Loadings and Item Means and Standard Deviations for the Measure of Science Inquiry</td>
<td>50</td>
</tr>
<tr>
<td>4.1</td>
<td>Descriptive Statistics of Variables for Five Countries</td>
<td>62</td>
</tr>
<tr>
<td>4.2</td>
<td>Maximum Likelihood Estimates of the Variance Components at the Student- and Classroom-Level</td>
<td>63</td>
</tr>
<tr>
<td>4.3</td>
<td>The Proportion of Variance in Science Achievement Within and Between Classrooms</td>
<td>65</td>
</tr>
<tr>
<td>4.4</td>
<td>The Effects of Student-Level Variables on Science Achievement Within Countries</td>
<td>66</td>
</tr>
<tr>
<td>4.5</td>
<td>The Effects of Student- and Classroom-Level Variables on Science Achievement Within Countries</td>
<td>69</td>
</tr>
<tr>
<td>4.6</td>
<td>Average TIMSS 2003 Science Scores by Student-Level Home Resources</td>
<td>71</td>
</tr>
<tr>
<td>4.7</td>
<td>Explained Variances in Science Achievement Within Countries</td>
<td>72</td>
</tr>
<tr>
<td>4.8</td>
<td>Student-Level Effects Across Countries</td>
<td>75</td>
</tr>
<tr>
<td>4.9</td>
<td>Classroom-Level Effects Across Countries</td>
<td>77</td>
</tr>
<tr>
<td>4.10</td>
<td>Hypothesis Testing of Student- and Classroom-Level Effects Across Countries</td>
<td>78</td>
</tr>
<tr>
<td>5.1</td>
<td>Significant Variables at the Student- and Classroom-Level Within Countries</td>
<td>81</td>
</tr>
<tr>
<td>5.2</td>
<td>Significant Variables at the Student- and Classroom-Level Across Countries</td>
<td>82</td>
</tr>
<tr>
<td>A.1</td>
<td>Inter-Correlation Matrix Between Variables for Singapore: Results of the Bivariate Analysis</td>
<td>96</td>
</tr>
</tbody>
</table>
Table A.2  Inter-Correlation Matrix Between Variables for Japan: Results of the Bivariate Analysis ................................................................. 96

Table A.3  Inter-Correlation Matrix Between Variables for the U.S.: Results of the Bivariate Analysis ................................................................. 97

Table A.4  Inter-Correlation Matrix Between Variables for Australia: Results of the Bivariate Analysis ................................................................. 97

Table A.5  Inter-Correlation Matrix Between Variables for Scotland: Results of the Bivariate Analysis ................................................................. 98
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Conceptual Model of the Relationship Between Science Achievement and Student-Level and Classroom-Level Factors</td>
<td>9</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Major Variables in Bloom’s Theory of School Learning</td>
<td>12</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>A School Learning Model</td>
<td>13</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Factorial Structure of Science Inquiry</td>
<td>51</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Home Resources by Class Mean Self-Confidence Interaction Predicting Science Achievement in the U.S.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Home Resources by Class Mean Home Resources Interaction Predicting Science Achievement in Singapore</td>
<td>74</td>
</tr>
</tbody>
</table>
ABSTRACT

The interest in raising levels of achievement in math and science has led to a focus on investigating the factors that shape achievement in these subjects (Lamb & Fullarton, 2002) as well as understanding how these factors operate across countries (Baker, Fabrega, Galindo, & Mishook, 2004). The current study examined the individual student factors and classroom factors on fourth grade science achievement within and across five countries. Guided by the previous school learning models, the elements of students’ science learning were categorized as student-level and classroom-level factors. The student-level factors included gender, self-confidence in science, and home resources. The classroom-level factors included teacher characteristics, instructional variables and classroom composition. Results for the United States and four other countries, Singapore, Japan, Australia, and Scotland were reported. Multilevel effects of student and classroom variables were examined through Hierarchical Linear Modeling (HLM) using the Trends in International Mathematics and Science Study (TIMSS) 2003 fourth grade dataset. The outcome variable was the TIMSS 2003 science score.

Overall, the results of this study showed that selected student background characteristics were consistently related to elementary science achievement in countries investigated. At the student-level, higher levels of home resources and self-confidence and at the classroom-level, higher levels of class mean home resources yielded higher science scores on the TIMSS 2003. In general, teacher and instructional variables were minimally related to science achievement. There was evidence of positive effects of teacher support in the U.S. and Singapore. The emphasis on science inquiry was positively related to science achievement in Singapore and negatively related in the U.S. and Australia. Experimental studies that investigate the impacts of teacher and instructional factors on elementary science achievement are needed. For all the countries investigated, with the exception of Singapore, the between-class variance was much smaller than the within-class variance. Japan had the smallest variation in science achievement among classrooms which indicates the homogeneity across classrooms in Japan.

Increasing awareness and knowledge of gender neutral instructional techniques, providing a non-threatening, rich and supportive environment for both genders in classrooms by elementary teachers are to be encouraged. To improve students’ self beliefs about science,
it is recommended that teachers model science activities and accommodate students’ needs and abilities (Bandura, 1997; Britner & Pajares, 2006). Schools and teachers are recommended to develop a successful home-school partnership for improved student learning and positive attitudes toward science (Eccles & Harold, 1996; Epstein & Salinas, 2004). Furthermore, developing a knowledge base for teachers regarding the influences of classroom and school composition is highlighted (Honig, Kahne, & McLaughlin, 2001; Murrel, 2001). At the classroom- and school-level, policy efforts could focus on the distribution of educational resources (Condron & Roscigno, 2003; Goesling, 2003) to compensate for poor family background.
CHAPTER 1

INTRODUCTION

In recent years, a reform in science education in the United States has been highlighted by influential policy reports (Council of Chief State School Officers [CCSSO], 2006; National Academy of Sciences, 2006; National Science Board, 2007; U.S. Department of Education, 2000). It was emphasized that students in the U.S. are falling behind in science and mathematics compared to other developed countries. Improving K-12 science and mathematics education was the top priority of a recent report by National Academy of Sciences (2006). The report underlined the significance of receiving quality academic training for all students at all levels in science, technology, engineering, and mathematics (STEM) to be compatible in global economy.

The interest in raising levels of achievement in math and science has led to a focus on investigating the factors that shape achievement in these subjects (Lamb & Fullarton, 2002) as well as understanding how these factors operate across countries (Baker, Fabrega, Galindo, & Mishook, 2004). This study was an attempt to expand the research on student-level and classroom-level effects on science achievement in elementary schools within and across five different countries. The student-level factors, gender, home resources, and science self-confidence and classroom-level factors including teacher and instructional variables and classroom composition on science achievement were investigated by using a large scale international data set, the Trends in Mathematics and Science Study (TIMSS) 2003.

The Trends in Mathematics and Science Study (TIMSS), sponsored by the International Association for the Evaluation of the Education Achievement (IEA) measures mathematics and science achievement of nationally representative samples of students and collected background information from students, their teachers and schools. The TIMSS 2003 is the third in a continuing cycle of international mathematics and science assessments conducted every four years. Forty-eight countries participated in the TIMSS 2003 at the eighth grade level, and 26 countries participated at the fourth grade level. Secondary analyses of the TIMSS 2003 assessment were recommended to help individual countries to better understand their standings in mathematics and science, to
identify their strengths and weaknesses, and find ways to improve their educational systems (Martin & Mullis, 2004).

The TIMSS 2003 is a large international dataset. For handling the data more effectively, instead of using the whole dataset, four other countries along with the U.S. were selected. Two of these countries, Singapore and Japan, scored significantly higher than the U.S., and the other two, Australia and Scotland, scored significantly lower. Singapore was selected because it was the top scoring country. Japan was selected because it also scored high and it is a country that was often compared to the United States. Australia was selected as an English speaking country with a decentralized education system which is similar to the United States (Bray, 2007). Similarly, in Scotland, education is the responsibility of local authorities and schools. Scotland does not have a proscribed national curriculum though schools are expected to follow national guidelines (The Scottish Government, 2008).

Table 1.1 presents some education statistics for five countries according to the United Nations Educational Scientific and Cultural Organizations (UNESCO) statistics (2006).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (million)a</td>
<td>4,839</td>
<td>127,953</td>
<td>302,841</td>
<td>20,536</td>
<td>5,144</td>
</tr>
<tr>
<td>Youth Literacy (%)</td>
<td>99.7</td>
<td>98</td>
<td>99.5</td>
<td>98</td>
<td>99.5</td>
</tr>
<tr>
<td>Primary School Enrollment (%)b</td>
<td>96</td>
<td>99</td>
<td>92</td>
<td>96</td>
<td>98</td>
</tr>
<tr>
<td>Secondary School Enrollment (%)</td>
<td>76</td>
<td>99</td>
<td>88</td>
<td>88</td>
<td>92</td>
</tr>
<tr>
<td>Tertiary School Enrollment (%)c</td>
<td>39</td>
<td>57</td>
<td>82</td>
<td>73</td>
<td>59</td>
</tr>
<tr>
<td>Public Expenditure on Education as % of Total Gov. Exp.</td>
<td>18.2</td>
<td>9.2</td>
<td>13.2</td>
<td>13.3</td>
<td>12.5</td>
</tr>
</tbody>
</table>

   Data for Scotland is available at http://www.scotland.org/about/fact-file/population/
b. Percentages of children in school are represented by Gross Enrolment Ratios (GER) and Net Enrolment Ratios (NER). GER is the number of pupils enrolled in a given level of education regardless of age expressed as a percentage of the population in the theoretical age group for that level of education. NER is the number of pupils in the theoretical age group who are enrolled expressed as a percentage of the same population.
c. Data for Singapore is available at http://earthtrends.wri.org/pdf_library/country_profiles
Research investigating the effects of student factors on science achievement reported that students tend to perform better when they have higher levels of science self-efficacy (Andrew, 1998; Britner & Pajares, 2001, 2006; Gwilliam & Betz, 2001; House, 2004, 2008; Lau & Roeser, 2002) and more educational resources at home (Iverson & Walberg, 1982; Roscigno & Ainsworth-Darnell, 1999; Xin, Xu, & Tatsuoka, 2005) compared to their counterparts. In terms of gender effects on science achievement, the gap between boys and girls at the elementary school level was less pronounced than was the middle and high school level (Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996b; Greenfield, 1995; Kahle, 2004; Solomon, 1997).

Studies examining the effects of teacher characteristics on student achievement reported mixed results. Some indicated positive influences of teacher characteristics, such as experience (Fetler, 1999; Greenwald, Hedges, & Lane, 1996), subject major (Darling-Hammond, 2000; Goldhaber & Brewer, 1998, 2000) and education (Darling-Hammond) on students’ math and science achievement, others reported no relations between teachers’ experience (Ferguson & Ladd, 1996; Goldhaber & Brewer; Monk, 1994), education (Croninger, Rice, Rathbun, Nishio, 2007; Goldhaber & Brewer, 2000; Wenglinsky, 2002; Xu & Gullosino, 2006) and student achievement. In general, students tended to have higher levels of achievement (Brown, Anfara, & Roney, 2004; Klem & Connell, 2004) and motivation (Klem & Connell; Marks, 2000; Skinner & Belmont, 1993) when their teachers were more supportive.

In terms of instructional factors, the emphasis on science inquiry improved students’ science achievement (House, 2006; Lee, Deaktor, Hart, Cuevas, & Enders, 2006; Paris, Yambor, & Packard, 1998; Stohr-Hunt, 1996). The amount of time spent on subject content, in general, positively affected student achievement (Bodovsky & Farkas, 2007; Coates, 2003; Lewin, 1993; Fuller, 1987). However, the results were inconclusive for studies investigating the relations between the amount of science instruction and science achievement (Baker et al., 2004; Coates). Finally, while positive effects of smaller class-size on student achievement were documented (Finn & Achilles, 1999; Mitchell & Mitchell, 1999; Smith, Molnar & Zahorik, 2003), internationally, research suggested little evidence of the effect of class size on student achievement (Pong & Pallas, 2001; Woessman & West, 2002).
Purpose of the Study

The purpose of this study was to explore the effects of student- and classroom-level factors on students’ science achievement at the elementary level within and across five countries. Student-level factors include gender, self-confidence in science, and home resources. Classroom-level factors include teacher characteristics (teacher’s experience, education, major, and teacher support), instructional variables (science inquiry, minutes of weekly science instruction, and class size), class mean of home resources and class mean of self-confidence. Possible similarities and differences among five countries that participated in the TIMSS 2003 were examined. The hierarchical linear modeling (HLM) analysis was used seeking the answers to the research questions.

Research Questions

The research questions for this study are as follows:

1. What student-level factors are significantly related to science achievement at fourth grade level in Singapore, Japan, the U.S., Australia, and Scotland?
2. What classroom-level factors are significantly related to science achievement at fourth grade level in Singapore, Japan, the U.S., Australia, and Scotland?
3. How much of the variance in student achievement is explained by student- and classroom-level factors within and across countries?

Statement of the Problem

Over the past few years, many reports have criticized the competitiveness of the U.S. in global economy (see Council of Chief State School Officers [CCSSO], 2006; National Academy of Sciences, 2006; National Science Board, 2007; U.S. Department of Education, 2000). In fall 2005, the National Academy of Sciences, released its report entitled, ‘Rising above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future’. The committee identified and compiled a list of ‘worrisome indicators’ of diminishing global supremacy of the United State and made recommendations for federal action. Improving mathematics and science education at all levels was at the top of the list of recommendations.
The committee pointed out that scientific and technological development critical for the economic leadership of the United States might be at stake as many other nations gain strength. Since the vast majority of jobs in the 21st century will require higher levels of scientific and technological skills, it is essential for students to receive quality academic training in science, technology, engineering, and mathematics (STEM) to retain the best jobs in the United States. After all, the nation’s ability to advance in science and technology will be relatively limited without competence in these fields. It is highlighted that starting from early stages of schooling, public education should lay a foundation of gaining scientific and mathematical literacy (National Academy of Sciences, 2006).

In an earlier report by the U.S. National Commission on Mathematics and Science Teaching for the 21st Century, also known as the Glenn Commission, mathematics and science instruction in the U.S. were harshly criticized (U.S. Department of Education, 2000). The report pointed out that 78 percent of the twelfth graders failed to demonstrate more than a ‘basic’ level in science achievement on the 1996 National Assessment of Educational Progress (NAEP). The need to improve the preparation and practice of the U.S. science teachers was emphasized. In 2005 National Assessment of Educational Progress (NAEP), 32 percent of fourth grade students, 41 percent of eighth grade students, and 46 percent of twelfth grade students scored below the ‘basic’ level in science (National Center for Education Statistics, 2006).

Internationally, the performance of the U.S. students in mathematics and science has long been scrutinized. For example, in the Trends in International Math and Science Study (TIMSS), which assesses fourth and eighth grade students on math and science, average scores and the ranking of U.S. fourth graders dropped between 1995 and 2003 (Martin, Mullis, Gonzalez, & Chrostowski, 2004). The percentage of the U.S. fourth grade students achieving the ‘Advanced International Benchmark’ in science decreased from 19 to 13 percent between the 1995 and 2003 administrations of the test (Martin et al.).

The Program for International Student Assessment (PISA), given every year, measures 15-year-old students' ability to apply scientific and mathematical concepts and thinking skills to problems they might encounter, particularly in situations outside of a classroom. In the PISA 2006, the United States had an average score of 489 on the combined science literacy scale, lower than the Organization for Economic Co-operation
and Development (OECD) average score of 500. In relative terms, the U.S. students ranked 21st among 30 OECD jurisdictions. Eight of the 27 non-OECD jurisdictions also scored higher than the U.S. (Baldi, Jin, Skemer, Green, Herget, & Xie, 2007).

With the ever growing emphasis on science, technology, engineering, and mathematics (STEM) worldwide, it is worthwhile to examine more closely the factors that may contribute to students’ achievement in these subjects within and across nations. Such research is especially important if appropriate steps to improve students’ achievement in these subjects are to be undertaken in early stages of schooling.

Significance of the Study

Current political demands for maximizing student learning outcomes and accountability in education require the attention of educational researchers (National Education Association, 2007). As the testing of science mandated by No Child Left Behind Act of 2001 was implemented in 2007-2008 teachers are recommended to improve their instruction in science (U.S. Department of Education, 2004). Little research has been published addressing the relations between backgrounds and practices of science teachers and student outcomes. Considering the amount of research on mathematics and reading achievement (see Fetler, 1999, 2001; Darling-Hammond, 2000; Goldhaber & Brewer, 2000; Hill, Rowan, & Ball, 2005; Wayne & Young, 2003), the need to explore the effects of teacher characteristics and instructional variables on science achievement seems justified.

Furthermore, most of the existing research on classroom effects has focused on middle and secondary levels. Because of the growing recognition of the importance of developing scientific literacy skills starting from the earliest grades (National Academy of Sciences, 2006; National Research Council, 1996; American Association for Advancement of Science, 1993) there is a need to expand the work to elementary grade levels. This study examined the student- and classroom-level variables in U.S. elementary science classrooms in comparison with four other countries. Thus, the results might shed some light on actions to be taken for increasing the science achievement at early grades.

Many studies regarding classroom effects on student achievement ignore qualitative factors, such as instructional variables and teacher support. Research suggests that these factors might be highly conducive to student achievement (Connor, Son,
Hindman, & Morrison, 2005; Cohen, Raudenbush, & Ball, 2003). Wenglinsky’s (2000) analysis reported that instructional variables when combined with teacher characteristics can be as influential as student and family variables. If this is the case, then instructional variables may explain a sizeable portion of the variance in student achievement (Cochran-Smith, 2002; Connor et al; Cohen et al.). The current study seeks to explore this possibility. By adding instructional variables and teacher support to other teacher characteristics, it is intended to measure more precisely the contribution of classroom variables to students’ learning.

The TIMSS 2003 data set used in this study has a nested data structure, in which students are nested in classrooms and classrooms are nested in schools. Previous studies on the effects of classroom-level factors on science achievement were limited in their use of appropriate analytical methods to take into account nested data structures (Xin et al., 2004). Therefore the study employs a multilevel analysis that considers the nested data structure which provides an opportunity to analyze the data more precisely at student- and classroom-levels (Raudenbush & Bryk, 2002).

Using a comprehensive international data set, this study may help policy makers and practitioners to assess their comparative status and thus, to derive implications for improving science education. Data from the U.S. was compared with the data from two countries that scored significantly higher than the U.S. and two countries that scored significantly lower. Data from Singapore, the top scoring country in TIMSS 2003 fourth grade science; Japan is a country that has always been compared to and used as a benchmark by the U.S.; and Australia and Scotland, as English speaking countries that scored lower than the U.S. were analyzed.

**Conceptual Model**

There is a variety of school learning models that have been developed to organize the student-, classroom-, and school-based variables related to student learning. Unfortunately, no single model can completely explain ‘how learning occurs’ with all the aspects of the teaching/learning process. It is apparent, however, that the learner characteristics and classroom variables are the crucial components of this process.

Learner characteristics include gender (Biggs & Moore, 1993), socioeconomic status (SES) (Biggs & Moore; Walberg, 1986), and students’ affective characteristics,
such as academic self beliefs (Carroll, 1963; Biggs & Moore; Bloom, 1976). Other direct impacts on measures of school learning are of those variables related to the classroom characteristics. This category includes teacher characteristics (i.e., experience and knowledge) (Biggs & Moore; Carroll) and instructional variables including quality (Carroll; Biggs & Moore; Bloom; Walberg), quantity of instruction (Carroll; Walberg) and class size (Biggs & Moore).

To understand better how student and classroom factors relate to science achievement the following conceptual model was developed based on the previous school learning models (Figure 1.1). Within this model, student and home background factors were clustered under student-level factors while teacher and instructional factors were clustered under classroom-level factors. For student-level factors, gender, home resources, and science self-confidence and for classroom-level factors teacher characteristics (experience, education, major, and support) and instructional variables (science inquiry, minutes of weekly science instruction, and class size) were selected. Classroom composition which is measured by the class mean of home resources and class mean of science self-confidence was also included at the classroom-level. As Figure 1.1 indicates, both student-level and classroom-level factors directly effect students’ science outcome. It is further hypothesized that student-level factors are mediated by classroom-level factors. Details of the previous learning models that guided the study as well as literature review of the variables were discussed in Chapter 2.
Definitions

The Trends in Mathematics and Science Study (TIMSS) 2003: The TIMSS 2003, sponsored by the International Association for the Evaluation of Education Achievement (IEA), measured mathematics and science achievement of nationally representative samples of students and collected background information from students, their teachers and schools in 49 countries.

Elementary School Students: For this study, the students enrolled in the upper grade of the two adjacent grades that contained the largest proportion of 9-year-olds considered as elementary school students. They correspond to the fourth graders in practically every country participated in the TIMSS 2003 (Martin et al., 2004).

Self-Efficacy: “People’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” (Bandura, 1986, p.391). In the TIMSS 2003, the self-efficacy (i.e., self-confidence) index is based on students’ responses to four statements about their science ability: (a) I usually do well in science; (b) Science is more difficult for me than for many of my classmates; (c) Science is not one of my strengths; (d) I learn things quickly in science.

Home Resources: Derived variable based on three TIMSS student questionnaire items related to the availability of home resources: (a) about how many books are there in your
home? Do you have these items at home (b) computer, and (c) study desk/table for your 
own use?

**Teacher’s Experience:** Teacher’s response to TIMSS teacher questionnaire item: 
Number of years as a teacher.

**Teacher’s Education:** Teacher’s response to TIMSS teacher questionnaire item: Highest 
level of formal education completed.

**Teacher’s Major:** Teachers’ reports about having a science or mathematics major.

**Teacher Support:** Teacher support generally involves characteristics such as caring, 
friendliness, understanding, dedication, and dependability (Ryan & Patrick, 2001). For 
the measure of teacher support, this study used the composite score of students’ responses 
to the following statements: (a) teacher cares about the students, and (b) teacher wants 
students to do their best.

**Science Inquiry:** The National Science Standards (NRC, 1996) describe science inquiry 
as “the activities of students in which they develop knowledge and understanding of 
scientific ideas, as well as an understanding of how scientists study the natural world” (p. 
23). As the measure of science inquiry, the study utilized the composite score of students’ 
responses to the statements regarding science inquiry activities, such as ‘We do science 
experiments or investigations’, and ‘We work in groups to do experiments or 
investigations.’

**Class Mean of Science Self-Confidence:** Mean value of students’ self-confidence in 
science for each classroom. This value is computed by dividing the total score of 
students’ self-confidence in each classroom by the number of students in that classroom.

**Class Mean of Home Resources:** Mean value of students’ home resources for each 
classroom. This value is computed by dividing the total score of students’ home resources 
in each classroom by the number of students in that classroom.
CHAPTER 2

LITERATURE REVIEW

The first section of this chapter is a description of the school learning models that guided the development of the conceptual model for the study. Next, student-level variables namely, gender, self-efficacy, and home resources which potentially affect student achievement were briefly discussed. Even though the study used science self-confidence as a student-level variable, ‘self-efficacy’ was the most commonly used term in the literature. Therefore, studies that examined self-efficacy with regards to science achievement were reviewed. The third section contains reviews of empirical studies that are related to classroom-level teacher characteristics and instructional variables. Finally, the characteristics of the TIMSS 2003 database were presented.

School Learning Models

To explain trends in student outcomes and investigate systemic factors that have impact on groups or individuals, researchers rely on previously developed learning models. John Carroll (1963) and Benjamin Bloom (1976) developed prominent learning models addressing various factors. Students’ cognitive abilities, motivation and self-concept, quantity and quality of instruction were the major elements affecting student achievement in these models. In his model of Mastery Learning, Carroll suggested that time is the basic component of learning process and the time needed by learners to master the academic content is contingent upon three factors: ability, aptitude, and the quality of instruction. Carroll pointed out that quality of instructional materials, teachers’ subject matter knowledge and their ability to diagnose learner difficulties are important aspects of instructional quality. He did not include the influences of family, community, or society.

Inspired by Carroll’s work, Bloom (1976) developed the Theory of School Learning. Accordingly, the major variables of school learning were learner characteristics, instruction and learning outcomes (see Figure 2.1). In this model, Bloom stated that “… variations in learning and the level of learning of students are determined by the students’ learning history and the quality of instruction they receive.” (p.16)
Student characteristics were defined by cognitive and affective entry behaviors. Cognitive entry behaviors referred to prior knowledge of students needed for a specific learning task. Affective entry behaviors referred to students’ attitudes toward school and specific school subjects, as well as academic self-concept. Bloom (1976) stated that teachers are responsible for providing the quality of instruction which will help students to master the learning task. Critical elements of the quality of instruction included clear communication of the learning expectations, giving feedback, and providing additional time and help as needed by students (Bloom).

Guided by Carroll’s and Bloom’s theories, Harvey Walberg (1986) developed a learning model that included nine factors that play important role in children’s learning. Walberg classified these factors in three groups: aptitude factors, which included ability, level of maturation/development, and motivation to learn; instruction factors, which included the amount and quality of instruction; and environment factors, which included the home environment, classroom environment, peer influences, and the amount of television viewed by a child.

The aptitude and instruction factors were prominent in Carroll and Bloom’s models. Walberg (1986) added four environment factors which were ignored by both Carroll and Bloom. By doing so, Walberg indicated that learning takes place within a social context of the media, home, classroom, and peer group. Interdependence of these nine factors was one of the major features of Walberg’s model. This interdependence
suggested that efforts to improve student outcome might have a better chance of success because changing one factor simultaneously affects other factors (Reynolds & Walberg, 1992).

More recently, Biggs and Moore (1993) developed a School Learning Model underlying the effects of teaching and learner characteristics besides teaching/learning process on learning outcomes within a classroom context (see Figure 2.2). Biggs and Moore’s model is also called ‘3P Model’ with the components of Presage (learner and teaching characteristics), Process (teaching-learning process) and Product (outcomes).

![Figure 2.2 A School Learning Model (adapted from Biggs & Moore, 1993)](image)

Learner presage factors include student characteristics, such as ability, SES, gender, ethnicity, conception of learning, and interest in task. Teaching presage factors include teacher’s personality and expertise, curriculum, method of teaching, and classroom climate. The teaching/learning process consist all the interactions between student and teaching presage factors. Biggs and Moore, (1993) suggested that students learn better in smaller classrooms where teachers vary their teaching methods and focus on student-driven learning through cooperative group work and ensure that there is a warm classroom climate. The learning outcome may be described quantitatively (how much is learned), qualitatively (how well is learned), or institutionally in which both quantitative and qualitative aspects are considered in deriving student grades.

Presage factors can directly affect outcomes as well as they can be mediated by teaching-learning process. For example, more time on task improves learning outcomes
and it also affects the process (Biggs & Moore, 1993). Biggs and Moore indicated that classrooms consist of students with different backgrounds and performing at different levels; therefore, it is important to differentiate curriculum, instructional methods, and assessment techniques based on students’ needs. Coherence and balance among student-based and classroom-based presage factors is required for optimum teaching and learning processes.

After investigating these learning models, for the current study, the basic elements of students’ learning were categorized as the student-level and classroom-level factors which were drawn from the TIMSS 2003 data set. Student-level factors include gender, self-confidence in science, and home resources. Classroom-level factors include teacher characteristics (teacher’s experience, education, major, and teacher support) and instructional variables (science inquiry, minutes of weekly science instruction, and class size) and classroom composition (class mean of home resources and class mean of science self-confidence). The TIMSS 2003 fourth grade science score (science plausible values) was the outcome variable.

The classroom processes seem to be another important factor in teaching and learning process, however, the TIMSS 2003 data set that is used in this study is not a powerful source of factors that are related to the classroom processes because it is difficult to capture them through survey questions (Burnstein et al., 1995; Rowan, Correnti, & Miller, 2002). Therefore, the conceptual model developed for the current study was limited to the variables available in the TIMSS 2003 dataset. The following section presents an overview of three student-level variables and their relations with student achievement.

**Student-Level Factors and Achievement**

**Gender**

In the recent National Assessment of Education Progress (NAEP) 2005 science test, boys scored significantly higher than girls at all grade levels (National Center for Education Statistics, 2006). However, previous studies investigating the effects of gender on science achievement reported mixed results. Even though, in general, boys outperform girls in science (DeBacker & Nelson, 2000; Eisenhart, Finkel, & Marion, 1996; Greenfield, 1996; Jovanovich & King, 1998; Kahle, 2004; Kahle, Parker, Rennie, &
Riley, 1993; Morrell & Lederman, 1998), results may vary depending on the grade level (Greenfield, 1995; Kahle). Gender differences in science at elementary school level were less common (Beaton et al., 1996b; Greenfield; Kahle).

The reasons for girls’ lower performance in science have been attributed to various factors such as cultural stereotypes of roles in science (DeBacker & Nelson, 2000; Eisenhart, et al., 1996), girls’ lack of exposure to science-related activities inside and outside the classroom (Kahle et al., 1993), decrease in girls’ science ability perception over time (Jovanovich & King, 1998), and gender biases of teachers (Greenfield, 1996), and differences in cognitive abilities (Meyer & Koehler, 1990; Spelke, 2005). It is argued that parents’ actions, beliefs and the language they use at home on gender roles in science also influence children’s attitudes toward science and their achievement in the short and long term (Jacobi, Wittreich & Houge, 2003).

In a study with fifth grade students, Dimitrov (1999) examined patterns of gender differences in science achievement of fifth graders, taking into account the role of ethnicity, ability, and response format. Data from over 2,500 fifth grade students was analyzed by using the statewide science test scores as the outcome. The test included 32 multiple-choice and 10 open-ended items. Multivariate analysis of variance (MANOVA) results showed that the difference between boys and girls in science achievement depended on response formats and ability levels. For the high-ability students, boys outperformed girls only on the open-ended format for physical sciences. There was no significant gender difference for the low- and medium- ability students.

Chapin (2007) investigated gender and racial differences in science and social studies achievement in early grades. Data from the Early Childhood Longitudinal Study (ECLS) analyzed through t-test and correlations. The sample included 13,820 students who completed the General Knowledge Test in kindergarten and first grade. The General Knowledge Test consisted of science and social studies questions. Results showed no substantial gender differences in test scores. Only in kindergarten girls scored slightly higher than boys. By first grade, for all racial groups, boys scored slightly higher than girls. The author cautioned that even though the difference between boys and girls in science achievement was negligible in early grades, the achievement gap might increase during elementary school.
Gender differences in science achievement were more apparent in secondary school than they were at the elementary levels. Penner (2003) used data from the TIMSS 1995 to examine whether there is an interaction between gender and item difficulty. Analysis of high school mathematics and science data from 10 countries revealed significant gender differences in mathematics and science test scores. For both mathematics and science, the male advantages were minimal on easy questions and increased as questions grew more difficult. Beaton and colleagues (1996b) reported that at the 8th grade level, overall, boys performed significantly higher than girls in science across most countries in the TIMSS 1995.

Summary

In general, boys outperform girls in science (DeBacker & Nelson, 2000; Eisenhart et al., 1996; Greenfield, 1996; Jovanovich & King, 1998; Kahle, 2004; Kahle et al., 1993; Morrell & Lederman, 1998). Gender differences in science at the elementary school level are less common than the middle school level (Beaton et al., 1996b; Greenfield, 1995; Kahle, 2004; Solomon, 1997). This study used gender as a student-level variable to examine its possible influences on fourth grade science achievement in the TIMSS 2003.

Self-Efficacy

Bandura (1986) defined self-efficacy as “people’s judgments of their capabilities to organize and execute courses of action required to attain designated types of performances”. (p.391) People with low self-efficacy related to an activity usually avoid that activity, whereas people with high self-efficacy show constant effort to do the activity and they are more likely to be successful (Bandura, 1997; Britner & Pajares, 2001). Bandura describes students who believe that they can not be successful in science related activities put minimal effort to complete those tasks, and when they face a challenge, they usually give up or experience anxiety.

Bandura (1997) suggested that limited experiences and unfamiliarity with the materials usually determine students’ efficacy judgments. To overcome students’ fear and anxiety in science, teachers were recommended to scaffold science activities, accommodate students’ needs and abilities, and provide them with familiar tasks and materials. It was further emphasized to be aware of students’ self-efficacy beliefs to minimize failures that might diminish confidence in new abilities (Bandura; Britner & Pajares, 2006).
In a meta-analysis conducted with 36 studies published between 1977 and 1988, Multon and colleagues (1991) found that 14 percent of the variance in academic performance is accounted for by self-efficacy. The large majority of samples in these 36 studies consisted of elementary school children. Lau and Roeser (2002) found that science self-efficacy predicts high school students’ achievement and engagement with science-related activities in and out of the classroom better than gender, ethnicity, and parental background. At the college level, science self-efficacy was a strong predictor of achievement (Andrew, 1998) and persistence in science-related majors and career choices (Gwilliam & Betz, 2001; Luzzo, Hasper, Albert, Bibby, & Martinelli, 1999).

To determine whether science self-efficacy beliefs predicted science achievement, Britner and Pajares (2001) conducted a study with 262 seventh grade students from four urban public schools. The motivational constructs instrument measured the science self-efficacy, science self-concept, science anxiety, value of science, self-efficacy for self-regulation, and achievement goal orientations. Students' GPA in science class at the end of the academic year was used as the outcome variable. Analyses of variance (ANOVA) indicated that girls and Caucasian students had stronger self-efficacy. Multiple regression analyses were conducted to reveal whether science self-efficacy makes an independent contribution to students’ science achievement. Results indicated that self-efficacy was a strong predictor of science achievement, regardless of gender and race.

Britner and Pajares (2006) replicated the previous study by investigating the effects of science self-efficacy on students’ science achievement. Participants were 319 students (155 boys, 164 girls) in grades 5 - 8 in a public middle school in a middle class, Midwestern city. The Sources of Science Self-Efficacy Scale was adapted from an earlier study in the domain of mathematics (see Lent, Lopez, et al., 1996). Students’ grade in science class at the end of the grading period was used as the outcome variable. Multivariate analyses of covariance (MANCOVA) and regression analyses results showed that science self-efficacy was the strongest predictor of students’ science grade. Approximately 39 percent of the variance in science achievement was explained by science self-efficacy. The results of this study should be taken cautiously due to the demographics of the sample. The findings may not be generalizable to more diverse populations.
Analysis with the 8th grade TIMSS 1999 data investigated the relationship between self-beliefs, instructional strategies, and science achievement over 4,500 students from Japan (House, 2008). The analysis (ANOVA) results indicated that students who expressed less positive perceptions about learning science (i.e., Although I do my best, science is more difficult for me than for many of my classmates) also tended to show lower achievement levels. In a separate analysis, with the TIMSS 1995 and TIMSS 1999 eighth grade science data from Thailand, House found that students who had positive attitudes toward science tended to achieve higher science scores. The multiple regression analysis results were consistent for students in both the TIMSS 1995 and TIMSS 1999 International Samples. Even though these studies revealed important associations between science beliefs and achievement they did not consider the multi-level influences (i.e., student- and classroom-level).

Summary

Several studies reported a strong positive relation between students’ science self-efficacy and achievement across grade levels (Andrew, 1998; Britner & Pajares, 2001, 2006; Gwilliam & Betz, 2001; House; 2004, 2008; Lau & Roeser, 2002; Luzzo et al. 1999). Recent research on science self-efficacy has focused exclusively on middle and secondary level students. The aim of the current study is to extend existing findings regarding the influence of self-efficacy on science achievement at the elementary school level. The variable of science self-confidence derived by the TIMSS 2003 was used as the indicator of self-efficacy.

Home Resources

Studies investigating the effects of socioeconomic status (SES) on student achievement used different factors. Some showed that parental education, income, and occupation were powerful predictors of student achievement (Lytton & Pyryt, 1998; Ma & Klinger, 2004; Manning, 1998; Sammons, West, & Hind, 1997; Thomas, Sammons, Mortimore, & Smees, 1997). However, others argued that achievement is more closely linked to the measures of the social-psychological environment and intellectual stimulation (Iverson & Walberg, 1982; Stevenson & Stigler, 1992). Educational resources and parental support are considered to be critical factors for intellectual stimulation (Campbell & Wu, 1994).
While investigating the factors influencing student learning, Iverson and Walberg (1982) synthesized 18 studies involving home influences. They concluded that intellectual stimulation at home, such as educational resources have stronger influences on children's ability and achievement than the other indicators of socio-economic status. Researchers highlighted that Asian families tend to take a more active role in creating supportive and rich environments with educational resources for their children which results in higher student achievement (Muller & Kerbow, 1993; Stevenson & Stigler, 1992).

Roscigno and Ainsworth-Darnell (1999) investigated the differential effects of cultural capital and home resources on student achievement. Data was drawn from the National Education Longitudinal Study of 1988 (NELS 88) database with over 16,000 students. Students’ grade point averages (GPA’s) were used as the outcome variable as well as the standard test scores. The home resources measure reflected the number of items in a student's household that are conducive to learning, such as computer, books, pocket calculator, and dictionary. Multiple regression analyses indicated that educational resources have strong and positive effects on both GPA’s and standardized achievement.

In the TIMSS 1995 assessment, strong positive relations were found between availability of educational resources and science and mathematics achievement at both elementary and secondary level across countries (Beaton et al., 1996a; Beaton et al., 1996b). Xin and colleagues (2005) conducted multilevel analyses by using the TIMSS 1995 eighth grade mathematics scores from four countries including Japan, Korea, Netherlands, and the U.S. Significant positive effects of educational resources, including books, computer, dictionary, and study desk on mathematics achievement were reported for all countries. However, Van Den Broeck and colleagues (2003) indicated that the effect of educational resources on students’ achievement may have been underestimated in TIMSS assessments because the information is based only on students’ answers. The authors emphasized the need for a more valid measure for SES, such as a parent questionnaire in the TIMSS and other international studies.

Mokshein (2002) examined the relationships between the Malaysian students’ achievement in science in the TIMSS 1999 with student- and school-level factors. The selected student-level factors included educational resources at home, and student background characteristics. The school-level factors included factors related to teachers
and instruction. The TIMSS 1999 data for Malaysia with approximately 5,400 eight-grade students and 150 schools was analyzed by using 2-level HLM. Home resources had a significant positive effect on science achievement at both student- and school-level. School mean of home resources explained approximately 16 percent of variance between schools.

**Summary**

Studies investigating the relations between home resources and student achievement found consistent positive relations. Students coming from homes with more educational resources tended to perform better than their counterparts (Iverson & Walberg, 1982; Mokshein, 2002; Roscigno & Ainsworth-Darnell, 1999; Xin et al., 2005). The current study used the composite variable of educational resources in students’ homes (number of books and the availability of computer and study desk) based on the TIMSS 2003 student questionnaire. The next section presents an overview of the selected classroom-level variables and their effects on student achievement.

**Classroom-Level Factors and Achievement**

The debates over classroom effects on students’ performance started with the Coleman Report (Coleman et al., 1966). The study, which included approximately 600,000 students, 60,000 teachers, and 4,000 public schools, attempted to relate family background (including race and SES) and school variables to students’ test results and their attitudes toward attending higher education. Coleman and colleagues found that student outcomes were unrelated to the school characteristics (i.e., the quality of school facilities, programs, and teachers). Instead, academic results were significantly linked to the characteristics of student body in schools (i.e., the proportion of students with encyclopedias in their homes and the proportion with high aspirations) (Coleman et al.) The results raised questions about school policy development and spending. However, according to the critiques there were some flaws in the Coleman Report including improper sampling, multicollinearity among variables, and not being longitudinal (Kahlenberg, 2001; Ladd, 2003).

Lamb and Fullarton (2002) examined the influences of student, classroom, and school factors on eighth grade students’ mathematics achievement in the U.S. and Australia through 3-level HLM analysis with the TIMSS data set. They used teacher
attributes, such as educational qualifications, years in teaching, and time teaching math in addition to classroom composition, such as class mean SES and student grouping practices. Results showed that, most of the variation between classrooms was due to the classroom compositional factors in both countries and very little of it was due to differences between teachers. The authors highlighted that for improved achievement, policy practices should consider student grouping practices and classroom learning environments rather than only targeting teachers. The following section will present an overview of some teacher characteristics that are chosen for the current study and their relations with student achievement.

**Teacher Characteristics**

*Teacher’s Experience*

Studies that examined the effects of teacher experience on student achievement found equivocal results. Greenwald, Hedges, and Lane (1996) reviewed 60 studies on the relations between school inputs and student outcomes. In these studies teacher experience was strongly related to student achievement. However, in a meta-analysis of teacher characteristics and achievement, Hanushek (1997) did not find evidence to support a relation between teaching experience and student achievement in more than 70 percent of the studies.

Fetler (1999) investigated the effects of teacher characteristics on students’ mathematics scores from Grade 9 through 11 by using a large set of data from California. Approximately 1.3 million students from 795 schools were included in the study. School average mathematics test scores and the Stanford Achievement Test scores were used as student outcomes and teacher characteristics (years of teaching, educational level, and certification status) were provided from California Basic Educational Data System. Correlational analysis revealed that the average number of years in service was positively correlated with students’ math scores at grades 9 through 11. In their multilevel analysis of the TIMSS 1999 eighth grade mathematics data for four countries, which was described earlier, Xin and colleagues (2004) concluded that teaching experience had positive significant effect on eighth grade math scores in the U.S.

Teachers’ experience did not always bring advantages to classrooms. In their analysis of the Study of Instructional Improvement (SII) longitudinal study with over 1,700 third graders and 365 teachers, Hill and colleagues (2005) investigated the relations
between teachers’ mathematical knowledge and student achievement. The measure of students’ math achievement was drawn from McGraw-Hill’s Terra Nova Survey. The results of linear mixed models indicated that there was not any significant relation between the number of years of teaching and math achievement. The impact of novice teachers, who were in their first or second year of teaching, was negative in first grade but insignificant.

Ferguson and Ladd (1996) analyzed achievement scores of students in grades 4, 8, and 9 in Alabama in respect to their relations with school factors. Multi-level regression models concluded that when all other variables were controlled there was no significant relation between teacher experience and students’ state-wide standardized test scores in mathematics and reading. There was evidence, however, that smaller class sizes and higher teacher scores on certification tests positively affected students’ scores.

Goldhaber and Brewer (1998) investigated the influences of teacher preparation on student achievement in different subjects in their analysis of NELS 88 data. F-tests for subject specific teacher models suggested that the years of teaching experience variable was not statistically significant in math, science, English, or history. Likewise, Monk (1994) used Longitudinal Survey of American Youth data to examine the effects of science and mathematics teachers’ subject matter preparation on student performance. Data from approximately 3,000 students and their teachers (N=483) was analyzed by using weighted least square regression analyses. The composite measures of student performance in science and mathematics were used as dependent variables in sophomore and junior years. Achievement tests for students utilized items developed by NAEP. Results indicated little evidence that additional teacher experience contributed to student performance. Indeed, in the junior year, there was a significant negative effect of teacher experience on students’ science performance.

**Summary**

Studies that examined the effects of teacher experience on student achievement found equivocal results. While some studies have demonstrated that students whose teachers are more experienced tended to have higher achievement than their counterparts (Fetler, 1999; Greenwald et al., 1996), others reported no relations between teacher experience and student achievement (Ferguson & Ladd, 1996; Goldhaber & Brewer, 1998; Monk, 1994) and some reported mixed results (Hill, et al., 2005; Xin et al., 2004).
This study used the number of years of teaching experience as a teacher variable at the classroom-level to examine its effects on the TIMSS 2003 science scores.

Teacher’s Education and Subject Major

Following the Coleman Report, a considerable amount of research has been conducted by using school resources and teacher characteristics in respect to their relations with student achievement. Some characteristics, such as level of education, and subject major, received much more attention than others. It is explained that these variables were used frequently because they were easy to measure (Hanushek, 1997) and that the research results can easily be translated into recommendations for policy implications (Darling-Hammond, 2000).

In an extensive report on teacher quality and student achievement, Darling-Hammond (2000) analyzed the data from a 50-state survey of policies, state case study analyses, the 1993-94 Schools and Staffing Surveys (SASS), and the National Assessment of Educational Progress (NAEP), taking student characteristics into account. Correlation and regression analyses indicated strong positive relations between teacher characteristics and student achievement even after controlling for student poverty and language backgrounds. Teacher variables, which included major in the field, and education level, had stronger effects on student achievement than class size, overall spending levels, and teacher salaries. These results should be interpreted cautiously because the author did not include some background variables that are critical for students’ learning, such as parents’ education. Furthermore, since state level aggregated data was used for the analyses, the results might be different when similar kinds of data were used at the individual student, teacher, or school level.

Lederman and Flick (2003) pointed out that, even though having a higher degree in a certain discipline might be the indication of subject matter knowledge, researchers have not yet concluded that teachers with higher degrees are better teachers. In the previously described Goldhaber and Brewer (2000) study with NELS 88 data, in both mathematics and science, students whose teachers had a PhD were not found to have higher scores than the students whose teachers did not hold this credential. Similarly,

Using data from over 7,000 eighth graders who took the 1996 NAEP mathematics assessment, Wenglinsky (2002) investigated the relations between the teacher characteristics and student achievement. As teacher characteristics, the measures of
teachers’ education level and whether the teacher majored or minored in the relevant subject area (mathematics or math education) were used. Multilevel structural equation modeling (MSEM) was used for the analysis. Results showed that teacher’s education level did not have a significant impact on student mathematics performance at eighth grade.

In a recent analysis with Early Childhood Longitudinal Study (ECLS) data, Croninger and colleagues (2007) analyzed the relations between elementary school teacher characteristics and first-grade achievement in reading and mathematics. They used 3-level HLM analyses with a sample of 5,000 students taught by over 1,300 teachers. It is reported that teachers with advanced degrees (i.e., master’s) had no significant effects on students’ achievement. On the other hand, students taught by teachers with an elementary education degree achieved significantly better scores in reading than did students taught by teachers with a different type of degree (i.e., early childhood education).

Using the same data set, Xu and Gulosino (2006) investigated the effects of teacher characteristics and teacher-parent partnership on early childhood performance in public and private schools. Regression analyses with over 9,000 students revealed that teachers’ highest level of education showed no relations with first grade mathematics, reading, and general knowledge scores after controlling for individual, family and school characteristics.

Teacher’s holding a subject major seems to be more consistently related to student achievement in that subject. In the previously described study, Wenglinsky (2002) reported that students significantly outperformed their peers when their teachers had major or minor degree in mathematics. Similarly, Darling-Hammond (2000) found that teacher’s major degree in the field was positively correlated with students’ reading and math scores on NAEP test at both fourth and eighth grade.

Monk and King (1994) used hierarchical linear models (HLM) to analyze data from the Longitudinal Study of American Youth (LSAY), a nationally representative panel survey including a base-year sample of almost 3,000 students from 51 randomly selected public high schools. Survey instruments were completed by sampled students, their teachers, and parents. Students also completed achievement tests in math and science. The results suggested that high school math and science teachers with a major in
their field of instruction have higher achieving students than teachers who are teaching out-of-field.

Goldhaber and Brewer (2000) documented the positive effects of holding a major degree in mathematics on students’ achievement in this subject. They used NELS 88 mathematics and science scores with a sample of over 16,000 12th graders and over 3,400 teachers. Multiple regression analyses showed that students who had teachers with a major degree in mathematics were more successful than those whose teachers had out-of-subject degrees. In science, there was no impact of teachers’ subject specific degrees on student scores.

In a separate analysis with NELS 88 data Goldhaber and Brewer (1998) investigated the influences of teacher preparation on student achievement in math, science, English, and history. F-tests for subject specific teacher models suggested that in math and science, holding subject-specific degrees was found to have a positive impact on student test scores in those subjects. However, Darling-Hammond (1999) indicated that since more than 20 variables simultaneously tested in the equations and many of them highly correlated with certification status, it is difficult to interpret the findings of the Goldhaber and Brewer’s study. In cases like this, multicollinearity problems often cause mis estimations of the effects (Darling-Hammond).

Summary

Empirical studies that investigated the impacts of teachers’ subject major on students’ achievement indicated that in general, students tended to perform better when their teachers had a major degree in the subject taught (Darling-Hammond, 2000; Goldhaber & Brewer, 1998, 2000). On the other hand, studies examined the impacts of teachers’ education on achievement yielded mixed results. While Darling-Hammond (2000) reported that students perform better when their teachers have advanced degrees, others reported no relations between advanced degrees and student achievement (Croninger et al., 2007; Goldhaber & Brewer; Wenglinsky, 2002; Xu & Gulosino, 2006). The current study examined whether teachers’ science or math majors and advanced degrees affect students science achievement on the TIMSS 2003 science test.

Teacher Support

Teacher support generally involves characteristics such as caring, friendliness, understanding, dedication, and dependability (Ryan & Patrick, 2001). It is stated that
when teachers care about their students and believe in their potential, the students sense it and put forth greater efforts to succeed (Lumpkin, 2007; Noddings, 1992). As Noddings emphasized, “caring is the very bedrock of all successful education.” (p.27) Supportive teachers usually build positive relationships with their students. They believe in each student’s ability to achieve and shape the instruction by placing the learner at the center of teaching/learning process (Noddings).

Deci and Ryan (1985) indicated that students have psychological needs inside the classrooms, such as relatedness and attentiveness. Teachers’ practices play important roles in meeting these needs. Although teachers might report that they make efforts to meet students’ psychological needs, students’ perceptions and interpretations of these efforts are important (Daniels & Perry, 2003). Therefore, this study utilized student responses to those statements about teacher care and support.

Klem and Connell (2004) examined the links between teacher support, student engagement, and student academic success in elementary and middle schools. Longitudinal data sets collected by the Institute for Research and Reform in Education were used with over 4,000 elementary and secondary students. Teacher support items which were derived from student surveys included statements, such as ‘My teacher cares about how I do in school’ or ‘My teacher’s expectations of me are way off base.’ The threshold analysis examined the effects of teacher support on student engagement and performance. Results showed that students who perceive teachers as caring, encouraging, and fair were more likely to have higher attendance and test scores on variables that strongly predict whether youth will successfully complete school and pursue post-secondary education.

Brown, Anfara, and Roney (2004) conducted a qualitative multi-site case study in order to investigate the possible explanations for the difference in student achievement in six high performing and six low performing middle schools. Apart from the great disparity in SES and funding between two types of schools, teachers in high performing schools had higher expectations for their students and higher levels of commitment for student success than those in low performing schools. The authors recommended holding high expectations for all students in low achieving schools, for improved school climate, rather than outlining reform initiatives with hard to achieve goals.
Teacher support seems to be influencing students’ engagement in instructional activities as well as their achievement. Marks (2000) investigated the factors influencing students’ engagement across grade levels and subjects. Over 3,500 elementary, middle, and high school students were surveyed on items about their attitudes, behaviors, and experiences in class, as well as their background. The three-level HLM analysis indicated that supportive classroom environments, in which students experience high expectations and receive help from teachers increase student engagement. The effect was the greatest among elementary school students and in mathematics. Mark stated that enhancing student engagement through care and support could be the key in improving achievement.

On the basis of student engagement, Skinner and Belmont (1993) examined the effects of three dimensions of teacher behavior (involvement, structure, and autonomy support) on 144 students’ (Grades 3-5) behavioral and emotional engagement across a school year. Teacher context was assessed through teachers’ and students’ report of classroom interactions. Student questionnaire included items, such as ‘My teacher does not know the real me’ and ‘I know what my teacher expects of me in class.’ Student engagement was assessed with children's and teachers’ reports of behavior and emotion in the classroom. Co-relational analyses indicated that there was a strong relation between teacher/student interactions and students’ behavioral and emotional engagement in the classroom.

Summary

In general, more supportive teachers not only increase the levels of student engagement and motivation (Klem & Connell, 2004; Marks, 2000; Skinner & Belmont, 1993) but also improve student performance (Brown et al., 2004; Klem & Connell) in elementary and middle schools. The current study used the composite of student responses to those statements regarding teacher support on the TIMSS 2003 student questionnaire.

Instructional Variables

Science Inquiry

The National Science Standards (National Research Council, 1996) and Project 2061’s benchmarks (American Association for the Advancement of Science, 1993) suggest that it is important to provide opportunities for students to conduct inquiry which
involves investigating, experimenting, and problem solving. The National Science Standards (National Research Council, 1996) describe science inquiry as “the activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world.” (p. 23) The nature of science inquiry entails using investigative skills; actively seeking answers to questions about specific science concepts; and developing students’ ability to explore (Barman, 2002; Von Secker, 2002; Yore, 1984). National Science Standards suggest that students learn science more effectively by engaging in meaningful activities through manipulating materials (National Research Council).

Lee and colleagues (2006) examined the impact of an instructional intervention to promote achievement and equity in science and literacy for culturally and linguistically diverse over 1,500 elementary students. Science intervention was designed to promote inquiry-based learning which moved progressively from a teacher-explicit instruction to student-initiated exploration. Intervention was supported with teacher workshops and materials for hands-on activities. To assess students’ knowledge of science concepts and inquiry, pre- and post-unit tests as well as NAEP and TIMSS tests were administered. Data were analyzed using dependent t-test and Cohen’s $d$ effect sizes were computed. The students in the intervention group demonstrated statistically significant gains and large effect sizes ($d > .80$) on all measures of science and literacy achievement at the end of the school year. The third and fourth graders in the research performed lower than the national and international samples of students on NAEP and TIMSS items for grades 3 and 4 at the beginning of the school year; however, they generally performed better than the samples.

Paris and colleagues (1998) assessed the effects of a hands-on science program on students’ interest and learning about biology. One hundred and eighty four students in Grades 3, 4 and 5 participated in a six-week curriculum involving inquiry-guided experiments and open-ended worksheets. Effects of these experiences were assessed before and after the program with a test of scientific problem solving and an attitude survey. Analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) results revealed that there were significant improvements in their problem-solving skills and significant increases in students' interest in science and at all grade
levels. Girls had higher problem-solving scores and reported more positive attitudes about science and than boys.

Stohr-Hunt (1996) examined the relation between the amount of time students spent experiencing hands-on science and science achievement by using NELS 88 sample of approximately 25,000 eighth grade students from over 1,000 public and private schools. Student achievement in science was measured by a cognitive test battery developed by the Educational Testing Service. Information regarding the frequency of hands-on experience was collected through a self-administered teacher questionnaire, which included a series of questions specific to the science curriculum. The results of ANOVA concluded that significant differences existed across the hands-on frequency variable with respect to science achievement. Specifically, students who engaged in hands-on activities every day or once a week scored significantly higher on a standardized test of science achievement than students who engaged in hands-on activities once a month, less than once a month, or never.

House (2006) investigated the effects of classroom instructional strategies on science achievement of elementary school students in Japan by using the TIMSS 1995 data for Population 1 (9-year-olds). According to multiple regression analysis results students who reported that they conducted experiments in class more frequently earned higher test scores than their counterparts.

Inquiry-based science did not always promote positive results in student outcomes. In an effort to measure the degree to which students can do inquiries, Pine and colleagues (2006) compared the performance of about 1,000 fifth grade students in inquiry-based hands-on curricula with an equal number of students with textbook curricula. Participants included students from 41 classrooms from nine school districts and lessons required one to three class sessions. At the completion of lessons, students were given four performance assessments. A 25-item Third International Math and Science Study (TIMSS) test and a 65-question short-answer ‘Cognitive Abilities’ test on literacy, mathematics, and figure analysis were also administered. Multivariate HLM analysis results showed that hands-on students performed better on only one of the performance tests controlling for school and teacher variables. There was a strong positive effect of students’ cognitive abilities on their test scores. The authors stated that
the lack of curriculum effect on student performance might be due to the way it is being implemented by the teachers.

One of the components of science inquiry described by the TIMSS 2003 was the group work. Small groups are sites in which the individual ideas are shared and negotiated (Roth & Bowen, 1995; Roth, 1995). Lazarowitz, Hertz-Lazarowitz, and Baird (1994) indicated that since science activities usually require students to work in groups, science courses are natural curriculum areas for focusing on cooperative learning practices. During cooperative small group activities students are actively engaged in the material by sharing insights, ideas, and representations, giving feedback, and teaching each other (Towns & Grant, 1997).

In their meta-analysis, Qin and colleagues (1995) compared the effects of cooperative and competitive efforts on problem solving. They investigated 46 studies published between 1929 and 1993. Findings revealed that members of cooperative teams from all ages outperformed individuals competing with each other. The superiority of cooperation was greater on non-linguistic problems that involved symbols, math, motor activities and actions than linguistic problems that were solved through written and oral language.

In an analysis with the TIMSS 1995 Population 1 (9-year-olds) data for Japan, House (2006) found that more frequent use of cooperative learning activities (i.e., working together in pairs or small groups) during science lessons was significantly related to higher science test scores. In addition, analysis with the TIMSS 1995 Population 2 (13-year-olds) data for Japan and the U.S. revealed that cooperative learning activities were significantly associated with student enjoyment for learning science (House, 2003).

**Summary**

In general, the emphasis on science inquiry (House, 2006; Lee et al., 2006; Paris et al., 1998; Stohr-Hunt, 1996) and group-work (House, 2003, 2006; Qin et al., 1995) positively influenced students’ science achievement. However, there were exceptions when students did not perform better than their counterparts when science inquiry (see Pine et al., 2006) was emphasized. This study examined whether the emphasis on science inquiry, reported by students, in elementary science classrooms effects science achievement. The composite score based on the student reports about inquiry related
activities, including group work, on the TIMSS 2003 student questionnaire was used as a classroom-level variable.

**Amount of Science Instruction**

The U.S. Department of Education (2000) and the National Research Council (1996) emphasize equal time for science, reading, writing and mathematics teaching. Recently, the National Academy of Sciences (2006) called for ‘vastly’ increased efforts in K-12 science education so the U.S. does not fall behind in the global economy. Yet, in national comparisons teachers reported spending consistently less time in science instruction compared to language arts and math (Center on Education Policy, 2007; Weiss & Pasley, 2004). The 2000 National Survey of Science and Mathematics Education reported that time spent per day on science teaching for grades K-2 science was 21 minutes compared to 119 minutes spent per day on reading/language arts, and 49 minutes spent on mathematics. In grades three through five, the average number of minutes per day devoted to reading/language arts was 108 minutes; for mathematics, 58 minutes; and for science, 30 minutes (Fulp, 2002).

The Center on Education Policy (CEP) recently published a report (July, 2007) which included average amount of time spent on different subjects in 349 school districts nationally. The report showed that since the enactment of the No Child Left Behind act 2001-2002, the average number of minutes of instruction in English language arts and mathematics increased 42 percent. On the other hand, an average of 31 percent decrease in minutes spent in other subjects or activities was observed. The average decrease in science instruction per week was 28 percent. With the demands of the No Child Left Behind Act and an ever increasing focus on students’ reading performance, teachers may believe that time spent on science impedes students’ reading development. Thus, they devote less time for science instruction compared to language arts and math (Marx & Harris, 2006).

In general, time spent on subject content was positively related to student achievement. Two meta-analyses of school effects on student achievement (Lewin, 1993; Fuller, 1987) concluded positive effects of time spent on subject content on student achievement. In the first meta-analysis by Lewin, 86 percent of the studies reported the length of instructional time to be an important factor influencing student achievement. In the second meta-analysis of 60 studies (Fuller) that examined school resources in relation
to student achievement, most studies reported positive effects of daily or yearly instructional time on achievement in various subjects.

In their analysis of Early Childhood Longitudinal Study—Kindergarten Cohort (ECLS-K) data Bodovski and Farkas (2007) examined the effects of the kindergarten mathematics knowledge, instruction, and engagement on subsequent achievement growth through third grade. Data included over 13,000 students and was analyzed by using ordinary least-squares (OLS) regression method. The amount of math instruction was measured in minutes per day based on teachers’ reports. Results showed that the time spent on mathematics instruction was positively associated with gains in achievement, though with a small effect size. Time on math instruction increased achievement for all students equally, regardless of their beginning achievement.

Coates (2003) investigated the effects of instructional time in four subjects on elementary students’ standard test scores in mathematics, reading and writing. Data from 800 school districts with more than 2,000 schools in Illinois provided information about the minutes of instruction per day in mathematics, English, social studies, and science. In addition to the instructional time, teacher characteristics and student background variables were used in multiple regression analyses. The results of the analyses generally indicated that more instruction in mathematics and reading improved scores in these subjects. On the contrary, additional time spent in science instruction had no beneficial effect on any test scores. The author recommended further investigation of the effects of time spent in science instruction on students’ science scores.

Baker and colleagues (2004) examined the effects of instructional time on eighth, ninth and tenth grade student achievement by using three cross national datasets, Programme for International Student Assessment (PISA) 2000, the TIMSS 1999, and the International Study of Civic Education (CIVICS) 1999. Co-relational analyses revealed that there was no significant relation, at the cross-national level, between achievement test scores and the amount of instructional time in science. Within each country, the average correlation between total time of instruction and science achievement was weak. The association between instructional time and achievement in science accounted for, on average, 5 percent of the variance in each country. In three countries more science instruction was associated with lower test scores. The authors recommended improving
teaching and the curriculum for improved student achievement, rather than using resources in marginal increases in instructional time.

**Summary**

In general, time spent on subject content positively affected student achievement (Bodovski & Farkas, 2007; Coates, 2003; Lewin, 1993; Fuller, 1987). However, the results are inconclusive for studies investigating the relations between the amount of science instruction and science achievement (Baker et al., 2004; Coates). To provide further empirical information between the amount of science instruction and achievement, the current study used the amount of weekly science instruction (in minutes) as a classroom-level variable based on the TIMSS 2003 teacher questionnaire.

**Class Size**

Experimental class size studies reported improved student achievement associated with small class sizes (Finn & Achilles, 1999; Mitchell & Mitchell, 1999; Smith, Molnar & Zahorik, 2003). For non-experimental studies, researchers suggest that the effects of class size should be investigated with caution because of nonrandom assignment of students to classrooms (Boozer & Rouse, 1995; Pong & Pallas, 2001). Students might be assigned to classrooms based on their academic achievement or special needs. Moreover, families with high SES might prefer schools with small class sizes. Therefore the observed effect might be due to the SES rather than the class size (Boozer & Rouse).

The most widely known class size study, Tennessee’s Project STAR, revealed that small class size increased students’ achievement significantly in the primary years (Finn & Achilles, 1999). Approximately 12,000 students participated in the course of the 4-year study. Children were assigned to one of the three class types and stayed in the same experimental condition for four years. The researchers were able to follow participants from third through seventh grade. Analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) showed that in small classes (13-17 students) provided improved teaching conditions and student learning; fewer discipline problems, and fewer student retentions. In order to investigate the long terms effects of small class sizes Nye, Hedges, and Konstantopoulos, (1999) conducted analyses by using the same students’ California Test of Basic Skills (CTBS) science, mathematics, and reading scores at forth, sixth, and eighth grade. Two-level HLM analyses results
indicated that the positive effects of small classes in early grades persisted through eight grade in all three subjects.

Similar studies in Wisconsin, California, and North Carolina also showed that students achieve higher test scores when they are in smaller classes (Mitchell & Mitchell, 1999; Smith et al., 2003). Wisconsin’s Student Achievement Guarantee in Education (SAGE) program was designed for low-income students to increase the academic achievement by reducing the K-3 class size to 15 students. Students were tracked from 1996 to 2001 and student scores from 30 schools that participated in the program were compared with those of students from 17 schools with larger class sizes. Overall, first and third graders in the SAGE program scored significantly higher on reading, language arts, and mathematics subtests of CTBS than their counterparts in larger classes (Smith et al.).

Miller-Whitehead (2002) examined the effects of class size on eighth grade students’ science achievement by using Terra Nova science scores in 138 school districts. The results of regression analyses showed that as the percentage of small classrooms in a district increased students’ science scores also increased significantly. Wenglinsky (1997) explored the possibility that economic resources are linked to student achievement through the impact of class size. The path analyses of the National Assessment of Educational Progress (NAEP) mathematics data for eight graders showed significant indirect effect of class-size on student achievement. Smaller class-size was positively related to school environment which had a significant positive effect of math achievement.

Smaller class-size did not always provide evidence for higher academic levels. In a meta-analysis, Hanushek (1996) reported that only 15 of 277 estimates of the effects of teacher-student ratio on student performance were significantly positive. Influenced by the STAR project in the U.S., a large scale class-size reduction initiative in the United Kingdom involved more than 1,000 elementary schools (Hargreaves, Galton, Pell, 1998). The report concluded that no simple links existed between class-size and students’ learning. The researchers suggested that small class sizes benefited only students in the early years of elementary education probably because teachers are able to provide more individual attention in smaller classrooms which is critical for the development of younger children.
Similarly, Bosker (1998) investigated the effects of class-size on students’ math and language achievement in 416 elementary and middle schools in the Netherlands. Multilevel analysis included SES, gender, and IQ, as well as class size. The students in second grade classrooms with 25-29 and 35-39 students scored significant lower than the students in smaller classrooms. The associations between class size and student achievement in grades four, six and eight were less pronounced, although there were some indications that students in classes with 35 slightly lagged behind other students. Both studies from the UK and the Netherlands recommended policy makers to make their investments wisely and focus on early elementary grades rather than upper grades.

Internationally, research suggested little evidence of the class size effects on student achievement (Blatchford, Goldstein, & Mortimore, 1998; Pong & Pallas, 2001; Woessman & West, 2002). In their study, Pong and Pallas investigated the relations between class size and eighth grade mathematics achievement in nine countries, including the U.S., which participated in the TIMSS 1995. Multilevel analyses indicated that the U.S. was the only country in which there was a negative association between math achievement after controlling for teacher, classroom, and school effects. There was not a linear association between class size and mathematics achievement in eight other countries examined. One reason for this result was attributed to the fact that in most countries small classes were possible remedial classes with low achieving students and students from low socioeconomic backgrounds.

Woessman and West (2002) estimated the effects of class size on students’ mathematics and science scores in 18 countries by using the TIMSS 1995 seventh and eighth grade dataset. Clustering-robust linear regression method was used to estimate student- and school-level effects. In 11 countries, the findings ruled out large class-size effects. While there were sizeable beneficial effects of smaller classes in Greece and Iceland, no effects were found in Singapore and Japan. The authors pointed out that in countries where class-size had significant effects on student achievement teachers had lower credentials and education.

**Summary**

Research on the effects of class-size studies on student achievement yielded mixed results. While some research documented positive effects of smaller class-size on student achievement (Finn & Achilles, 1999; Mitchell & Mitchell, 1999; Smith et al.,
others indicated that smaller classrooms are beneficial only for younger students (Bosker, 1998; Hargreaves et al., 1998). Internationally, research suggested little evidence of the effect of the class size on student achievement (Pong & Pallas, 2001; Woessman & West; 2002). To investigate the effects of class size in five countries, the current study used class size at the classroom-level.

Although previous research on teacher and instructional characteristics used large scale data sets and appropriate analyses, few of them used elementary school data. Furthermore, most of the previous research focused on mathematics and reading achievement. In an effort to address the gaps in the literature, this study drew on international data from the TIMSS 2003 to examine the relations between elementary school teacher and instructional variables and science achievement, taking student-level factors and the classroom composition into consideration. In doing so, this study attempted to provide further empirical evidence about the classroom effects on science achievement at the elementary school level. The next section provides an overview of the characteristics of the TIMSS 2003 database.

**The Trends in International Mathematics and Science Study (TIMSS) 2003**

The TIMSS, sponsored by the International Association for the Evaluation of the Education Achievement (IEA), has measured mathematics and science achievement of nationally representative samples of students and collected background information from students, their teachers, and schools. The TIMSS 2003 is the third in a continuing cycle of international mathematics and science assessments conducted every four years. The TIMSS assesses achievement in countries around the world and collects a rich array of information about the educational contexts for learning mathematics and science. Forty-eight countries participated in the TIMSS 2003 at the eighth grade and 26 countries participated at the fourth grade. The database includes data from over 360,000 students, about 25,000 teachers, about 12,000 school principals and National Research Coordinators of each country (Martin & Mullis, 2004).

The TIMSS was administered in 1995, 1999, and 2003 to eighth graders and in 1995 and 2003 to fourth graders.

For countries that participated in previous assessments, the TIMSS 2003 provides three-cycle trends at the eighth grade (1995, 1999, 2003) and data over two points
in time at the fourth grade (1995 and 2003). In countries new to the study, the 2003 results can help policy makers and practitioners assess their comparative standing and gauge the rigor and effectiveness of their mathematics and science programs. (Martin & Mullis, 2004, p.1-1)

Table 2.1 *The TIMSS 2003 Fourth Grade Science Results* (adapted from Martin & Mullis, 2004)

<table>
<thead>
<tr>
<th>TIMSS 2003 Rank</th>
<th>Country</th>
<th>Average Score 2003</th>
<th>Average Score 1995</th>
<th>TIMSS 1995 Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Singapore</td>
<td>565</td>
<td>523</td>
<td>10 ▲</td>
</tr>
<tr>
<td>2</td>
<td>Chinese Taipei</td>
<td>551</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Japan</td>
<td>543</td>
<td>553</td>
<td>2 ▼</td>
</tr>
<tr>
<td>4</td>
<td>Hong Kong SAR</td>
<td>542</td>
<td>508</td>
<td>14 ▲</td>
</tr>
<tr>
<td>5</td>
<td>England</td>
<td>540</td>
<td>528</td>
<td>8 ▲</td>
</tr>
<tr>
<td>6</td>
<td>United States</td>
<td>536</td>
<td>542</td>
<td>3 ▼</td>
</tr>
<tr>
<td>7</td>
<td>Latvia</td>
<td>532</td>
<td>486</td>
<td>18 ▲</td>
</tr>
<tr>
<td>8</td>
<td>Hungary</td>
<td>530</td>
<td>508</td>
<td>15 ▲</td>
</tr>
<tr>
<td>9</td>
<td>Russian Federation</td>
<td>526</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>Netherlands</td>
<td>525</td>
<td>530</td>
<td>6 ▼</td>
</tr>
<tr>
<td>11</td>
<td>Australia</td>
<td>521</td>
<td>521</td>
<td>5 ▼</td>
</tr>
<tr>
<td>12</td>
<td>New Zealand</td>
<td>520</td>
<td>505</td>
<td>16 ▲</td>
</tr>
<tr>
<td>13</td>
<td>Belgium-Flemish</td>
<td>518</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>Italy</td>
<td>516</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>Lithuania</td>
<td>512</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>Scotland</td>
<td>502</td>
<td>514</td>
<td>13 ▼</td>
</tr>
<tr>
<td>17</td>
<td>Republic of Moldova</td>
<td>496</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>Slovenia</td>
<td>490</td>
<td>464</td>
<td>11 ▼</td>
</tr>
<tr>
<td>19</td>
<td>Cyprus</td>
<td>480</td>
<td>450</td>
<td>23 ▲</td>
</tr>
<tr>
<td>20</td>
<td>Norway</td>
<td>466</td>
<td>504</td>
<td>17 ▼</td>
</tr>
<tr>
<td>21</td>
<td>Armenia</td>
<td>437</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>Islamic Republic of Iran</td>
<td>414</td>
<td>380</td>
<td>25 ▲</td>
</tr>
<tr>
<td>23</td>
<td>Philippines</td>
<td>332</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>Tunisia</td>
<td>314</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>Morocco</td>
<td>304</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td><strong>International Average</strong></td>
<td><strong>489</strong></td>
<td><strong>524</strong></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* In Yemen, classroom sampling procedures did not meet the TIMSS sampling standards and so could not be approved by the International Study Centre.

The TIMSS is a comprehensive international study which involves complex design and procedures for assessing students’ mathematics and science achievement, as well as drawing samples, analyzing and reporting the data. The TIMSS 2003 was administered near the end of the school year in the participating countries. Data for the TIMSS participants have been released by the National Center for Educational Statistics.
(NCES) since 1997 (Martin & Mullis, 2004). The Table 2.1 lists the countries which participated in the TIMSS 2003 at the fourth grade and their rankings in science in 2003 and in 1995.

Sixteen of the 26 countries participated in the TIMSS 2003 at fourth grade, also participated in 1995 and provided data at two points in time. Similar to previous TIMSS results, Asian countries performed at the top of other countries in 2003. At the fourth grade, Singapore and Chinese Taipei were the top performing countries having significantly higher mean achievement than the rest of the participating countries. The U.S. placed sixth after England (Martin & Mullis, 2004).

To represent the range of performance shown by students internationally, the advanced benchmark was selected to be 625, the high benchmark was 550, the intermediate benchmark was 475, and the low benchmark was 400 on the TIMSS 2003 assessment. To describe the performances at these benchmarks, the TIMSS worked with the Science and Mathematics Item Review Committee. At the fourth grade, students at the advanced benchmark were able to apply their knowledge and understanding in beginning scientific inquiry whereas those at the low benchmark demonstrated only some elementary knowledge of the earth, life, and physical sciences (Martin et al., 2004).

The TIMSS 2003 has been carefully designed and implemented across countries which helped consistency and comparability of data for participating countries. To help readers understand the TIMSS 2003 database, the next section briefly describes the instrument development and test administration, background questionnaires, sampling design, and scaling procedures.

**Instrument Development and Test Administration**

Developing the TIMSS tests for 2003 was a cooperative effort involving all of the National Research Coordinators in participating countries during the entire process. The TIMSS International Study Center began the process with an item-writing workshop for the National Research Coordinators and their colleagues. Participating countries later submitted items that were reviewed by science subject matter specialists. The items were field-tested with representative samples of students in each country, and all of the potential new items were reviewed by the Science and Mathematics Item Review Committee. The National Research Coordinators reviewed the items and the scoring criteria (Martin & Mullis, 2004).
The TIMSS 2003 science curriculum frameworks included the content dimensions as well as the cognitive dimensions. There were three content domains at fourth grade level: (1) life science, (2) earth science, and (3) physical science. Student achievement was reported in terms of performance in each content area as well as in science overall. Science cognitive domains defined the sets of behaviors expected of students as they engaged with the science content as factual knowledge, conceptual understanding, and reasoning and analysis (Martin & Mullis, 2004).

The TIMSS 2003 science tests contained 189 items at the eighth grade and 152 items at the fourth grade. To ensure reliable measurement of trends over time, the assessment included also items that had been used in the 1995 and 1999 assessments. Each student took one booklet containing both mathematics and science items. There were 12 student booklets at each grade level, with six blocks of items in each booklet. Even though questions may vary from booklet to booklet, they were designed with the same level of difficulty. The Cronbach’s alpha reliabilities across the 12 test booklets generally were high, with an international median of 0.84 at both 4th and 8th grades. The assessment time for individual students was 72 minutes at fourth grade and 90 minutes at eighth grade (Martin & Mullis).

Science tests contained multiple choice and constructed-response format items (see Appendix B). Between one-third and two-fifths of the items were in constructed-response format. In constructed-response items, students were asked to generate and write their answers. These included short answer items and extended response items. Correct answers to most questions were worth one point. Some constructed-response questions were evaluated for partial credit, with a fully correct answer being given two points. For reliably scoring student responses to constructed-response items, detailed rubrics were provided and reliability checks were conducted within and across countries. The agreements among scorers for science were 97 percent within country and 87 percent across countries (Martin & Mullis, 2004).

The TIMSS 2003 data collection instruments were prepared in English and translated into 34 languages. In addition to translation, the international versions of some items were modified for cultural reasons, even in the countries that tested in English. For quality assurance, this process involved several groups of people such as translators,
subject matter experts, International Study Center consultants and statisticians (Martin & Mullis, 2004).

Each participating country was responsible for the data collection and quality control procedures. Instructions about handling the materials, test rules, directions, timing, security and rules for answering students’ questions were provided for test administrators in training manuals (Martin & Mullis, 2004).

**Background Questionnaires**

In order to identify factors or combination of factors related to student achievement, the TIMSS 2003 administered a wide range of questionnaires. The TIMSS 2003 background questionnaires collected intensive information from students, teachers, school principals, and curriculum specialists. The data collected through these questionnaires aimed to provide information about similarities and differences in educational systems, policies, and practices among participating countries (Martin et al., 2004).

Across the two grades and two subjects, the TIMSS 2003 involved 11 questionnaires. The students who took the TIMSS 2003 test answered questions pertaining to their home backgrounds and their experiences in learning mathematics and science. The teachers of the sampled students responded to questions regarding teaching emphasis on the topics in the curriculum frameworks, instructional practices, professional training and education, and their views on mathematics and science. The school questionnaires were completed by principals and consisted of questions about school contexts for the teaching and learning of mathematics and science. Finally, the curriculum questionnaire included questions regarding the organization, emphasis, and content coverage of the mathematics and science curriculum at fourth and eighth grades (Martin, 2005). The current study utilized the items from student and teacher questionnaires.

**Sampling Procedure**

“In IEA studies, the target population for all countries is known as the international desired population” (Foy & Joncas, 2004, p.110). The TIMSS 2003 targeted students at the end of their fourth (Population 1) and eighth (Population 2) years of formal schooling in the participating countries. For the purpose of this study Population 1 data was used. In each country, representative samples of students were
selected using a two-stage stratified cluster design. First, countries selected at least 150 schools using probability-proportional-to-size sampling. Stratification variables, such as urbanicity, region or school type were used for the first stage. At the second stage, one or two classes were randomly selected from each school. Average class size differed from 23 students per class in the U.S. to 38 students per class in Singapore. As a result, at least 4,000 students were sampled in each country (Martin, 2005).

In the TIMSS 2003 schools, classrooms, and students were all considered as sampling units in the sample design in order to meet specific requirements for data quality and sampling precision at all levels. Participants were expected to include 95 percent of the national desired populations in their sample. However, exclusions occurred at the school and classroom levels, or both. Schools in geographically remote regions, extremely small schools, schools with a curriculum different from the mainstream educational system, and schools which were serving to only certain type of students (i.e., students with disabilities) were excluded from the sampling process in the first stage. Due to the possibility of non-participation at the school level, a priori, replacement schools were identified and replaced with the sampled school when needed (Foy & Joncas, 2004).

In the second stage TIMSS preferred to sample intact classrooms because it makes the linking between students and teachers simpler. At fourth grade, students in most countries were taught all subjects by the same teacher. Therefore, sampling intact classrooms was straightforward in fourth grade. At eighth grade, however, because classrooms were organized by subjects, TIMSS chose the mathematics class as the sampling unit. As a rule, in each country, one classroom per school was sampled, although some participants preferred to sample two classrooms. Additionally, some participants were required to sample two or more classrooms per school in order to meet the minimum requirement of 4,000 sampled students (Foy & Joncas, 2004).

Scaling Procedure

In order to provide comparison between 1995 and 2003 scores for fourth graders, TIMSS proficiency scale scores were placed on the same metric. The TIMSS scale average was set at 500 and the standard deviation at 100. Because of its complex design, with each student responding only to a sample of items and not the entire assessment, TIMSS relies primarily on item response theory (IRT) scaling methods to measure trends.
in students’ mathematics and science achievement (Gonzalez, Galia, & Li, 2004). The IRT analysis provides common mathematics and science scales for comparing student performance across countries and over time. The IRT scaling method computes a score by averaging each student’s responses for items that he or she took by taking into account the difficulty and discriminating power of each item. The IRT method is more reliable when a student responds to a large number of items (Martin & Mullis, 2004).

In TIMSS assessments students are given only a fraction of the entire test, therefore, it is challenging to obtain reliable estimates of students’ scores. To overcome this problem, TIMSS uses multiple imputations or ‘plausible values’ procedure. TIMSS produces achievement distribution for each student by combining information about item characteristics, student responses, and student background information. To account for error in this imputation process, TIMSS draws five ‘plausible values’ for each student, and incorporates the variability between the five estimates. The difference between the five values reflects the degree of uncertainty in the imputation process. When the process yields consistent results, the differences between the five values are very small. (Gonzalez et al., 2004).

**Summary**

The Trends in International Mathematics and Science Study (TIMSS 2003) was a large and rigorous international comparison of student performance in mathematics and science. Data included background information from students, their teachers, and schools as well as students’ mathematics and science achievement scores at fourth and eighth grade. Similar to previous TIMSS results, Asian countries performed at the top of other countries in 2003. It is aimed to help policy makers and practitioners to assess their comparative standing and evaluate their mathematics and science programs (Martin & Mullis, 2004).

This chapter provided a literature review of the previous school learning models, student- and classroom-level factors utilized in this study as well as the characteristics of the TIMSS 2003 database. The following chapter discusses the methodology employed in this study.
CHAPTER 3

METHODOLOGY

This study examined the effects of student-level and classroom-level factors on fourth grade science achievement. Due to the nested data structure, multilevel analysis was used with a large scale data set. The effects of teacher characteristics, instructional variables, and classroom composition along with student characteristics on science achievement at the elementary level in five countries were explored. The results were reported for Singapore, Japan, the U.S., Australia, and Scotland. Hierarchical Linear Modeling (HLM) was used in analyzing data from the Trends in International Mathematics and Science Study (TIMSS) 2003.

The following research questions guided the statistical analyses and their interpretations:

1. What student-level factors are significantly related to science achievement at fourth grade level in Singapore, Japan, the U.S., Australia, and Scotland?
2. What classroom-level factors are significantly related to science achievement at fourth grade level in Singapore, Japan, the U.S., Australia, and Scotland?
3. How much of the variance in student achievement is explained by student- and classroom-level factors within and across countries?

The science score (five plausible values) on the TIMSS 2003 was the outcome variable of interest. Three variables were used at the student-level, which are gender, home resources, and science self-confidence. Seven variables related to teacher characteristics and instructional variables were added to the classroom-level. Those variables are, namely, teacher’s experience, education, major and teacher support, the amount of weekly science instruction, emphasis on science inquiry and class size. Two student-level variables, home resources and science self-confidence were aggregated at the classroom-level and used as classroom composition variables. The descriptions of these variables were further discussed in Variables section.
Data Source

The Trends in International Mathematics and Science Study (TIMSS) 2003, fourth grade (Population 1) data was the main data source for this study. Approximately 360,000 students and about 25,000 teachers from 26 countries participated in the TIMSS 2003 at the fourth grade level. The TIMSS 2003 provided data from a nationally representative sample of students for each country. Therefore, the results can be generalized to a specific country’s students. The TIMSS 2003 study provided information about teacher characteristics and some instructional variables. Furthermore, this data set was appropriate for multilevel analysis in which student- and classroom (teacher)-level variables can be included.

Sample

The TIMSS 2003 is a large international dataset. For handling the data more effectively, instead of using the whole dataset, four other countries along with the U.S. were selected. Two of these countries scored significantly higher than the U.S., and two of them scored significantly lower. The following criteria were used in the selection of the countries:

Singapore: The top scoring country on the TIMSS 2003 fourth grade science. ($N_S$ = 6,122, $N_T$ = 171).

Japan: A country which has long been used as a benchmark by and compared to the U.S. Japan scored significantly higher than the U.S. on the TIMSS 2003 fourth grade science. ($N_S$ = 4,250, $N_T$ = 144).

Australia: An English speaking country which scored significantly lower than the U.S. Similar to the U.S., educational system is decentralized in Australia (Bray, 2007) ($N_S$ = 3,573, $N_T$ = 206).

Scotland: Like Australia, a primarily English speaking country which scored significantly lower than the U.S. Educational system is also decentralized in Scotland (The Scottish Government, 2008) ($N_S$ = 2,665, $N_T$ = 120).

It is important to note that teachers in the sample from each country were teachers who happened to be teaching the representative samples of students selected. That is, the teachers were not selected and therefore do not constitute a representative sample of teachers from each country. The U.S. sample included 7,623 students and 359 teachers.
The sample sizes for other countries were presented in the descriptive statistics of the Results section.

**Sampling Weights**

“The sampling weight is the inverse of this selection probability. In a properly selected and weighted sample, the sum of the weights for the sample approximates the size of the population” (Martin, 2005 p. 45). In the TIMSS 2003, students in each country had different probability of selection due to the probability-proportional-to-size sampling. Sampling weights were used for more accurate estimates of population parameters. The sampling weights reflect the probability of selection of each school, classroom and student, take into account any stratification or disproportional sampling of subgroups, and include adjustments for non-response (Joncas, 2004).

For the TIMSS 2003, sampling weights were calculated according to a three-step procedure involving selection probabilities for schools, classrooms, and students. Each of these sampling units had a corresponding sampling weight. In most cases, different schools had different weights based on their sizes; however, classroom weights were set to unity in cases where a single classroom was sampled in each school. Classrooms within schools generally were selected with equal probabilities. Several countries chose to sample more than one classroom from each sampled school. For most countries, because the entire classroom was sampled, each student in the sampled classrooms had the same student weight which was 1.0. The basic sampling weight attached to each student record was the product of the school, classroom, and student weights (Joncas, 2004). For the current study, the House Weight (HOUWGT) was used at the level-1 in multilevel analyses.

**Plausible Values**

Since each student is administered only a fraction of the items in the test, the TIMSS 2003 utilized plausible values to obtain scores for each student. A plausible value is an estimate of how the individual student would have performed on a test that included all possible items. The process derives performance distributions for each student using information about performance of other students with similar background characteristics. (Martin & Mullis, 2004).
Due to the complexity of the TIMSS data sets, raw scores for the dependent variable were not used in this study. Rather, five plausible values were used in the HLM analyses. The HLM takes the plausible values into account in generating the HLM estimates. For each HLM model, the program runs each of the five (or the number specified) plausible values internally, and produces their average value and the correct standard errors (Raudenbush, Bryk, Cheong, & Congdon, 2004). For independent variables, however, raw scores of the individual items were used.

Variables

Guided by previously developed models of school learning this study examined some student-level and classroom-level factors that pertain to children’s learning. Three variables at the student-level and nine variables at the classroom-level are included as independent variables. Science achievement score (5 plausible values) at fourth grade on the TIMSS 2003 test was the dependent variable.

The TIMSS data provided some information about student and home characteristics for each country. For this study, a range of specific student-level variables that were potentially associated with student achievement were considered. As for the student variables, student gender (GENDER), home resources (HOMERES) and science self-confidence (SELFCON) were chosen primarily because the school learning models and the science education literature emphasizes their influences on student achievement. Home resources variable was an index based on students’ responses to questions about educational resources (number of books at home, having a computer, and study desk) at home. Higher scores indicate students have more educational resources at home. Science self-confidence is another index created by the TIMSS 2003.

Based on the literature, three teacher characteristics related to their educational and professional attainment were drawn from the TIMSS 2003 teacher questionnaire. These variables are teaching experience (TEXP), the highest level of formal education (TEDUC), and the major area of study (TMAJOR). The teacher support (TSUPPORT) variable was created based on students’ responses to the statements about teacher care and expectation. Raw scores were added to create this variable.

The TIMSS 2003 emphasized five instructional variables as science inquiry (SCINQ) variables. These variables were parallel with National Science Education
Standards’ guidelines on quality science instruction (NRC, 1996). To assure the construct validity, factor analysis was conducted on these five variables for each country. Accordingly, the frequency of science experiments done by teachers, science experiments planned by students, science experiments done by students, working in groups, and giving explanations of a phenomenon were analyzed by using explanatory factor analysis procedure. The amount of science instruction was also emphasized by NSES and by the learning models discussed in Chapter 2. Thus, the amount of weekly science instruction (SCIMIN) was used as another instructional variable. Class size (CLASSIZE) was used as a possible mediating variable. Class mean of home resources (M_HOMERES) and science self-confidence (M_SELFCON) were also included at the classroom-level. The variables are displayed in Table 3.1.

Table 3.1 List of Variables

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
</tr>
<tr>
<td>SCIENCE PVs</td>
<td>Science plausible values for the TIMSS 2003 4th grade</td>
</tr>
<tr>
<td><strong>Student-Level</strong></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>Gender of Student</td>
</tr>
<tr>
<td>SELFCON</td>
<td>Science Self-Confidence</td>
</tr>
<tr>
<td>HOMERES</td>
<td>Home Resources</td>
</tr>
<tr>
<td><strong>Classroom-Level</strong></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>Teacher’s Experience</td>
</tr>
<tr>
<td>TEDUC</td>
<td>Teacher’s Education</td>
</tr>
<tr>
<td>TMAJOR</td>
<td>Teacher’s Major</td>
</tr>
<tr>
<td>TSUPPORT</td>
<td>Teacher Support</td>
</tr>
<tr>
<td>SCINQ</td>
<td>Science Inquiry</td>
</tr>
<tr>
<td>SCIMIN</td>
<td>Minutes of Weekly Science Instruction</td>
</tr>
<tr>
<td>CLASSIZE</td>
<td>Class Size</td>
</tr>
<tr>
<td>M_SELFCON</td>
<td>Class Mean of Science Self-Confidence</td>
</tr>
<tr>
<td>M_HOMERES</td>
<td>Class Mean of Home Resources</td>
</tr>
</tbody>
</table>

Following are the descriptions of student variables:

**Gender (GENDER):** Gender of student is a dichotomous variable [1= Female, 0= Male].

**Science Self-Confidence (SELFCON):** This index is based on students’ responses to four statements about their science ability: (a) I usually do well in science; (b) Science is
more difficult for me than for many of my classmates; (c) Science is not one of my strengths; (d) I learn things quickly in science.

The index was computed from students’ responses to the statements above on a 4-point Likert scale (1= Agree a lot, 2= Agree a little, 3= Disagree a little, 4= Disagree a lot). The science self-confidence index had three categories:
1 = High: Average is less than or equal to 2.
2 = Medium: Average is greater than 2 and less than 3.
3 = Low: Average is greater than or equal to 3 (Martin, 2005).

The index was reversed so that higher values mean higher self-confidence in science.

**Home Resources (HOMERES):** The index based on students' responses to three questions about home resources: (a) About how many books are there in your home? [1= None or very few (0-10 books), 2= Enough to fill one shelf (11-25 books), 3= Enough to fill one bookcase (26-100 books), 4= Enough to fill two bookcases (101-200 books), 5= Enough to fill three bookcases or more (more than 200 books)]; and (b) Do you have these items at home – computer and study desk/table for your own use? [1= Yes, 0= No].

The ‘High level on this index indicates more than 100 books in the home, computer and study desk. The ‘Low’ level indicates 25 or fewer books in the home and no computer and study desk. The ‘Medium’ level includes all other possible combinations of responses (Xin et al., 2004).

At the classroom-level, teacher characteristics as well as instructional variables, reported either by students or teachers, were introduced into the model. Using the same teacher background variables for all countries might be challenging since each country has different requirements for teachers. Comparative studies found cultural differences in teachers’ roles and identities across countries (Anderson-Levitt, 2001; LeTendre, 1999). However, recent research showed considerable homogenization of curricula within core subject areas across nations (Benavot & Braslavsky, 2006; LeTendre, Baker, Akiba, Goesling, & Wiseman, 2001; Stigler & Hiebert, 1999). Furthermore, a report by Wang and colleagues (2003) indicated that the requirements for teachers and the structure of teacher education programs were similar across eight countries, Australia, England, Hong Kong, Japan, Korea, the Netherlands, Singapore, and the United States. Therefore, a cross-national comparison by using the same teacher characteristics in each country seems reasonable.
Teacher’s Experience (TEXP): Number of years as a teacher (Continuous variable).
Teacher’s Education (TEDUC): The highest level of formal education completed by the teacher (1= Have master’s or doctoral degree, 0= Do not have master’s or doctoral degree).
Teacher’s Major (TMAJOR): Teacher’s major area of study during post-secondary education (1= Science or Math major, 0= Non-Science or Non-Math major).

For creating teacher support (TSUPPORT) and science inquiry (SCINQ) variables student questionnaires were used. Although teachers might report that they make efforts to meet students’ needs and provide appropriate instruction, students’ perceptions and interpretations of these efforts are more crucial (Daniels & Perry, 2003).

Teacher Support (TSUPPORT): The composite score of students’ responses to the following statements: (a) Teacher cares about the students, and (b) Teacher wants students to do their best [1= Disagree a lot, 2= Disagree a little, 3= Agree a little, 4= Agree a lot].

Science Inquiry (SCINQ): The composite score of students’ responses to the following statements about weekly science activities: (a) We watch teacher do science experiment, (b) We plan science experiments or investigations, (c) We do science experiments or investigations, (d) We work in groups experiments or investigations, (e) We give explanations for something we are studying in science [1= Never, 2= A few times a year, 3= Once or twice a month, 4= At least once a week]. Figure 3.1 describes the factorial structure of science inquiry.
Table 3.2 *Factor Loadings and Item Means and Standard Deviations for the Measure of Science Inquiry*

<table>
<thead>
<tr>
<th>Factor 1 (Science Inquiry)</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. We watch teacher do science experiment</td>
<td>0.48 3.29(0.83)</td>
<td>0.35 3.36(0.78)</td>
<td>0.57 2.84(1.07)</td>
<td>0.58 2.80(1.04)</td>
<td>0.58 2.91(1.02)</td>
</tr>
<tr>
<td>2. We plan science experiments or investigations</td>
<td>0.63 2.04(1.05)</td>
<td>0.59 3.05(0.93)</td>
<td>0.66 2.30(1.09)</td>
<td>0.72 2.45(1.04)</td>
<td>0.68 2.41(1.07)</td>
</tr>
<tr>
<td>3. We do science experiments or investigations</td>
<td>0.74 2.44(1.06)</td>
<td>0.64 3.06(0.95)</td>
<td>0.77 2.57(1.04)</td>
<td>0.78 2.57(1.00)</td>
<td>0.75 2.56(1.02)</td>
</tr>
<tr>
<td>4. We work in group experiments or investigations</td>
<td>0.56 2.88(1.02)</td>
<td>0.60 3.44(0.77)</td>
<td>0.61 2.85(1.08)</td>
<td>0.68 2.81(1.02)</td>
<td>0.55 2.77(1.06)</td>
</tr>
<tr>
<td>5. We give explanations for something we study in science</td>
<td>0.47 2.82(1.13)</td>
<td>0.49 3.19(0.90)</td>
<td>0.47 3.06(1.00)</td>
<td>0.61 2.91(0.96)</td>
<td>0.55 2.92(1.06)</td>
</tr>
</tbody>
</table>

Variance Explained for 1-Factor Solution (%)  
46.69 43.07 51.10 56.18 51.10

*Note.* M = mean; SD = standard deviation
To examine the construct validity of science instructional variables defined by the TIMSS 2003, exploratory factor analysis was conducted on the five instructional variables in each country. Exploratory factor analysis is used to explore data to determine the number and nature of factors that account for the co-variation between variables when the researcher does not have an a priori hypothesis about the number of factors underlying the data. Therefore, exploratory factor analysis is generally thought of as more of a theory-generating procedure (Stevens, 1996). The analysis extracted only one factor with maximum likelihood extraction method in each country which indicated that using Science Inquiry as a single construct for those five variables would be appropriate. The factor loadings and explained variances for each country were reported in Table 3.2. Figure 3.1 displays the factorial structure of science inquiry variables.

**Figure 3.1 Factorial Structure of Science Inquiry**

- **Minutes of Weekly Science Instruction (SCIMIN):** Teacher’s response to the following open-ended question: How many minutes per week do you teach science? (Continuous variable).
- **Class size (CLASSIZE):** Teachers’ response to the following question: How many students are in class for science instruction? [1= 1-19 students, 2= 20-26 students, 3= 27-32 students, 4= more than 33 students].
- **Class Mean of Science Self-Confidence (M_HOMERES):** Mean value of students’ self-confidence in science for each classroom. This value was computed by dividing the
total score of students’ self-confidence in each classroom by the number of students in that classroom.

**Class Mean of Home Resources (M_SELFCON):** Mean value of students’ home resources for each classroom. This value was computed by dividing the total score of students’ home resources in each classroom by the number of students in that classroom.

**Data Analysis**

Education studies commonly have nested data structures, which is the structure in this study. In the TIMSS 2003, students were nested within classrooms and classrooms were nested within schools. Many previous non-experimental studies on classroom effects treated classroom and teacher variables as individual student’s characteristics (Pong & Pallas, 2001). Using a standard multiple linear regression analysis at the student-level to examine sources of influence related to student performance would be problematic because there are multilevel influences (i.e., student-level, classroom-level) that need to be analyzed using multilevel models (Mendro, 1998; Raudenbush & Bryk, 2002). The estimated standard error is usually much smaller when the individual student is treated as an independent observation. Students sampled from the same classroom share many classroom characteristics. Therefore, a hierarchical approach that partials student and classroom variances separately is more appropriate (Raudenbush & Bryk).

The hierarchical linear modeling or the multilevel approach is capable of analyzing data simultaneously at different levels of the educational hierarchy (i.e., student, classroom, school levels) while the one level regression model uses the assumption of independent observations and ignores the nesting of students. Hierarchical linear modeling determines the error variance for the individual student-level as well as the classroom-level (Arnold & Sedlacek, 1995; Raudenbush & Bryk, 2002). HLM specifies a statistical model with a unique random effect for each organizational unit and variability in each level random effect is taken into account when computing the standard errors. Applying maximum likelihood procedure, the HLM allows the estimation of student-level and classroom-level effects simultaneously by hypothesizing relationships at both the student and the aggregate level between classrooms (Raudenbush & Bryk).

Within- and between-teacher variances of student achievement were compared through multi-level data analyses. The within-class variance is due to individual student
factors. Between-class variance might be accounted for by classroom-level factors. In other words, different achievement scores among classes might be attributed to the variables related to classroom variables (Xin et al. 2004).

For data analyses, a multiple imputation HLM procedure was used in which results from the analyses with five plausible values were combined to estimate parameters of correlates of science achievement (Arnold & Sedlacek, 1995; Raudenbush & Bryk, 2002). Data were analyzed by using the HLM software version 6.07 and SPSS 15. The HLM analyses were carried out for each country in several stages and then with the data from all five countries. Missing data was handled by list wise deletion and mean imputation. Prior to actual analyses, inter-correlation matrix, and descriptive statistics of the data are reported to examine the nature of the data and multi-collinearity among variables.

A Two-Level Hierarchical Model

A two-level HLM analysis in which students are the level-1 units and classrooms are the level-2 units was conducted in this study. Hierarchical linear models follow the three-stage approach of multi-level analysis. In the first stage, the analysis produces the ‘unconditional’ model with no independent variables at the student and the classroom levels. With only the student-level outcome measure, this model is similar to a random effect ANOVA model, providing a measure of the variances within and between classrooms for science achievement. At the second stage (random coefficients model), student-level variables are added to the unconditional model to determine whether their relationships with achievement varied significantly across classrooms. The last stage refers to the ‘full’ or ‘conditional’ model (intercepts and slopes as outcomes) when classroom-level variables (teacher and instructional variables) are added to the model (Raudenbush & Bryk, 2002).

At the first stage of the analysis, results from the following basic model were used to determine how much of the variation in students’ science scores is at the student-level or within-class and how much is at the classroom-level or between-class.

Unconditional Model (Model 1)

Level-1 (Student-level)

\[ Y_{ij} = Y_{ij}^{(SCIENCE PVs_{ij})} = \beta_{0j} + r_{ij} \]  

(1)
Where, \( i \) is a subscript for an individual student and \( j \) for the classroom. \( Y_{ij} \) represents the science score of the student \( i \) in the school \( j \). \( \beta_{ij} \) is the function of average science score and the classroom-level variance for the students in classroom \( j \). The error variance at the student-level is represented by \( r_{ij} \).

**Level-2 (Classroom-level)**

\[
\beta_{0j} = \gamma_{00} + u_{0j}
\]

Where,

\( \gamma_{00} \) = The intercept or the grand mean of the dependent variable (i.e., science score) for all classrooms.
\( u_{0j} \) = The error variance at the classroom-level.

The statistics of interest in this stage are the variance components. The sum of level-1 and level-2 variance provides the total variance. The proportion of level-1 and level-2 variances then is computed to obtain the proportion of within-class (\( Var(r_{ij}) = \sigma^2 \)) and between-class (\( Var(u_{0j}) = \tau_{00} \)) variances. The proportion of the total variance that is between classes is called the intra-class correlation (ICC):

\[
\rho \text{ (Unconditional)} = \frac{\tau_{00}(\text{Unconditional})}{\tau_{00}(\text{Unconditional}) + \sigma^2(\text{Unconditional})}
\]

Where, \( \sigma^2 \) is the level-1 variance and \( \tau_{00} \) is the level-2 variance component.

For example, the intra-class correlation, \( \rho = .24 \), shows that 24 percent of the variation in student achievement is between classes and the remaining 76 percent is within classes, at the student-level (Raudenbush & Bryk, 2002).

In the second stage, student-level variables were introduced into the level-1 equation to examine the relations between science achievement and the selected student variables. At the level-1, variables are centered around the group mean (\( X_{ij} - \overline{X}_{.j} \)) and at the level-2 they were centered around the grand mean (\( X_{ij} - \overline{X}_{..} \)). Centering method makes the intercept more meaningful and eases the interpretation of results. When no centering is applied and the raw scores are used for all the variables, the intercept provides the expected outcome when all independent variables are set equal to zero. With group mean centering, the intercept would be the expected outcome when all independent variables are set equal to their group means. With grand mean centering, the intercept
would be the expected outcome for a subject whose values on the other independent variables are set equal to the grand mean (Raudenbush & Bryk, 2002). Following are the level-1 (student-level) and level-2 (classroom-level) equations based on the hypothesized HLM model.

Random Coefficients Model (Model 2)

Level-1

\[ \text{SCIENCE PVs}_{ij} = \beta_{0j} + \beta_{1j}(\text{GENDER}_i) + \beta_{2j}(\text{SELFCON}_i) + \beta_{3j}(\text{HOMERES}_i) + r_{ij} \]  

Where,

\[ \text{SCIENCE PVs}_{ij} = \text{Science achievement of student } i \text{ in classroom } j. \]
\[ r_{ij} = \text{The error variance at the student-level.} \]

Level-2

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} \]
\[ \beta_{2j} = \gamma_{20} \]
\[ \beta_{3j} = \gamma_{30} \]

Where,

\[ \gamma_{00} = \text{The intercept or the grand mean of the science score for all classrooms.} \]
\[ \gamma_{10} = \text{The effect of GENDER on science achievement in classroom } j. \]
\[ \gamma_{20} = \text{The effect of SELFCON on science achievement in classroom } j. \]
\[ \gamma_{30} = \text{The effect of HOMERES on science achievement in classroom } j. \]
\[ u_{0j} = \text{The error variance at the classroom-level.} \]

The coefficients and the error variances were examined to understand the nature of the relations. The reduction of the level-1 variance in the basic model was computed to obtain the proportion of the variance explained by student-level variables. With different variables entered in the level-1 model, changes in the within-class random variance were examined by using the following formula:

\[ \text{Reduction} = \frac{\sigma^2(\text{Unconditional}) - \sigma^2(\text{Model2})}{\sigma^2(\text{Unconditional})} \]

In the third stage, classroom-level variables were introduced in the classroom-level equation to examine the relations between science achievement and the selected classroom-level variables. Following are the level-1 (student-level) and level-2 (classroom-level) equations based on the hypothesized HLM model.
**Conditional Model (Model 3)**

**Level-1**

\[ \text{SCIENCE PVs}_{ij} = \beta_{0j} + \beta_{1j}(\text{GENDER}_i) + \beta_{2j}(\text{SELFCON}_i) + \beta_{3j}(\text{HOMERES}_i) + r_{ij} \]

**Level-2**

\[
\begin{align*}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\text{TEXP})_j + \gamma_{03}(\text{TEDUC})_j + \gamma_{04}(\text{TMAJOR})_j + \gamma_{05}(\text{SCINQ})_j + \gamma_{06}(\text{SCIMIN})_j + \gamma_{07}(\text{CLASSIZE})_j + \gamma_{08}(\text{M_SELFCON})_j + \gamma_{09}(\text{M_HOMERES})_j + u_{0j} \\
\beta_{1j} &= \gamma_{10} \\
\beta_{2j} &= \gamma_{20} \\
\beta_{3j} &= \gamma_{30}
\end{align*}
\]

Where,

- \( \gamma_{00} \) = The intercept or the grand mean of the science scores for all classrooms.
- \( \gamma_{01} \) = The effect of TEXP on class mean achievement (the effect of one year change in experience on mean achievement across classes).
- \( \gamma_{02} \) = The effect of TEDUC on class mean achievement.
- \( \gamma_{03} \) = The effect of TMAJOR on class mean achievement.
- \( \gamma_{04} \) = The effect of TSUPPORT on class mean achievement.
- \( \gamma_{05} \) = The effect of SCINQ on class mean achievement.
- \( \gamma_{06} \) = The effect of SCIMIN on class mean achievement.
- \( \gamma_{07} \) = The effect of CLASSIZE on class mean achievement.
- \( \gamma_{08} \) = The effect of M_SELFCON on class mean achievement.
- \( \gamma_{09} \) = The effect of M_HOMERES on class mean achievement
- \( u_{0j} \) = The error variance at the classroom-level.

As with level-1 analysis, both the fixed effects and the variance components were examined to understand the nature of the relations. The effects and the variance components were estimated with robust standard errors in the HLM analyses. Robust standard errors are relatively insensitive to misspecification at the levels of the model and the distributional assumptions at each level. If the robust and model-based standard errors differ substantially, that indicates model misspecifications and requires further investigation of HLM assumptions. The robust standard errors are appropriate for datasets having a moderate to large number of level 2 units (Raudenbush & Bryk, 2002).

The results were used to identify some variables along with interactions that significantly predict students’ science outcomes and how much they account for the between-class variance. The proportion of variance explained by the final model was computed by subtracting the total variance in the final model (conditional) from the total
variance in the unconditional model and dividing by the total variance in the unconditional model.

\[
\text{Reduction} = \frac{\tau_{00} \text{(Unconditional)} - \tau_{00} \text{(Conditional)}}{\tau_{00} \text{(Unconditional)}}
\]

Finally, data from all five countries was analyzed in a similar fashion, by using 2-level HLM to make cross-country comparisons. The same three variables were used at the student-level (GENDER, SELFCON, HOMERES). Country specific variables were created and the variables were centered around the grand mean. Followings are the equations at the student- and classroom-level for this analysis.

**Level-1**

\[
\text{SCIENCE PVs}_{ij} = \beta_{1j} \text{(SING)} + \beta_{2j} \text{(JAPAN)} + \beta_{3j} \text{(US)} + \beta_{4j} \text{(AUST)} + \beta_{5j} \text{(SCOT)} + \beta_{6j} \text{(SING\_GENDER)}_{i} + \beta_{7j} \text{(JAPAN\_GENDER)}_{i} + \beta_{8j} \text{(US\_GENDER)}_{i} + \beta_{9j} \text{(AUST\_GENDER)}_{i} + \beta_{10j} \text{(SCOT\_GENDER)}_{i} + \ldots + r_{ij}
\]

**Level-2**

\[
\beta_{1j} = \gamma_{10} + u_{1j} \\
\beta_{2j} = \gamma_{20} + u_{2j} \\
\beta_{3j} = \gamma_{30} + u_{3j} \\
\beta_{4j} = \gamma_{40} + u_{4j} \\
\beta_{5j} = \gamma_{50} + u_{5j} \\
\beta_{6j} = \gamma_{60} \\
\beta_{7j} = \gamma_{70} \\
\beta_{8j} = \gamma_{80} \\
\beta_{9j} = \gamma_{90} \\
\beta_{10j} = \gamma_{100} \\
\ldots
\]

Where,

\[
\gamma_{10} = \text{The grand mean of science scores for Singapore} \\
\gamma_{20} = \text{The grand mean of science scores for Japan} \\
\gamma_{30} = \text{The grand mean of science scores for the U.S.} \\
\gamma_{40} = \text{The grand mean of science scores for Australia} \\
\gamma_{50} = \text{The grand mean of science scores for the Scotland} \\
\gamma_{60} = \text{The effect of GENDER on science achievement in Singapore compared to other countries} \\
\gamma_{70} = \text{The effect of GENDER on science achievement in Japan compared to other countries} \\
\gamma_{80} = \text{The effect of GENDER on science achievement in the U.S. compared to other countries}
\]
\( \gamma_{90} = \) The effect of GENDER on science achievement in Australia compared to other countries
\( \gamma_{100} = \) The effect of GENDER on science achievement in the Scotland compared to other countries

\[ \ldots \]
\( u_{ij} - u_{5j} = \) Random error of mean achievement in each country.

Based on the country-specific analyses results, at the classroom-level, the effects of teacher support (TSUPPORT), science inquiry (SINQ), class mean of science self-confidence (M_SELFCON) and home resources (M_HOMERES) were compared in five countries. Followings are the equations at the student- and classroom-level for the final analysis.

**Level-1**

\[
\text{SCIENCE PVS}_{ij} = \beta_0 + \beta_1(GENDER_i) + \beta_2(SELFCON_i) + \beta_3(HOMERES_i) + r_{ij}
\]

**Level-2**

\[
\beta_0 = \gamma_{01}(\text{SING})_j + \gamma_{03}(\text{JAPAN})_j + \gamma_{05}(\text{US})_j + \gamma_{04}(\text{AUST})_j + \gamma_{05}(\text{SCOT})_j + \gamma_{06}(\text{SING}\_\text{TSUPPORT})_j + \gamma_{07}(\text{JAPAN}\_\text{TSUPPORT})_j + \gamma_{08}(\text{US}\_\text{TSUPPORT})_j + \gamma_{09}(\text{AUST}\_\text{TSUPPORT})_j + \gamma_{10}(\text{SCOT}\_\text{TSUPPORT})_j + \ldots + u_{0j}
\]

Where,

\( \gamma_{01} = \) The grand mean of science scores for Singapore
\( \gamma_{02} = \) The grand mean of science scores for Japan
\( \gamma_{03} = \) The grand mean of science scores for the U.S.
\( \gamma_{04} = \) The grand mean of science scores for Australia
\( \gamma_{05} = \) The grand mean of science scores for the Scotland
\( \gamma_{06} = \) The effect of TSUPPORT on class mean achievement in Singapore compared to other countries
\( \gamma_{07} = \) The effect of TSUPPORT on class mean achievement in Japan compared to other countries
\( \gamma_{08} = \) The effect of TSUPPORT on class mean achievement in the U.S. compared to other countries
\( \gamma_{09} = \) The effect of TSUPPORT on class mean achievement in Australia compared to other countries
\( \gamma_{10} = \) The effect of TSUPPORT on class mean achievement in the Scotland compared to other countries

\[ \ldots \]
\( u_{0j} = \) Classroom specific random error on mean achievement of five countries.
Assumptions and Limitations

This study assumed that participants of TIMSS 2003 study answered the questionnaire items in a manner that reflect their true feelings and correct information about themselves. It is also assumed that the sample of elementary school students participating in the TIMSS 2003 was the representative of the elementary school population in each country.

Furthermore, it is assumed that the science plausible values, which constitute the outcome variable of the current study, are reliable measures to estimate population characteristics, and they are more appropriate than point estimates of achievement. Wu (2005) pointed out that plausible values perform well in recovering population parameters such as the mean, variance and percentiles, even when short tests are administered. For this study, multiple imputation HLM procedure was used in which results from the analyses with five plausible values were combined to estimate parameters of correlates of science achievement (Arnold & Sedlacek, 1995; Raudenbush et al., 2002).

In terms of the HLM analyses, several key assumptions were held in this study: (a) each \( r_{ij} \) is independent and normally distributed with a mean of zero and variance \( \sigma^2 \) for every level-1 unit \( i \) within each level-2 unit; (b) the level-1 predictors \( X_{ij} \) are independent of \( r_{ij} \); (c) the random errors at level-2 (\( u_1, u_2 \), etc.) are multivariate normal, each with a mean of zero and some variance \( \tau_{qq} \). The random error vectors are independent among the \( j \) level-2 units; (d) the level-2 predictors are independent of every \( u_{ij} \); (e) the predictors at each level are not correlated with the random effects at the other level; (f) the errors at level-1, \( r_{ij} \) and level-2, \( u_{ij} \) are also independent (Raudenbush & Bryk, 2002).

The sampling procedures utilized in data collection for the TIMSS 2003 and the method of this study revealed important associations between student- and classroom-level factors and science achievement. However, there were some limitations which should be kept in mind when interpreting the results. First, some of the data used in this study was questionnaires from students and teachers who participated in the TIMSS 2003. Thus, the study is exploratory in nature and could not provide direct evidence about causal effects. This is a major limitation of all studies that use questionnaire survey
method. The variables, especially the ones related to science instructional practices, might be crude and may not reflect the real classroom applications.

Second, the TIMSS does not provide a ‘value added model’ in which students’ previous performance is taken into account. Therefore, students’ prior academic level could not be used in HLM analyses. In value added models, improvement in performance could be more easily related to classroom and school inputs (Hanushek, 1997). On the contrary, in studies that do not use previous achievement, student performance correlates with other variables (Raudenbush & Bryk, 2002; Hanushek; Xin et al., 2004). If successful students, for example, were assigned to teachers with higher qualifications, one would expect an upward bias for teacher effect (Xin et al.).

Third, the TIMSS 2003 fourth grade data did not provide adequate information about family background, especially SES and parents’ education. Research indicates that family background variables explain a substantial amount of variance in student achievement (Goldhaber, 1999; Lamb & Fullarton, 2002; Ma & Klinger, 2004). Therefore, this study failed to explain a large portion of variance at the student-level.
CHAPTER IV

RESULTS

This chapter presents the results of the hierarchical linear modeling (HLM) analyses and the interpretations of the results. It is organized around the research questions presented in Chapter 1. Descriptive statistics of the variables were presented in the first section. In the following sections, variance components, the fixed effects, effect sizes, and explained variances were examined within and across Singapore, Japan, the U.S, Australia, and Scotland. Interactions between student-level and classroom-level variables were also investigated. Hypothesis testing of the student- and classroom-level effects across countries were reported in the last section.

Descriptive Statistics

This study examined a number of student-level and classroom-level variables with respect to their effects on fourth grade science achievement on the TIMSS 2003 assessment in five countries. The descriptive statistics for each country were presented in Table 4.1. Science achievement is represented as the average of five plausible values.
### Table 4.1 Descriptive Statistics of Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 6,122 (Level-1)</td>
<td>n = 4,250 (Level-1)</td>
<td>n = 7,623 (Level-1)</td>
<td>n = 3,573 (Level-1)</td>
<td>n = 2,665 (Level-1)</td>
</tr>
<tr>
<td></td>
<td>n = 171 (Level-2)</td>
<td>n = 144 (Level-2)</td>
<td>n = 359 (Level-2)</td>
<td>n = 206 (Level-2)</td>
<td>n = 120 (Level-2)</td>
</tr>
<tr>
<td><strong>Student-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science Achievement (Average of 5 PVs)</td>
<td>564.57 80.34</td>
<td>544.05 68.15</td>
<td>530.50 75.93</td>
<td>525.73 74.62</td>
<td>510.02 72.11</td>
</tr>
<tr>
<td>Gender</td>
<td>0.50 0.50</td>
<td>0.49 0.50</td>
<td>0.50 0.50</td>
<td>0.51 0.50</td>
<td>0.51 0.50</td>
</tr>
<tr>
<td>Self-Confidence</td>
<td>2.05 0.77</td>
<td>2.33 0.69</td>
<td>2.56 0.64</td>
<td>2.61 0.61</td>
<td>2.48 0.70</td>
</tr>
<tr>
<td>Home Resources</td>
<td>2.22 0.45</td>
<td>2.16 0.40</td>
<td>2.22 0.47</td>
<td>2.40 0.51</td>
<td>2.30 0.50</td>
</tr>
<tr>
<td><strong>Classroom-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Experience</td>
<td>10.82 12.66</td>
<td>20.21 9.38</td>
<td>13.23 10.36</td>
<td>16.46 9.89</td>
<td>15.36 10.29</td>
</tr>
<tr>
<td>Teacher’s Education</td>
<td>0.01 0.08</td>
<td>0.03 0.18</td>
<td>0.53 0.50</td>
<td>0.26 0.44</td>
<td>0.12 0.32</td>
</tr>
<tr>
<td>Teacher’s Major</td>
<td>0.50 0.50</td>
<td>0.20 0.40</td>
<td>0.15 0.36</td>
<td>0.21 0.41</td>
<td>0.13 0.34</td>
</tr>
<tr>
<td>Teacher Support</td>
<td>7.51 0.32</td>
<td>6.84 0.44</td>
<td>7.65 0.31</td>
<td>7.55 0.34</td>
<td>7.58 0.36</td>
</tr>
<tr>
<td>Science Inquiry</td>
<td>13.47 1.58</td>
<td>16.07 1.13</td>
<td>13.53 1.95</td>
<td>13.52 2.20</td>
<td>13.65 2.09</td>
</tr>
<tr>
<td>Min. of Sci. Inst.</td>
<td>123.00 14.88</td>
<td>120.48 19.36</td>
<td>137.00 68.86</td>
<td>71.58 39.52</td>
<td>88.36 56.34</td>
</tr>
<tr>
<td>Class Size</td>
<td>3.88 0.45</td>
<td>3.26 0.93</td>
<td>2.00 0.75</td>
<td>2.37 0.76</td>
<td>2.50 0.83</td>
</tr>
<tr>
<td>Class Mean Self-Con.</td>
<td>2.06 0.20</td>
<td>2.32 0.20</td>
<td>2.56 0.23</td>
<td>2.60 0.21</td>
<td>2.48 0.21</td>
</tr>
<tr>
<td>Class Mean Home Res.</td>
<td>2.21 0.16</td>
<td>2.16 0.11</td>
<td>2.21 0.19</td>
<td>2.37 0.23</td>
<td>2.30 0.19</td>
</tr>
</tbody>
</table>

*Note. n = sample size; SD = standard deviation*
Variation in Science Achievement: Within vs. Between Classrooms

To determine the partitioning of variation in science achievement at the fourth grade within-classrooms and between-classrooms, a random ANOVA model was run with HLM. No variable was introduced at either the student- or classroom-level. The following unconditional model was used to examine the variations in each country:

**Level 1 (Student-Level):**

\[ \text{SCIENCE PV}_{ij} = \beta_{0j} + r_{ij} \]

Where \( r_{ij} \sim N(0, \sigma^2) \)

**Level 2 (Classroom-Level):**

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]

Where \( u_{0j} \sim N(0, \tau_{00}) \)

Where \( r_{ij} \) is the error variance for student \( i \) in classroom \( j \), and \( u_{0j} \) is the error variance for classroom \( j \). Sigma squared (\( \sigma^2 \)) is the student-level random error variance and tau (\( \tau_{00} \)) is the classroom-level random error variance. The sum of \( \sigma^2 \) and \( \tau_{00} \) equals the total variance. Table 4.2 lists the maximum likelihood estimates of the variance components at the student- and classroom-level in five countries.

**Table 4.2 Maximum Likelihood Estimates of the Variance Components at the Student-and Classroom-Level**

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>SD</th>
<th>Variance Component</th>
<th>df</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT1, ( \tau_{00} )</td>
<td>63.36</td>
<td>4013.87</td>
<td>170</td>
<td>6507.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level-1, ( \sigma^2 )</td>
<td>59.77</td>
<td>3572.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT1, ( \tau_{00} )</td>
<td>13.80</td>
<td>190.45</td>
<td>143</td>
<td>312.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level-1, ( \sigma^2 )</td>
<td>71.13</td>
<td>5058.77</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT1, ( \tau_{00} )</td>
<td>44.13</td>
<td>1947.55</td>
<td>358</td>
<td>3797.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level-1, ( \sigma^2 )</td>
<td>65.57</td>
<td>4298.83</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT1, ( \tau_{00} )</td>
<td>39.45</td>
<td>1550.29</td>
<td>204</td>
<td>1267.30</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level-1, ( \sigma^2 )</td>
<td>71.60</td>
<td>5124.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INTERCEPT1, ( \tau_{00} )</td>
<td>30.84</td>
<td>951.24</td>
<td>199</td>
<td>661.78</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Level-1, ( \sigma^2 )</td>
<td>70.16</td>
<td>4922.95</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SD = standard deviation; df = degrees of freedom*
The proportion of variance in science achievement within- and between-classrooms is computed as below:

\[ \rho \text{ (Within-Class)} = \frac{\sigma^2}{\tau_{00} + \sigma^2} \times 100\% \quad \text{and} \]

\[ \rho \text{ (Between-Class)} = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \times 100\% \]

Accordingly, the proportion of student-level or within-classroom variance in the U.S.:

\[ \rho \text{ (Within-Class)} = \frac{4298.83}{1947.55 + 4298.83} \times 100\% = 69\% \]

The proportion of classroom-level or between-classroom variance in the U.S.:

\[ \rho \text{ (Between-Class)} = \frac{1947.55}{1947.55 + 4298.83} \times 100\% = 31\% \]

For the U.S. fourth grade classrooms, the intra-class correlation (ICC) was 0.31. In other words, in U.S. fourth grade classrooms, 69 percent of the variance in science achievement was attributable to the differences among students in the same classroom and 31 percent of the variance was attributable to the differences between classrooms. Table 4.7 lists the proportion of variance in science achievement within- and between-classrooms in the other countries in the study.

Within-class variance in Japan, the U.S., Australia, and Scotland was larger than between-class variance. Especially in Japan, most of the variance was within classrooms, among the students in the same classroom, taught by the same teacher. Only in Singapore, was within-class variance (47 percent) smaller than between-class variance (53 percent). The next section describes the effects of the student-level variables on science achievement in each country.
### Table 4.3 The Proportion of Variance in Science Achievement Within and Between Classrooms

<table>
<thead>
<tr>
<th>Country</th>
<th>Random Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>T-ratio</th>
<th>Within Class Variance (%)</th>
<th>Between Class Variance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singapore</td>
<td>INTERCEPT1, β&lt;sub&gt;0&lt;/sub&gt;</td>
<td>563.02</td>
<td>5.38</td>
<td>104.78</td>
<td>47.09</td>
<td>52.91</td>
</tr>
<tr>
<td></td>
<td>INTERCEPT1, γ&lt;sub&gt;00&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>INTERCEPT1, β&lt;sub&gt;0&lt;/sub&gt;</td>
<td>544.00</td>
<td>1.80</td>
<td>301.51</td>
<td>96.37</td>
<td>3.63</td>
</tr>
<tr>
<td></td>
<td>INTERCEPT1, γ&lt;sub&gt;00&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S.</td>
<td>INTERCEPT1, β&lt;sub&gt;0&lt;/sub&gt;</td>
<td>538.24</td>
<td>3.06</td>
<td>176.15</td>
<td>68.82</td>
<td>31.18</td>
</tr>
<tr>
<td></td>
<td>INTERCEPT1, γ&lt;sub&gt;00&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>INTERCEPT1, β&lt;sub&gt;0&lt;/sub&gt;</td>
<td>522.88</td>
<td>4.48</td>
<td>116.66</td>
<td>76.71</td>
<td>23.29</td>
</tr>
<tr>
<td></td>
<td>INTERCEPT1, γ&lt;sub&gt;00&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scotland</td>
<td>INTERCEPT1, β&lt;sub&gt;0&lt;/sub&gt;</td>
<td>509.36</td>
<td>3.88</td>
<td>131.41</td>
<td>83.81</td>
<td>16.19</td>
</tr>
<tr>
<td></td>
<td>INTERCEPT1, γ&lt;sub&gt;00&lt;/sub&gt;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: SE = standard error*

### The Effects of Student-Level Variables

The first research question examined the effects of the selected student-level factors on science scores at the fourth grade in five countries. To answer this question, a random coefficients model was run with HLM. The student-level variables, gender, self-confidence, and home resources were introduced at the student-level (level-1). No variables were used at the classroom-level. The correlations among student-level variables were low (see Appendix A). The following model was used in examining the effects of student-level variables on the TIMSS 2003 science scores:

**Level-1 (Student-Level)**

\[
\text{SCIENCE PV}_{ij} = \beta_0 + \beta_1 \text{(GENDER}_i) + \beta_2 \text{(SELFCON}_i) + \beta_3 \text{(HOMERES}_i) + r_{ij}
\]

**Level-2 (Classroom-Level)**

\[
\beta_0 = \gamma_{00} + u_{0j}
\]

\[
\beta_1 = \gamma_{10}
\]

\[
\beta_2 = \gamma_{20}
\]

\[
\beta_3 = \gamma_{30}
\]
Table 4.4 The Effects of Student-Level Variables on Science Achievement Within Countries

<table>
<thead>
<tr>
<th>Type of Effect</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>Coefficient</td>
<td>SE</td>
<td>Coefficient</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>566.38**</td>
<td>5.71</td>
<td>544.34**</td>
<td>2.24</td>
<td>539.59**</td>
</tr>
<tr>
<td>Student-Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-6.93*</td>
<td>2.31</td>
<td>-0.69</td>
<td>2.41</td>
<td>-2.63</td>
</tr>
<tr>
<td>Self-Confidence</td>
<td>16.50***</td>
<td>1.36</td>
<td>20.39***</td>
<td>1.88</td>
<td>21.56***</td>
</tr>
<tr>
<td>Home Resources</td>
<td>17.11***</td>
<td>2.35</td>
<td>22.67***</td>
<td>3.37</td>
<td>18.51***</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td>Variance Component</td>
<td>SD</td>
<td>Variance Component</td>
<td>SD</td>
<td>Variance Component</td>
</tr>
<tr>
<td>Between Class, $u_0$</td>
<td>4047.07</td>
<td>63.62</td>
<td>200.73</td>
<td>14.17</td>
<td>1960.41</td>
</tr>
<tr>
<td>Within Class, $r$</td>
<td>3322.45</td>
<td>57.64</td>
<td>4767.83</td>
<td>69.05</td>
<td>4033.42</td>
</tr>
<tr>
<td>Level-1 Variance Explained (%)</td>
<td>7.00</td>
<td>5.75</td>
<td>6.17</td>
<td>7.50</td>
<td>4.33</td>
</tr>
</tbody>
</table>

*Note. SE = standard error; SD = standard deviation.*
*p < 0.05 level, **p < 0.01, ***p < 0.001.
The results of the Random Coefficients HLM analyses showed that there were no significant differences between the TIMSS 2003 fourth grade science scores of boys and girls in Japan, the U.S., and Australia. The difference was significant at p< 0.05 level in Singapore and Scotland. On average, girls scored approximately 7 points lower than boys in Singapore and 12 points lower in Scotland, controlling for the science self-confidence and home resources (see Table 4.4).

The science self-confidence (SELFCON) and home resources (HOMERES) positively affected the TIMSS 2003 science scores (see Table 4.4). The effects were significant at the p < 0.001 level in all five countries. On average, the increase in the science score associated with one point increase in science self-confidence were 16.50 in Singapore, 20.39 in Japan, 21.56 in the U.S., 24.44 in Australia, and 15.46 in Scotland, controlling for the gender and home resources. The increases in the science score associated with one point increase in home resources were 17.11, 22.67, 18.51, 23.07, and 15.91 respectively in these countries, controlling for the gender and science self-confidence. The cross-country comparisons of student-level effects were examined later in this chapter. The following section describes the effects of classroom-level variables on science achievement in each country.

The Effects of Classroom-Level Variables

The second research question examined the effects of the selected classroom-level factors on science scores at the fourth grade in five countries. To answer this question, a conditional model with all the student and classroom-level variables was run with HLM. The classroom-level variables, teacher’s experience, education, major, teacher support, science inquiry, minutes of science instruction, class size, class mean of home resources, and class mean of science self-confidence were introduced at the classroom-level (level-2). The correlations among classroom-level variables were presented in Appendix A. The following model was used in examining the effects of the classroom-level variables on the TIMSS 2003 science scores:

**Level-1 (Student-Level)**

\[ \text{SCIENCE PV}_{ij} = \beta_{0j} + \beta_{1j}(\text{GENDER}_i) + \beta_{2j}(\text{SELFCON}_i) + \beta_{3j}(\text{HOMERES}_i) + r_{ij} \]
Level-2 (Classroom-Level)

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{TEXP})_j + \gamma_{03}(\text{TEDUC})_j + \gamma_{03}(\text{TMAJOR})_j + \gamma_{04}(\text{TSUPPORT})_j + \gamma_{05}(\text{SCINQ})_j + \gamma_{06}(\text{SCIMIN})_j + \gamma_{07}(\text{CLASSIZE})_j + \gamma_{08}(\text{M_SELFCON})_j + \gamma_{09}(\text{M_HOMERES})_j + u_{0j} \]

\[ \beta_{1j} = \gamma_{10} \]

\[ \beta_{2j} = \gamma_{20} \]

\[ \beta_{3j} = \gamma_{30} \]

In general, there were no significant relations between the teacher’s education and major and the TIMSS 2003 science scores (see Table 4.5). There was a positive relation between teacher’s experience and students’ science scores only in Japan \((p < 0.05)\). Students achieved 0.40 points higher with one year increase in their teacher’s experience controlling for other variables. Teacher support was significantly related to science scores in Singapore and the U.S. \((p < 0.001)\). Students who perceived their teachers as more supportive tend to achieve higher scores on the science test compared to their counterparts in other classrooms. In Singapore, students scored 15 points \((46.47 \times 0.32)\) higher with one standard deviation increase in the perception of teacher support when all the other variables were held constant. In the U.S., there was an 11-point \((35.50 \times 0.31)\) increase in the science scores with one standard deviation increase in the perception of teacher support controlling for the other variables (see Table 4.5).

The effects of science inquiry \((\text{SCINQ})\) were inconsistent. In Singapore, classrooms that conducted more science inquiry activities scored significantly higher than did classrooms that conducted less \((p < 0.05)\) (see Table 4.9). In the U.S., however, the effect of science inquiry on science scores was negative and significant at \(p < 0.05\) level. In Singapore, there was a 7-point \((4.50 \times 1.58)\) increase in class mean achievement with one standard deviation increase in the emphasis on science inquiry, when other variables were held constant. In contrast, for U.S. fourth graders, there was a 7-point \((-3.74 \times 1.95)\) decrease in class mean achievement with every unit increase in the emphasis on science inquiry. In Australia, Japan and Scotland, there was no relation between science inquiry and the TIMSS 2003 fourth grade science scores. Minutes of weekly science instruction \((\text{SCIMIN})\) and class size \((\text{CLASSIZE})\) did not have significant effects on science scores in any of the five countries examined (see Table 4.5).
Table 4.5 *The Effects of Student- and Classroom-Level Variables on Science Achievement Within Countries*

<table>
<thead>
<tr>
<th>Type of Effect</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>SE</td>
<td>Coefficient</td>
<td>SE</td>
<td>Coefficient</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>564.63***</td>
<td>3.96</td>
<td>543.09***</td>
<td>2.14</td>
<td>533.49***</td>
</tr>
<tr>
<td><strong>Student-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender Slope</td>
<td>-6.97**</td>
<td>2.30</td>
<td>-0.84</td>
<td>2.42</td>
<td>-2.43</td>
</tr>
<tr>
<td>Self-Conf. Slope</td>
<td>16.47***</td>
<td>1.36</td>
<td>20.30***</td>
<td>1.86</td>
<td>21.57***</td>
</tr>
<tr>
<td>Home Res. Slope</td>
<td>17.12***</td>
<td>2.35</td>
<td>22.69***</td>
<td>3.37</td>
<td>18.50***</td>
</tr>
<tr>
<td><strong>Classroom-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Experience</td>
<td>0.31</td>
<td>0.26</td>
<td>0.40*</td>
<td>0.20</td>
<td>0.28</td>
</tr>
<tr>
<td>Teacher’s Education*</td>
<td>14.37</td>
<td>8.14</td>
<td>0.24</td>
<td>4.01</td>
<td>-2.75</td>
</tr>
<tr>
<td>Teacher’s Major</td>
<td>-1.78</td>
<td>5.54</td>
<td>3.03</td>
<td>4.27</td>
<td>1.55</td>
</tr>
<tr>
<td>Teacher Support</td>
<td>46.47***</td>
<td>11.65</td>
<td>6.51</td>
<td>4.28</td>
<td>35.50***</td>
</tr>
<tr>
<td>Science Inquiry</td>
<td>4.50**</td>
<td>1.82</td>
<td>1.18</td>
<td>1.49</td>
<td>-3.74**</td>
</tr>
<tr>
<td>Min. of Sci. Inst.</td>
<td>0.56</td>
<td>0.31</td>
<td>0.06</td>
<td>0.07</td>
<td>0.03</td>
</tr>
<tr>
<td>Class Size</td>
<td>18.84</td>
<td>9.76</td>
<td>-2.78</td>
<td>1.84</td>
<td>-3.18</td>
</tr>
<tr>
<td>Class Mean Self-Con.</td>
<td>15.66</td>
<td>16.02</td>
<td>11.45</td>
<td>10.37</td>
<td>38.73*</td>
</tr>
<tr>
<td>Class Mean Home Res.</td>
<td>289.24***</td>
<td>21.74</td>
<td>37.39*</td>
<td>17.30</td>
<td>122.85***</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td>Variance Component</td>
<td>SD</td>
<td>Variance Component</td>
<td>SD</td>
<td>Variance Component</td>
</tr>
<tr>
<td>Between Class, u₀</td>
<td>1135.61</td>
<td>33.87</td>
<td>132.65</td>
<td>11.52</td>
<td>729.26</td>
</tr>
<tr>
<td>Within Class, r</td>
<td>3322.99</td>
<td>57.64</td>
<td>4771.91</td>
<td>69.08</td>
<td>4044.72</td>
</tr>
<tr>
<td>Level-2 Variance Explained (%)</td>
<td>71.71</td>
<td>30.35</td>
<td>62.56</td>
<td>42.53</td>
<td>57.19</td>
</tr>
</tbody>
</table>

*Note. SE = standard error; SD = standard deviation.*

a. Teacher’s Education for Singapore was not included in the final HLM model because there was only one teacher with a master’s or doctoral degree.

* p < 0.05 level, **p < 0.01, *** p < 0.001.
The effect of class mean self-confidence at the classroom-level was significant only in the U.S. \( (p < 0.05) \) and Scotland \( (p < 0.01) \). On average, students achieved 9 points \( (38.73*0.23) \) higher in the U.S. and 8 points \( (38.80*0.21) \) higher in Scotland on the TIMSS 2003 science test with one standard deviation increase in the class mean self-confidence, controlling for other variables (see Table 4.5). The effect of class mean home resources \( (M_{\text{HOMERES}}) \) was positive and significant in four out of five countries at \( p < 0.001 \) level. In Singapore, classrooms with one standard deviation increase in class mean home resources achieved, on average, 46 \( (289.24*0.16) \) points higher, controlling for other variables. The increase was 23 points \( (122.85*0.19) \) in the U.S, 26 points \( (112.03*0.23) \) in Australia, and 18 points \( (96.50*0.19) \) in Scotland. In Japan, the effect of class mean home resources was not significant. The cross-country comparisons of the classroom-level effects were examined later in this chapter.

According to the interpretation of Cohen’s \( d \), the effect size of gender was trivial in Singapore \( (d= 0.12) \) and Scotland \( (d= 0.17) \). The effect sizes of student-level science self-confidence and home resources were small \( (0.2 < d < 0.5) \) in all five countries. At the classroom-level, teacher support had medium effect sizes in Singapore \( (d= 0.77) \) and the U.S. \( (d= 0.54) \). Mean science self-confidence had also medium effect sizes in the U.S. \( (d= 0.59) \) and Scotland \( (d= 0.55) \). The effect size of science inquiry was trivial \( (d < 0.2) \) in both Singapore and the U.S. Finally, the effect sizes of class mean home resources were high in Singapore \( (d= 4.82) \), the U.S. \( (d= 1.86) \), Australia \( (d= 1.56) \), and Scotland \( (d= 1.38) \) and medium in Japan \( (d= 0.53) \) (Cohen, 1988). Student-level standard deviations from the unconditional models were used in effect size estimations.

Finally, to examine students’ scores by different levels of home resources at the student-level, the average science plausible values for three levels of home resources (Low, Medium, High) were computed for each country (see Table 4.6). Japanese students with low levels of home resources scored higher than the students with low levels of home resources in other countries. Singaporean students with low levels of home resources scored lower than their counterparts in Japan, the U.S., and Scotland. The gap between the students with low levels of home resources and the students with high levels of home resources was 145 points in Singapore, higher than the gap in the other countries. The following section describes the explained variances in science achievement by student- and classroom-level variables in each country.
### Table 4.6 Average TIMSS 2003 Science Scores by Student-Level Home Resources

<table>
<thead>
<tr>
<th>Home Resources</th>
<th>Singapore M</th>
<th>SD</th>
<th>Japan M</th>
<th>SD</th>
<th>U.S. M</th>
<th>SD</th>
<th>Australia M</th>
<th>SD</th>
<th>Scotland M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>458.55</td>
<td>89.76</td>
<td>508.28</td>
<td>72.85</td>
<td>467.41</td>
<td>66.94</td>
<td>436.70</td>
<td>68.96</td>
<td>474.04</td>
<td>70.71</td>
</tr>
<tr>
<td>Medium</td>
<td>554.53</td>
<td>78.41</td>
<td>539.97</td>
<td>67.74</td>
<td>521.73</td>
<td>73.68</td>
<td>512.24</td>
<td>74.69</td>
<td>501.71</td>
<td>71.27</td>
</tr>
<tr>
<td>High</td>
<td>603.23</td>
<td>69.49</td>
<td>566.04</td>
<td>64.92</td>
<td>563.81</td>
<td>71.73</td>
<td>546.94</td>
<td>68.46</td>
<td>529.76</td>
<td>71.73</td>
</tr>
<tr>
<td>Average</td>
<td><strong>564.57</strong></td>
<td><strong>80.34</strong></td>
<td><strong>544.05</strong></td>
<td><strong>68.15</strong></td>
<td><strong>530.50</strong></td>
<td><strong>75.93</strong></td>
<td><strong>525.73</strong></td>
<td><strong>74.62</strong></td>
<td><strong>510.02</strong></td>
<td><strong>72.11</strong></td>
</tr>
</tbody>
</table>

*Note. M = mean; SD = standard deviation*

#### Explained Variances in Science Achievement

The third research question examined the explained variances in students’ science achievement by the selected student- and classroom-level factors in five countries. For this question, the reductions in the random error variances at the level-1 and level-2 were computed. The reduction in the random error variance at the level-1 was computed as below:

\[
\text{Reduction} = \frac{\sigma^2(\text{Unconditional}) - \sigma^2(\text{Model2})}{\sigma^2(\text{Unconditional})} \times 100
\]

In the U.S.,

\[
\text{Reduction} = \frac{4298.83 - 4033.41}{4298.83} \times 100 = 6.17 \%
\]

Introducing the selected student-level variables into the model reduced 6.17 percent of the level-1 error variance in the U.S. In other words, about 6 percent of the student-level variance in the U.S. was explained by science self-confidence and home resources both of which had significant effects on science achievement. The remaining 94 percent however, is accounted for by other variables not included in the model. Note that 69 percent of the variance in science scores was within classes in the U.S. In other words, student-level variables explained only 4 percent (6 x 69%) of the total variance. Similarly, explained variances at the level-1 for the four other countries in the study were 7 percent in Singapore, 5.75 percent in Japan, 7.5 percent in Australia, and 4.33 percent in Scotland (see Table 4.4).

The reduction in the random error variance at the level-2 (classroom-level) was computed:
In the U.S.,

\[
\text{Reduction} = \frac{\tau_{00}(\text{Unconditional}) - \tau_{00}(\text{Conditional})}{\tau_{00}(\text{Unconditional})} \times 100
\]

Introducing the selected classroom-level variables into the model reduced about 63 percent of the level-2 error variance. In other words, about 63 percent of the classroom-level variance was explained by teacher support, science inquiry, class mean home resources, and class mean of science self-confidence which had significant effects on science achievement. The remaining 37 percent however, is accounted for by other variables not included in the model. Explained variances at level-2 for other countries are listed in Table 4.7. Note that 31 percent of the variance in science scores was between classes in the U.S. Therefore, the classroom variables in the study explained the total of 19 percent \((63 \times 31\%)\) of the variance in the TIMSS 2003 science scores in the U.S.

Table 4.7 Explained Variances in Science Achievement Within Countries

<table>
<thead>
<tr>
<th>Explained Variance (%)</th>
<th>Singapore</th>
<th>Japan</th>
<th>U.S.</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student-Level</td>
<td>3.3</td>
<td>5.5</td>
<td>4.0</td>
<td>5.8</td>
<td>3.6</td>
</tr>
<tr>
<td>Classroom-Level</td>
<td>37.9</td>
<td>1.1</td>
<td>19.0</td>
<td>10.0</td>
<td>9.3</td>
</tr>
<tr>
<td>Total</td>
<td>41.2</td>
<td>6.6</td>
<td>23.0</td>
<td>15.8</td>
<td>12.9</td>
</tr>
</tbody>
</table>

Table 4.7 displays the explained variances in science achievement in each country at the student- and classroom-level, and as total. Overall, the selected student- and classroom-level variables explained as much as 41.2 percent of the variance in fourth graders’ science achievement in Singapore and as little as 6.6 percent in Japan.
Interactions

Several models were run to investigate possible interaction effects at the student- and classroom-level for each country. Only two interactions were significant and those were the interaction between home resources at the level-1 and class mean science self-confidence at the level-2 in the U.S. and the interaction between home resources at the level-1 and class mean home resources at the level-2 in Singapore (see Figure 4.1 and Figure 4.2).

![Figure 4.1 Home Resources by Class Mean Self-Confidence Interaction Predicting Science Achievement in the U.S.](image)

As seen in the Figure 4.1, for the U.S. fourth graders, the effect of home resources depended on the classroom mean of students’ self-confidence in science. On average, the more educational resources a student has at home the higher his/her score on the TIMSS 2003 assessment was. The interaction effect indicated that the effect of home resources was the greatest when a student was in a classroom with higher mean science self-confidence (i.e., 75th percentile of the sample) and the smallest when (s)he was in a classroom with lower mean science self-confidence (i.e., 25th percentile of the sample).
As seen in the Figure 4.2, for the Singaporean fourth graders, the effect of home resources depended on the classroom mean of home resources. On average, the more educational resources a student has at home the higher his/her score on the TIMSS 2003 assessment was. The interaction effect indicated that the effect of home resources was the greatest when a student was in a classroom with lower mean home resources and the smallest when (s)he was in a classroom with higher mean resources.

**Figure 4.2** Home Resources by Class Mean Home Resources Interaction Predicting Science Achievement in Singapore

**Cross-Country Analyses**

Combined data from all five countries were analyzed in a similar fashion, by using 2-level HLM to make cross-country comparisons. The same three variables are used at the student-level (gender, self-confidence, home resources). At the classroom-level, the variables with some significant effects, which were teacher support, science inquiry, class mean of science self-confidence and home resources, were compared. At the student- and classroom-level, country specific, dummy coded variables were created. Data from five countries included 24,233 students at the level-1 and 913 teachers at the level-2. Table 4.8 shows the cross-country effects of student-level variables. The results
of the combined analyses were slightly different from the country specific analyses (see Table 4.5).

**Table 4.8 Student-Level Effects Across Countries**

<table>
<thead>
<tr>
<th>Type of Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>T-ratio</th>
<th>Effect Size (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singapore</td>
<td>572.78***</td>
<td>5.34</td>
<td>107.35</td>
<td>-</td>
</tr>
<tr>
<td>Japan</td>
<td>547.68***</td>
<td>2.22</td>
<td>246.50</td>
<td>-</td>
</tr>
<tr>
<td>U.S.</td>
<td>535.18***</td>
<td>2.62</td>
<td>204.65</td>
<td>-</td>
</tr>
<tr>
<td>Australia</td>
<td>515.35***</td>
<td>4.64</td>
<td>111.15</td>
<td>-</td>
</tr>
<tr>
<td>Scotland</td>
<td>512.70***</td>
<td>3.99</td>
<td>128.53</td>
<td>-</td>
</tr>
<tr>
<td><strong>Student-Level Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sing – Gender</td>
<td>0.30</td>
<td>3.63</td>
<td>0.08</td>
<td>-</td>
</tr>
<tr>
<td>Japan – Gender</td>
<td>-0.85</td>
<td>2.38</td>
<td>-0.36</td>
<td>-</td>
</tr>
<tr>
<td>US – Gender</td>
<td>-4.86**</td>
<td>1.82</td>
<td>-2.68</td>
<td>0.06</td>
</tr>
<tr>
<td>Aust – Gender</td>
<td>-1.17</td>
<td>3.69</td>
<td>-0.32</td>
<td>-</td>
</tr>
<tr>
<td>Scot – Gender</td>
<td>-13.67**</td>
<td>4.20</td>
<td>-3.25</td>
<td>0.18</td>
</tr>
<tr>
<td>Sing – Self-Con.</td>
<td>17.08***</td>
<td>2.03</td>
<td>8.41</td>
<td>0.22</td>
</tr>
<tr>
<td>Japan – Self-Con.</td>
<td>19.85***</td>
<td>1.83</td>
<td>10.83</td>
<td>0.25</td>
</tr>
<tr>
<td>US – Self-Con.</td>
<td>26.44***</td>
<td>1.84</td>
<td>14.39</td>
<td>0.34</td>
</tr>
<tr>
<td>Aust – Self-Con.</td>
<td>24.50***</td>
<td>2.99</td>
<td>8.19</td>
<td>0.31</td>
</tr>
<tr>
<td>Scot – Self-Con.</td>
<td>17.03***</td>
<td>2.61</td>
<td>6.51</td>
<td>0.22</td>
</tr>
<tr>
<td>Sing – Home Res.</td>
<td>50.91***</td>
<td>4.21</td>
<td>12.09</td>
<td>0.65</td>
</tr>
<tr>
<td>Japan – Home Res.</td>
<td>24.13***</td>
<td>3.45</td>
<td>6.99</td>
<td>0.31</td>
</tr>
<tr>
<td>US – Home Res.</td>
<td>39.48***</td>
<td>2.52</td>
<td>15.68</td>
<td>0.51</td>
</tr>
<tr>
<td>Aust – Home Res.</td>
<td>26.29***</td>
<td>3.17</td>
<td>8.30</td>
<td>0.34</td>
</tr>
<tr>
<td>Scot – Home Res.</td>
<td>28.48***</td>
<td>4.44</td>
<td>6.41</td>
<td>0.37</td>
</tr>
<tr>
<td><strong>Random Effect</strong></td>
<td>Variance Component</td>
<td>SD</td>
<td>Chi-square</td>
<td>p-value</td>
</tr>
<tr>
<td>Within Class, r</td>
<td>5546.57</td>
<td>74.48</td>
<td>1014.44</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Level-1 Variance Explained (%)</td>
<td></td>
<td>5.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SE= standard error. *p < 0.05 level, **p < 0.01, *** p < 0.001.

In terms of gender, girls scored significantly lower than boys on the TIMSS 2003 science test in Scotland and the U.S. compared to other countries (see Table 4.8). In Singapore, Japan, and Australia, the differences between boys’ and girls’ scores were negligible. In the country specific analyses the gender effect was significant in Singapore.
and Scotland but not in the U.S. (see Table 4.5). Similar to country specific analyses, at the student-level, the effects of science self-confidence and home resources were positive and significant in all five countries (p< 0.001) (see Table 4.8). The three variables explained about 5 percent of the variance in the TIMSS 2003 fourth grade science scores in five countries at the student-level.

The effect size of gender was trivial (0.0 < d < 0.2) in Scotland and the U.S. The effect size of student-level science self-confidence was small (0.2 < d < 0.5), for one standard deviation, in all five countries. However, for two standard deviations, the effect size was small to moderate (0.5 < d < 0.8). For the student-level home resources, the effect sizes ranged from small to moderate for one standard deviation, and from moderate to high (0.8 ≤ d) for two standard deviations (Cohen, 1988).

Table 4.9 summarizes the effects of classroom-level variables across countries. Teacher support was positively related to students’ science scores in Singapore and the U.S compared to other countries; the effect was significant at p < 0.05 level. This result was consistent with the results of the country specific analyses. Science inquiry was negatively related to science scores in Australia and the U.S. and positively related in Singapore. In the country specific analyses, the effect of science inquiry on the TIMSS 2003 science scores was not significant in Australia (see Table 4.5).

The cross-country analysis revealed that the effect of class mean science self-confidence on the TIMSS 2003 scores was significant only in Australia compared to other countries. In the country specific analyses, the effect was significant in Scotland and the U.S. but not in Australia. Class mean of home resources was positively related to student scores (p< 0.001) in four countries. There was no significant relation between classroom-level home resources and the TIMSS 2003 science score in Japan compared to other countries. However, the country-specific analyses revealed a significant positive relation in Japan (see Table 4.5). The four variables, teacher support, science inquiry, class mean home resources, and class mean self-confidence explained approximately 60 percent of the variance in fourth grade science scores in five countries at the classroom-level. The student- and classroom-level variables explained approximately 22 percent of the total variance in science achievement in five countries.
Table 4.9 Classroom-Level Effects Across Countries

<table>
<thead>
<tr>
<th>Type of Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>T-ratio</th>
<th>Effect Size(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classroom-Level Fixed Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sing – Teacher Support</td>
<td>53.02***</td>
<td>13.19</td>
<td>4.02</td>
<td>0.80</td>
</tr>
<tr>
<td>Japan – Teacher Support</td>
<td>5.31</td>
<td>6.66</td>
<td>0.80</td>
<td>-</td>
</tr>
<tr>
<td>US – Teacher Support</td>
<td>36.07***</td>
<td>6.77</td>
<td>5.33</td>
<td>0.53</td>
</tr>
<tr>
<td>Aust – Teacher Support</td>
<td>13.57</td>
<td>8.71</td>
<td>1.56</td>
<td>-</td>
</tr>
<tr>
<td>Scot – Teacher Support</td>
<td>-10.59</td>
<td>10.07</td>
<td>-1.05</td>
<td>-</td>
</tr>
<tr>
<td>Sing – Science Inquiry</td>
<td>4.71*</td>
<td>2.25</td>
<td>2.10</td>
<td>0.07</td>
</tr>
<tr>
<td>Japan – Science Inquiry</td>
<td>0.75</td>
<td>1.93</td>
<td>0.39</td>
<td>-</td>
</tr>
<tr>
<td>US – Science Inquiry</td>
<td>-3.69*</td>
<td>1.29</td>
<td>-2.87</td>
<td>0.06</td>
</tr>
<tr>
<td>Aust – Science Inquiry</td>
<td>-2.73*</td>
<td>1.17</td>
<td>-2.34</td>
<td>0.04</td>
</tr>
<tr>
<td>Scot – Science Inquiry</td>
<td>-1.24</td>
<td>1.40</td>
<td>-0.89</td>
<td>-</td>
</tr>
<tr>
<td>Sing – Class Mean Self-Con.</td>
<td>-1.12</td>
<td>18.48</td>
<td>-0.06</td>
<td>-</td>
</tr>
<tr>
<td>Japan – Class Mean Self-Con.</td>
<td>-18.73</td>
<td>11.56</td>
<td>-1.62</td>
<td>-</td>
</tr>
<tr>
<td>US – Class Mean Self-Con.</td>
<td>19.83</td>
<td>15.82</td>
<td>1.25</td>
<td>-</td>
</tr>
<tr>
<td>Aust – Class Mean Self-Con.</td>
<td>27.38*</td>
<td>12.85</td>
<td>2.13</td>
<td>0.41</td>
</tr>
<tr>
<td>Scot – Class Mean Self-Con.</td>
<td>20.28</td>
<td>13.09</td>
<td>1.55</td>
<td>-</td>
</tr>
<tr>
<td>Sing – Class Mean Home Res.</td>
<td>287.97***</td>
<td>22.29</td>
<td>12.92</td>
<td>4.27</td>
</tr>
<tr>
<td>Japan – Class Mean Home Res.</td>
<td>17.00</td>
<td>20.72</td>
<td>0.82</td>
<td>-</td>
</tr>
<tr>
<td>US – Class Mean Home Res.</td>
<td>109.62***</td>
<td>16.24</td>
<td>6.75</td>
<td>1.62</td>
</tr>
<tr>
<td>Aust – Class Mean Home Res.</td>
<td>71.26***</td>
<td>14.50</td>
<td>4.92</td>
<td>1.06</td>
</tr>
<tr>
<td>Scot – Class Mean Home Res.</td>
<td>79.73***</td>
<td>18.04</td>
<td>4.42</td>
<td>1.18</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance</th>
<th>SD</th>
<th>Chi-square</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Class, u0</td>
<td>749.77</td>
<td>27.38</td>
<td>5212.11</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>Level-2 Variance Explained (%)</td>
<td>60.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Variance Explained (%)</td>
<td>21.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. SE = standard error.
*p < 0.05 level, **p < 0.01, *** p < 0.001.

As seen in Table 4.9, the effect size of teacher support was high in Singapore (0.8 ≤ d) and moderate in the U.S. (d = 0.53). Science inquiry had trivial effect sizes in Singapore, the U.S., and Australia. Classroom-level science self-confidence had a small effect size in Australia (d = 0.41). Finally, the effect size of classroom-level home resources was high in all four countries, Singapore, the U.S., Australia, and Scotland.
Hypothesis Testing of the Student- and Classroom-Level Effects

To compare the effects of the student-level and classroom-level variables in five countries, hypothesis testing was applied in HLM analyses. Accordingly, the effects of the significant variables in both levels in the U.S. were compared with the effects in other countries. Hypothesis testing was not conducted for variables that were not significant. The hypothesis testing on the effects of gender, science self-confidence, home resources, class mean of home resources, teacher support and the science inquiry were applied for the U.S. with respect to other countries. The results are displayed in Table 4.10.

<table>
<thead>
<tr>
<th>Hypothesis Testing</th>
<th>Coefficients</th>
<th>Chi-square</th>
<th>df</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effect of Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Scotland</td>
<td>-4.86 – (-13.67)</td>
<td>26.83</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>Effect of Self-Confidence</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Singapore</td>
<td>26.44 – 17.08</td>
<td>476.68</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Japan</td>
<td>26.44 – 19.85</td>
<td>385.02</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Australia</td>
<td>26.44 – 24.50</td>
<td>249.37</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Scotland</td>
<td>26.44 – 17.03</td>
<td>323.79</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>Effect of Home Resources</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Singapore</td>
<td>39.48 – 50.91</td>
<td>731.69</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Japan</td>
<td>39.48 – 24.13</td>
<td>333.85</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Australia</td>
<td>39.48 – 26.29</td>
<td>399.51</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Scotland</td>
<td>39.48 – 28.48</td>
<td>437.90</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>Classroom-Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Effect of Teacher Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Singapore</td>
<td>36.07 – 53.02</td>
<td>90.18</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>Effect of Science Inquiry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Australia</td>
<td>-3.69 – (-2.73)</td>
<td>17.80</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td><strong>Effect of Class Mean Home Resources</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U.S. vs. Singapore</td>
<td>109.62 – 287.97</td>
<td>376.51</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Australia</td>
<td>109.62 – 71.26</td>
<td>100.64</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
<tr>
<td>U.S. vs. Scotland</td>
<td>109.62 – 79.73</td>
<td>120.98</td>
<td>2</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

*Note.* df = degrees of freedom.

The results showed that, at the student-level, the effect of gender was larger in Scotland than in the U.S. (see also Table 4.8). In other words, the difference between the scores of boys and girls in Scotland on the TIMSS 2003 science test was larger than the
difference in the U.S. The effect of science self-confidence was larger in the U.S. compared to other countries. The effect of home resources was larger in the U.S. compared to Japan, Australia, and Scotland, but smaller compared to Singapore.

In terms of classroom-level effects, the effect of teacher support in the U.S. was smaller than the effect in Singapore. In other words, students in Singapore had higher scores when their teachers were more supportive compared to students in the U.S. The effect of science inquiry was larger in the U.S. compared to Australia. Given that the effects were negative in these countries; this result indicates students received lower scores in the U.S. than in Australia when science inquiry activities were emphasized. Finally, the effect of class mean of home resources on science achievement was larger in the U.S. compared to Australia and Scotland; smaller compared to Singapore (see also Table 4.12). The next chapter presents the summary of findings, discussions and recommendations for practice, policy, and future studies.
CHAPTER V

DISCUSSION

The results of the Trends in International Math and Science Study (TIMSS) 2003 assessment showed that the gap between the U.S. students and the students from other developed countries has increased since the 1995 assessment. This study was an attempt to examine possible similarities and differences between the U.S. and other countries in terms of select student- and classroom-level effects on elementary science achievement. Guided by previous school learning models, the current study conceptualized a model that presents two clusters of factors affecting student achievement (see Figure 1.1). Accordingly, student and home background factors were clustered under student-level factors while teacher, instructional, and compositional factors were clustered under classroom-level factors (see Table 5.1).

The TIMSS 2003 data for the U.S., in addition to four other countries, Singapore, Japan, Australia, and Scotland were utilized. Singapore and Japan were selected primarily because of their high performance on TIMSS assessments; Australia and Scotland were selected as English speaking countries that perform significantly lower than the U.S. and their decentralized education systems are similar to the U.S. In each country, two-level Hierarchical Linear Modeling (HLM) analyses were performed with student-level factors at level-1 and classroom-level factors at level-2. The outcome variable was the five plausible values that students achieved on the TIMSS 2003 science test. The data from five countries were combined to conduct the cross-country analyses. Using multilevel analysis was appropriate since students were nested within classrooms. The TIMSS 2003 fourth grade student and teacher background questionnaires were utilized to define variables. Explained variances in fourth grade science achievement within each country and across five countries were reported.
Summary

Table 5.1 summarizes the results of the country specific analyses.

Table 5.1 Significant Variables at the Student- and Classroom-Level Within Countries

<table>
<thead>
<tr>
<th>Variables</th>
<th>Singapore</th>
<th>Japan</th>
<th>United States</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student-Level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>**</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Self-Confidence</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Home Resources</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Level-1 Variance Explained (%)</td>
<td>7.0</td>
<td>6.0</td>
<td>6.3</td>
<td>7.3</td>
<td>4.3</td>
</tr>
<tr>
<td><strong>Classroom-Level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Education</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher’s Major</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Support</td>
<td>***</td>
<td></td>
<td></td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Science Inquiry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min of Sci Instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class Size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class Mean Self-Con</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>Class Mean Home Res</td>
<td>***</td>
<td>*</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Level-2 Variance Explained (%)</td>
<td>72</td>
<td>30</td>
<td>63</td>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td>**Total Variance Explained (%)</td>
<td><strong>41</strong></td>
<td>7</td>
<td><strong>23</strong></td>
<td><strong>16</strong></td>
<td><strong>13</strong></td>
</tr>
</tbody>
</table>

*p < 0.05 level, **p < 0.01 level, ***p < 0.001 level

The first research question of the study investigated the effects of student-level factors on science achievement in each country. As seen in Table 5.1, in each country, students with more resources at home and higher science self-confidence tended to have higher science scores at the fourth grade level. The gender difference was negligible in general, although, boys did outperformed girls in Singapore and Scotland. Similar results were found in the cross-country analysis. In this case, the countries were the U.S. and Scotland and the gender difference still favored boys (see Table 5.2).
Table 5.2 Significant Variables at the Student- and Classroom-Level Across Countries

<table>
<thead>
<tr>
<th>Variables</th>
<th>Singapore</th>
<th>Japan</th>
<th>United States</th>
<th>Australia</th>
<th>Scotland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student-Level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td>**</td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>Self-Confidence</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Home Resources</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td><strong>Classroom-Level Variables</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher Support</td>
<td>***</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science Inquiry</td>
<td>**</td>
<td>**</td>
<td>**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class Mean Self-Con</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class Mean Home Res</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
</tbody>
</table>

*p < 0.05 level, **p < 0.01 level, ***p < 0.001 level

The second research question examined the effects of classroom-level factors on the elementary science achievement. In terms of the effects of teacher and instructional variables on science achievement there were no clear patterns among countries. In general, students in the classrooms with higher levels of home resources tended to have higher science scores. It seems evident that, overall, other classroom-level factors were minimally related to science achievement. Classrooms that perceive the teachers to be more supportive tended to have higher science scores in Singapore and the U.S. (see Table 5.1 and Table 5.2). Finally, in the U.S. and Australia, classrooms whose teachers placed more emphasis on science inquiry activities scored significantly lower than their counterparts who were exposed to less science inquiry (see Table 5.2).

The third research question examined the explained variances in students’ science achievement by student- and classroom-level factors in five countries. Especially at the student-level, the three variables explained little variance in science achievement in the countries investigated. On the other hand, classroom-level variables explained slightly higher percentages of the variances compared to the student-level variables. Overall, the selected student- and classroom-level variables explained as much as 41 percent of the variance in fourth graders’ science achievement in Singapore and as little as seven percent in Japan (see Table 5.1). Finally, through cross-country analyses it was found that approximately 22 percent of the variance in student scores in five countries was explained by student- and classroom-level variables (see Table 5.2). Clearly, a large
portion of the variance in TIMSS 2003 4th grade science scores could not be explained by the variables examined in the study.

**Discussion**

Overall, the results of this study showed that selected student background characteristics were consistently related to elementary science achievement in countries investigated. At the student-level, higher levels of home resources and self-confidence and at the classroom-level, higher levels of class mean home resources yielded higher science achievement. In general, teacher and instructional variables were minimally related to science achievement. There was evidence of positive effects of teacher support on science achievement in the U.S. and Singapore. Science inquiry was positively related to science achievement in Singapore but negatively related in the U.S. and Australia. For all the countries investigated, with the exception of Singapore, the between-class variance was much smaller than the within-class variance. Japan had the smallest variation in science achievement among classrooms which indicates the homogeneity across classrooms in Japan.

The findings of this study did not support the hypothesized conceptual model in Chapter 1. The hypothesized model suggested that student-level factors, including gender, science self-confidence, and home resources, and classroom-level factors including teacher variables, instructional variables, and classroom composition would predict science achievement at fourth grade level. The results suggested that at the student-level, science self-confidence and home resources, and at the classroom-level, class mean home resources were consistent predictors of science achievement in all five countries investigated. As for the teacher and instructional variables, there were inconsistencies across countries. This finding could be due to the different levels of between-class variances in countries. In countries with smaller between-class variance, such as Japan (four percent) it was less likely to observe significant effects at the classroom-level due to the homogeneity across classrooms. Further testing of the role of these factors included in the conceptual model is suggested.

The results of this study were encouraging in terms of the small gender effects on science achievement. These results were consistent with the results of previous studies (see Chapin, 2007; Greenfield, 1996). Although, in general, boys outperformed girls, the
differences were significant only in the U.S. and Scotland compared to other countries in the study. Research indicates, though, the achievement gap between boys and girls widens as children move to middle school (Greenfield, 1995, 1996; Kahle, 2004). A recent TIMSS 2003 report (Martin et al., 2004) indicated that at fourth grade, boys outperformed girls in only six out of 25 countries whereas, at the eighth grade, boys scored higher than girls in 33 out of 49 countries.

In the current study, at the student-level, students’ self-confidence in science positively affected fourth grade science achievement, as established by prior research (see Andrew, 1998; Britner & Pajares, 2001, 2006; Gwilliam & Betz, 2001; House; 2004, 2008; Lau & Roeser, 2002; Luzzo et al. 1999). Understanding students’ self-beliefs early in schooling is important because, if these beliefs are carried through middle and high school, they could influence future academic achievement. Of particular concern to science educators and parents is that low self-confidence in science and mathematics can block the pursuit of careers in mathematics and science (Gwilliam & Betz, 2001; Luzzo et al., 1999).

Since parents’ educational level or SES was not available in the fourth grade TIMSS 2003 questionnaires, this study utilized home resources (books, computer, and study desk) as a student background variable. The variable of home resources was an important predictor of fourth grade science achievement at the student-level. Based on the results of hypothesis testing, the availability of educational resources at home seemed more important for science achievement of Singaporean students in comparison to other countries in the study.

The findings of the current study suggested that, in general, teachers’ background characteristics were not related to students’ science achievement. These results were aligned with previous studies investigating the teacher effects on student achievement (see Goldhaber & Brewer, 2000; Hanushek, 1997; Hill et al. 2005; Wayne & Young, 2003; Xin et al. 2004). It is worthwhile to mention, however, that positive influences of teacher support on fourth grade science achievement were found in Singapore and the U.S. The students of more supportive teachers tended to score higher than the students of less supportive teachers on the TIMSS 2003 science test in these countries. Note that between-class variances in Singapore (53 percent) and in the U.S. (31 percent) were higher than the other countries. Therefore, the likelihood of finding classroom variables
that predicted student outcomes was greater in Singapore and the U.S. compared to Japan, Australia, and Scotland.

In terms of instructional factors, minutes of weekly science instruction and class size were not related to fourth grade science achievement in any of the countries investigated. These results were consistent with the results of previous studies which investigated the effects of class-size (Pong & Pallas, 2001) and amount of instruction (Baker et al., 2004) on student achievement internationally. Emphasis on science inquiry had a positive effect on the TIMSS 2003 science scores in Singapore but negative effects in the U.S. and Australia, although the effect sizes were small.

This study found that the composition of classrooms measured by class mean of home resources is strongly predictive of science achievement at fourth grade level. The results were similar to those of Lamb and Fullarton (2002) which revealed that in Australia and the U.S., the TIMSS science achievement was highest in classrooms with higher concentrations of students from families with more books and higher levels of parental education. These results underscore the importance of controlling for student characteristics when studying classroom or teacher effects.

A large portion of the variance in science achievement at the student-level could not be explained for the countries investigated. On the other hand, relative to student level, explained variance between classrooms was larger. It is evident that the selected student-level factors, gender, science self-confidence, and home resources measured by the TIMSS 2003 questionnaire were not adequate in explaining the individual differences in science achievement. It should be kept in mind that most of the variation in achievement was among students within classrooms (i.e., 96 percent in Japan), which means that characteristics among individual students within a classroom were more diverse than the characteristics across classrooms. There could be various other factors that affect science achievement at the individual level that were not included in this study. For example, research findings suggest that parents’ education level (Lamb & Fullarton, 2002; Xin et al., 2004) and parental involvement in education (Aun et al., 2006; Bray, 1999; Khong & Ng, 2005; Russell, 1997) have positive influences on student outcomes.

The findings indicated some distinctive characteristics of countries. For example, in Japan the between-class variance was only four percent, reflecting very little difference
from one classroom to another perhaps a result of its uniform schooling system. Singapore, on the other hand, had the largest between-class variance (53 percent) among five countries. Most of the variance among classrooms in Singapore was explained by class mean home resources which indicates possible stratification of the schooling system in this country. The achievement gap between the students with low levels of home resources and high levels of home resources was also the largest (145 points) in Singapore. In Japan the difference was 58 points on the TIMSS scale (see Table 4.6).

As Husen and Postlethwaite (1994) suggested, in Japan which has a highly centralized education system, there was homogeneity across classrooms. In their analysis of the TIMSS Case Study Project, LeTendre and colleagues (2003) found remarkable differences between Japan and the U.S. in terms of curricular differentiation and ability grouping. In Japan, there was little or no differentiation in curriculum and no ability grouping until high school. Elementary and middle school curricula were uniform from class to class and school to school. In the U.S. curricular and organizational differentiation within classrooms and schools was widespread. It begins in early grades and persists through the end of high school. The authors pointed out that most Japanese believe all children should have a very similar education until high school. Given the emphasis on the high school entrance exam, Japanese parents prefer minimum differentiation in their children’s education (LeTendre et al.).

Despite its centralized education system, Singapore was much different from Japan with a large variation among classrooms. A possible explanation of Singapore’s heterogeneity across classrooms could be merit based distribution of scholarships (Mukhopadhaya, 2002) and tracking systems (Tan, 1998). In a case study analysis of income disparity, Mukhopadhaya (2002) pointed out that despite the allocation a substantial budget to education in Singapore, some educational policies are biased toward middle and high income classes. For example, merit scholarships and bursaries are aimed at the brightest students, and in most cases, the brightest students come from better educational environments with more resources. Therefore, the rich and middle class families are more likely to benefit from scholarships.

Tan (1998) speculated that the highly competitive schooling system in Singapore has led schools to become academically selective. There is a growing stratification of schools with admissions based on student ability and students from wealthier families are
usually overrepresented in well-established, prestigious schools. Moreover, Bracey (1998) stated that Singapore admits Malaysian students of high ability into its schools and many Singaporean families whose children are not succeeding in the local schools send them to schools in Malaysia. Bracey added long-term guest workers from the Philippines and Indonesia are forbidden to bring their families to Singapore. There is room for investigation of whether these factors are associated with Singapore’s superior performance in science.

In terms of explaining the achievement gap between the U.S. and other countries, the findings did not offer a lot answers. However, the U.S had the second largest between-class variance, a large portion of which was explained by class mean home resources. Home resources had significant impacts on students’ learning both at the student-and classroom-level. This finding draws attention to distribution of educational resources among schools in the U.S. In terms of the other classroom effects, emphasis on science inquiry had a negative effect on science achievement which is worth further investigation. Finally, at the student-level, self-confidence in science seemed to be the most influential on U.S. students’ achievement compared to other countries.

There could be other factors to investigate in explaining the variability in TIMSS 2003 science achievement between the U.S. and other countries. For example, Schmidt and McKnight (1998) indicated that differences in curriculum across countries might be playing an important role in different achievement levels. While comparing the U.S. and Asian education systems the TIMSS researchers found that the U.S. curriculum was less focused and less advanced than Asian countries (Schmidt, McKnight, & Raizen, 1997). Curriculum in the U.S. “… appears not only to have been unfocused but highly repetitive, lacking coherence, and providing little rigorous intellectual challenge …” (Schmidt & McKnight, 1998, p. 1830). Lee (1999) indicated that concise national curricula in Asian countries help children learn in depth rather than breadth.

Schmidt and McKnight (1998) also pointed out the inconsistencies in the U.S. curricula. For example, even though curricula in the U.S. consistently covered more topics than did the curricula of all other TIMSS countries, no solid foundation in physics and physical science was provided through the first eight years, as compared to other countries. This finding was evidenced by students’ low scores in this area and teacher reports of smaller amounts of time spent covering physical science topics. Schmidt and
McKnight (1998) suggested that curricular differences could be worthwhile to compare to establish policies for increasing international competitiveness.

Another possible explanation of the superiority of Asian students in science could be the parental involvement which was not examined in this study. Researchers argue that one of the crucial factors of students’ success in Singapore and Japan along with other Asian countries is parental interest and investments in education (Aun, et al., 2006; Bray, 1999; Russell, 1997). Parents have high expectations for their children’s education due to the increasingly competitive academic environment (Khong & Ng, 2005) and they invest heavily in private tutoring (Bray; Bray & Kwok, 2003). Private tutoring is much more intensive in Asia, compared to the countries in Western Europe, North America and Australia (Bray). For example, Russell (1997) reported that in East Asian countries, nearly 70 percent of all students had received tutoring by the time they had completed middle school. However, in the U.S. only 25 percent of students attend private tutoring outside of school (OECD, 2000). Moreover, Xin and colleagues (2005) suggested that very large within-classroom variance in student achievement in Asian countries (i.e., Japan) could be explained by the private tutoring factor.

Conclusions

The purpose of this study was to explore the effects of student- and classroom-level factors on students’ science achievement at the elementary level within and across five countries. The TIMSS 2003 fourth grade data from Singapore, Japan, the U.S., Australia, and Scotland were analyzed by using a two level HLM with student factors at the level-1 and classroom factors at the level-2. Considering the amount of research in math and reading, the findings contribute to the existing literature in understanding the relations between student and classroom factors and science achievement in different countries.

The findings showed that the variance between students in the same classroom was larger than the variance across classrooms in all countries with the exception of Singapore. Classrooms in Japan were more homogenous with very little between-class variance. The student background variables (i.e., home resources, self-confidence) and composition of classrooms were strong predictors of elementary science achievement on the TIMSS 2003 test within and across countries. In general, there was very little
indication of teacher and instructional effects on the TIMSS 2003 science scores. There was evidence of positive effects of teacher support on science achievement. Emphasis on science inquiry was negatively related to science achievement in the U.S. and Australia. Overall, a large amount of variance in elementary science achievement remained unexplained within and across countries.

In an effort to explain different achievement levels in elementary science among countries, the select student- and classroom-level variables were not adequate. There could be various other factors in explaining the science achievement gap among these countries. It was pointed out that curricular differences among countries might be playing an important role in different achievement levels (Schmidt & McKnight, 1998). The U.S. curriculum was regarded as having breadth rather than depth and therefore, being less focused and less advanced than Asian countries (Schmidt, McKnight, & Raizen, 1997). Furthermore, parental interest and investments in education (Aun, et al., 2006; Bray, 1999; Russell, 1997), especially in private tutoring (Bray; Bray & Kwok, 2003) were considered as crucial factors of students’ success in Asian countries.

**Recommendations for Practice and Policy**

Having used a large data set and appropriate multilevel analyses, the study revealed some important relations between student- and classroom-level variables and elementary science achievement. However, the analyses were based on questionnaire data which do not necessarily reflect the complexity of actual classroom and teaching variables. Therefore, the study is exploratory in nature and could not provide direct evidence required for determining causal effects. This is a major limitation of all studies that use questionnaire survey methods. Yet, the results of the study might suggest some implications for education practice and policy.

While gender was less of a factor in fourth grade science achievement, the evidence of a subsequent widening gender gap by eighth grade warrants early attention (see Beaton et al., 1996b; Greenfield, 1995; Kahle, 2004; Solomon, 1997). Therefore, being aware of the gender trajectories in science is important in terms of providing academic and motivational assistance. Increasing awareness and knowledge of gender neutral instructional techniques, providing a non-threatening, rich and supportive environment for both genders in classrooms by elementary teachers are to be encouraged.
Schools and teachers can help break the stereotypes associated with gender by encouraging them to learn science and making connections between current schoolwork and future occupations (Greene & DeBacker, 2004).

Students’ self-confidence was found to positively predict science scores. According to Bandura (1997), the sources of students’ self-efficacy judgments reflect a complex interaction of internal and environmental factors. While there are limits to the impact that educators can have on students’ home environments, school and classroom environments could help promoting science self-beliefs. Britner and Pajares (2006) suggest that a ‘science-friendly’ atmosphere in classrooms could promote awareness for science learning, and lifelong careers in science across genders and cultures.

To improve students’ self-confidence in science it is recommended that teachers scaffold science activities by accommodating students’ needs and abilities, and provide them with stimulating tasks and materials (Britner & Pajares, 2006). Through successful modeling, teachers can help students build their efficacy beliefs. Science teachers can have students in mixed-ability groups so that more able students can provide models for less able students in science skills (Britner & Pajares). Efforts may include science awareness week or month, science fairs and science camps with the active participation of students. Moreover, scientists from various demographic groups may be invited to classrooms to work with students and to share their work and the experiences that led them to careers in science.

The results of this study suggested that availability of educational resources at home strongly predicted students’ science achievement. According to Bordieu (1986) parents who provide more resources for their children tend to be more involved in their education. Parental involvement is credited being a significant predictor of student achievement regardless of race, ethnicity, and SES (AAAS, 1996; Lazar, et. al., 1999). By monitoring children’s schoolwork regularly, and being more involved in their education, parents can be an effective buffer against the negative impact of low SES status on children’s academic outcomes (White & Kaufman 1997).

It is recommended for schools and teachers to develop a successful home-school partnership for improved student learning and positive attitudes toward science. Parents usually demonstrate a willingness to take a proactive role in their children’s education when they perceive authentic demand from students and from teachers (Hoover-Dempsey
et al., 2001; Tichenor, 1998). When planning for parent involvement in the school context, teachers need to create an inviting and non-threatening school environment for parents to increase the level of involvement (Eccles & Harold, 1996; Epstein & Salinas, 2004). Home-school initiatives, such as family science nights, interactive science projects and homework, and information workshops could capture parents’ interest and induce positive attitudes toward involvement in students’ science learning.

The findings of this study suggest that what children bring to classrooms plays an important role in their learning; therefore, developing a knowledge base for teachers regarding the influences of classroom and school composition has been highlighted (Murrell, 2001; Swanson-Gehrke, 2005). Academic support for schools serving students from low SES families is recommended. Honig, Kahne, and McLaughlin (2001) proposed that instructional needs, such as coaching and mentoring can be targeted by becoming familiar with students’ families and neighborhoods.

In regards to classroom-level factors, classroom climate may be an area in which teachers may make changes to positively impact science achievement. Findings suggest that developing a warm learning environment through teacher support has potential for improving students’ science achievement. Deci and Ryan (1985) suggested that perceived relatedness is a psychological need that increases satisfaction and motivation in a given domain. Therefore, it is recommended that teachers become more aware of the importance of creating a warm classroom climate in which students feel cared and accepted by the teacher and the other students. As evidenced by the results of this study, providing a warm and supportive learning environment may be more influential on student outcomes than teachers’ background characteristics.

In terms of the instructional variables, the effect of science inquiry on achievement in Australia and the U.S. was negative indicating that classrooms whose teachers placed more emphasis on science inquiry scored significantly lower than their counterparts who were exposed to less science inquiry. Given the emphasis on hands-on, inquiry-based instruction in science education this finding was surprising. A plausible explanation for this finding could be that the science inquiry activities in the U.S. and Australia may not have the same impact on all children. In a recent study, Connor and colleagues (2008) found that, hands-on, inquiry-based science activities had no effect on the achievement of second graders with weaker vocabulary skills, whereas, positive
effects were observed when children had average and high vocabulary skills. The authors proposed that science instruction could be more effective when it is individualized based on students’ background knowledge and language skills. Experimental studies are needed to investigate the impact of inquiry-based activities on students’ achievement.

At the classroom- and school-level, policy efforts could focus on the distribution of educational resources and funding. Unfortunately, schools serving students from low SES families suffer the most from unequal distribution of educational funds in the U.S. (Condron & Roscigno, 2003; Darling-Hammond & Sykes, 2003; Goesling, 2003; Greenwald, Hedges, & Lane, 1996; Lee & Luykx, 2005). Darling-Hammond and Sykes stated that inequalities associated with family circumstances are multiplied by inequalities at school and continue throughout students’ educations. Condron and Roscigno (2003) indicated “… higher spending promotes achievement through particular school resources” (p.32). School spending could promote schools’ physical condition, organization, materials, and learning environment (Congron & Roscigno). In Japan, for example, government allocates more resources for schools serving low-SES communities to compensate for poor family background (Goesling, 2003) which could be another reason for the homogeneity of achievement in Japan.

It is recommended for teacher education programs that pre-service teachers become more informed about the influences of home environment on students’ learning. By working with students’ from low-SES families prospective teachers can develop a knowledge base about students’ needs. Field placement in urban schools (Cook & Van Cleaf. 2000; Olmedo, 1997) and mentoring (Darling-Hammond, & Sykes, 2003; Ingersoll, 2004) have some research support in becoming more familiar with the contexts and reducing pre-service teachers’ frustrations in teaching disadvantaged student populations. Moreover, opportunities to work with parents regarding students’ science learning may help pre-service teachers in developing a strong home-school partnership when they enter the teaching profession.

Finally, it is recommended that teacher education programs put more emphasis on creating positive interactions in classrooms and activities for students to enjoy science. By becoming more familiar with students’ self-beliefs about science pre-service teachers can address students’ fear and anxiety of science and help minimize failures (Bandura, 1997; Britner & Pajares, 2006) when they start full-time teaching. In addition, pre-service
teachers need to be aware that they can improve students’ effort to success through care, friendliness, and understanding (Lumpkin, 2007; Noddings, 1992; Ryan & Patrick, 2001). Opportunities to teach science during field experience can help them uncover students’ beliefs about learning science and become familiar with the methods that would improve their classroom climate and practices when they start full-time teaching.

**Recommendations for Future Research**

Even though this study explained higher percentage of classroom-level variances compared to student-level variances in each country, there is still room for further investigation. As indicated by Burstein et al., (1995) and Weiss et al., (2003), the quality of instruction cannot be assessed without observing the teacher-student interactions. Direct classroom observations are more sensitive measures of the instructional and teacher effects and provide better estimates of the relations between instructional practices and student outcomes (Palardy & Rumberger, 2008). Qualitative studies that reflect the characteristics of science instruction and pedagogical practices in elementary schools, particularly in the area of science inquiry are recommended. The TIMSS 1999 video study, for example, provided valuable information about teaching practices in mathematics and science at eighth grade in more than one thousand classrooms in various countries. Future TIMSS studies might provide similar information at fourth grade level.

The TIMSS does not provide a ‘value added model’ in which students’ previous performance is taken into account. In value added models, improvement in performance could be more easily related to classroom-level and school-level factors (Hanushek, 1997). On the contrary, in studies that do not use previous achievement, student performance correlates with other variables (Raudenbush & Bryk, 2002; Hanushek, 1997; Xin et al., 2004). Knowing students’ prior performance can also be helpful in determining child by instruction interactions (i.e., variations in responses of children with ranging ability levels and prior knowledge to the same instructional practices) (Connor, et al., 2008). Children with different ability levels might respond differently to certain instructional practices. Therefore, future TIMSS studies might consider including an index of students’ prior academic level.

The current study investigated the effects of student- and classroom-level factors on elementary students’ science achievement by using the TIMSS 2003 fourth grade
data-set. Future studies might also utilize the TIMSS 2003 eighth grade data set to examine the influences of those factors on middle school students’ achievement. Investigating the gender effect might be valuable, since the past research indicates that the achievement gap between boys and girls tend to increase as children move to middle and high school (Greenfield, 1995, 1996; Kahle, 2004). Furthermore, the TIMSS 2003 eighth grade data set includes parents’ education level. Thus, using this variable at the student-level might help explain more variance in science achievement.

Finally, to investigate school effects on students’ science achievement, future studies might utilize multilevel analysis by using the school-level factors. The TIMSS questionnaire completed by school heads or principals can be used as the source of school-level variables. Factors, such as school composition, environment, and resources might be investigated in relation to elementary and middle school science achievement in different countries.
APPENDIX A

INTER-CORRELATION MATRIX BETWEEN VARIABLES FOR SINGAPORE, JAPAN, THE U.S., AUSTRALIA, AND SCOTLAND
### Table A.1 Inter-Correlation Matrix Between Variables for Singapore: Results of the Bivariate Analysis (n = 6,122)

<table>
<thead>
<tr>
<th></th>
<th>Science PVs</th>
<th>GENDER</th>
<th>HOME RES</th>
<th>SELF CON</th>
<th>TEXP</th>
<th>TEDUC</th>
<th>TMAJOR</th>
<th>TSUPPOR</th>
<th>SCIMIN</th>
<th>SCINQ</th>
<th>CLAS SIZE</th>
<th>M_SELF CON</th>
<th>M_HOME RES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science PVs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>-0.003</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOME RES</td>
<td>0.295</td>
<td>0.019</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF CON</td>
<td>0.196</td>
<td>-0.121</td>
<td>0.122</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>0.100</td>
<td>0.062</td>
<td>0.024</td>
<td>0.015</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEDUC</td>
<td>-0.062</td>
<td>-0.002</td>
<td>-0.039</td>
<td>0.000</td>
<td>-0.043</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMAJOR</td>
<td>-0.029</td>
<td>-0.015</td>
<td>0.011</td>
<td>-0.003</td>
<td>-0.043</td>
<td>-0.079</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSUPPOR</td>
<td>0.098</td>
<td>0.114</td>
<td>-0.029</td>
<td>-0.025</td>
<td>0.016</td>
<td>-0.008</td>
<td>-0.020</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIMIN</td>
<td>0.140</td>
<td>0.018</td>
<td>0.026</td>
<td>-0.009</td>
<td>-0.027</td>
<td>-0.017</td>
<td>-0.139</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCINQ</td>
<td>0.096</td>
<td>0.059</td>
<td>0.121</td>
<td>0.137</td>
<td>0.003</td>
<td>0.009</td>
<td>0.013</td>
<td>0.092</td>
<td>0.045</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASSIZE</td>
<td>0.201</td>
<td>0.038</td>
<td>0.063</td>
<td>-0.001</td>
<td>0.039</td>
<td>0.018</td>
<td>-0.028</td>
<td>0.047</td>
<td>-0.015</td>
<td>-0.033</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_SELF CON</td>
<td>0.124</td>
<td>-0.050</td>
<td>0.064</td>
<td>0.259</td>
<td>0.058</td>
<td>-0.001</td>
<td>-0.011</td>
<td>0.002</td>
<td>-0.034</td>
<td>0.056</td>
<td>-0.004</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M_HOME RES</td>
<td>0.524</td>
<td>0.022</td>
<td>0.348</td>
<td>0.047</td>
<td>0.068</td>
<td>-0.113</td>
<td>0.032</td>
<td>-0.043</td>
<td>0.076</td>
<td>0.100</td>
<td>0.180</td>
<td>0.183</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table A.2 Inter-Correlation Matrix Between Variables for Japan: Results of the Bivariate Analysis (n = 4,250)

<table>
<thead>
<tr>
<th></th>
<th>Science PVs</th>
<th>GENDER</th>
<th>HOME RES</th>
<th>SELF CON</th>
<th>TEXP</th>
<th>TEDUC</th>
<th>TMAJOR</th>
<th>TSUPPOR</th>
<th>SCIMIN</th>
<th>SCINQ</th>
<th>CLAS SIZE</th>
<th>M_SELF CON</th>
<th>M_HOME RES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science PVs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>-0.031</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOME RES</td>
<td>0.156</td>
<td>-0.019</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF CON</td>
<td>0.209</td>
<td>-0.089</td>
<td>0.076</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>0.044</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.007</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEDUC</td>
<td>0.030</td>
<td>-0.021</td>
<td>0.000</td>
<td>-0.003</td>
<td>-0.138</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMAJOR</td>
<td>0.008</td>
<td>0.009</td>
<td>-0.002</td>
<td>0.022</td>
<td>-0.057</td>
<td>-0.096</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSUPPOR</td>
<td>0.118</td>
<td>0.112</td>
<td>0.034</td>
<td>0.109</td>
<td>-0.011</td>
<td>0.007</td>
<td>0.013</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIMIN</td>
<td>0.036</td>
<td>0.001</td>
<td>0.044</td>
<td>0.021</td>
<td>-0.030</td>
<td>-0.015</td>
<td>0.007</td>
<td>0.059</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCINQ</td>
<td>0.094</td>
<td>0.064</td>
<td>0.035</td>
<td>0.110</td>
<td>0.017</td>
<td>-0.015</td>
<td>0.003</td>
<td>0.165</td>
<td>0.048</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASSIZE</td>
<td>-0.019</td>
<td>0.001</td>
<td>0.014</td>
<td>0.027</td>
<td>0.084</td>
<td>-0.051</td>
<td>0.119</td>
<td>-0.006</td>
<td>-0.046</td>
<td>0.017</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_SELF CON</td>
<td>0.036</td>
<td>0.012</td>
<td>0.000</td>
<td>0.263</td>
<td>-0.027</td>
<td>-0.011</td>
<td>0.084</td>
<td>0.101</td>
<td>0.080</td>
<td>0.041</td>
<td>0.101</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M_HOME RES</td>
<td>0.068</td>
<td>0.026</td>
<td>0.270</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>-0.008</td>
<td>0.036</td>
<td>0.162</td>
<td>0.021</td>
<td>0.053</td>
<td>-0.001</td>
<td>1</td>
</tr>
</tbody>
</table>
Table A.3  Inter-Correlation Matrix Between Variables for the U.S.: Results of the Bivariate Analysis (n = 7,623)

<table>
<thead>
<tr>
<th></th>
<th>Science PVs</th>
<th>GENDER</th>
<th>HOME RES</th>
<th>SELF CON</th>
<th>TEXP</th>
<th>TEDUC</th>
<th>TMAJOR</th>
<th>TSUPPOR</th>
<th>SCIMIN</th>
<th>SCINQ</th>
<th>CLASSIZE</th>
<th>M_SELF CON</th>
<th>M_HOMERES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science PVs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>-0.030</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOME RES</td>
<td>0.273</td>
<td>0.026</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF CON</td>
<td>0.258</td>
<td>-0.043</td>
<td>0.116</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>0.108</td>
<td>0.005</td>
<td>0.066</td>
<td>0.013</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEDUC</td>
<td>0.048</td>
<td>-0.001</td>
<td>0.047</td>
<td>-0.009</td>
<td>0.323</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMAJOR</td>
<td>0.020</td>
<td>0.006</td>
<td>0.016</td>
<td>0.005</td>
<td>0.025</td>
<td>-0.066</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSUPPOR</td>
<td>0.164</td>
<td>0.106</td>
<td>0.050</td>
<td>0.097</td>
<td>0.001</td>
<td>0.022</td>
<td>0.003</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIMIN</td>
<td>0.129</td>
<td>0.010</td>
<td>0.039</td>
<td>0.052</td>
<td>0.087</td>
<td>-0.071</td>
<td>-0.010</td>
<td>0.052</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCINQ</td>
<td>-0.030</td>
<td>-0.021</td>
<td>0.074</td>
<td>0.148</td>
<td>0.000</td>
<td>0.021</td>
<td>0.006</td>
<td>0.097</td>
<td>0.086</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASSIZE</td>
<td>-0.084</td>
<td>-0.002</td>
<td>-0.022</td>
<td>-0.036</td>
<td>-0.086</td>
<td>0.134</td>
<td>0.041</td>
<td>-0.007</td>
<td>-0.108</td>
<td>0.013</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_SELF CON</td>
<td>0.265</td>
<td>-0.029</td>
<td>0.152</td>
<td>0.331</td>
<td>0.040</td>
<td>-0.028</td>
<td>0.015</td>
<td>0.093</td>
<td>0.158</td>
<td>0.203</td>
<td>-0.112</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M_HOMERES</td>
<td>0.438</td>
<td>-0.017</td>
<td>0.386</td>
<td>0.130</td>
<td>0.172</td>
<td>0.121</td>
<td>0.043</td>
<td>0.095</td>
<td>0.100</td>
<td>0.078</td>
<td>-0.057</td>
<td>0.393</td>
<td>1</td>
</tr>
</tbody>
</table>

Table A.4  Inter-Correlation Matrix Between Variables for Australia: Results of the Bivariate Analysis (n = 3,573)

<table>
<thead>
<tr>
<th></th>
<th>Science PVs</th>
<th>GENDER</th>
<th>HOME RES</th>
<th>SELF CON</th>
<th>TEXP</th>
<th>TEDUC</th>
<th>TMAJOR</th>
<th>TSUPPOR</th>
<th>SCIMIN</th>
<th>SCINQ</th>
<th>CLASSIZE</th>
<th>M_SELF CON</th>
<th>M_HOMERES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science PVs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td>-0.014</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOME RES</td>
<td>0.251</td>
<td>0.045</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF CON</td>
<td>0.204</td>
<td>0.001</td>
<td>0.101</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>0.063</td>
<td>-0.016</td>
<td>0.021</td>
<td>0.028</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEDUC</td>
<td>0.015</td>
<td>0.011</td>
<td>0.008</td>
<td>-0.008</td>
<td>0.063</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMAJOR</td>
<td>0.018</td>
<td>-0.036</td>
<td>0.038</td>
<td>-0.027</td>
<td>-0.018</td>
<td>-0.002</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSUPPOR</td>
<td>0.076</td>
<td>0.115</td>
<td>0.022</td>
<td>0.083</td>
<td>-0.055</td>
<td>-0.040</td>
<td>-0.029</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIMIN</td>
<td>-0.003</td>
<td>0.022</td>
<td>0.014</td>
<td>0.033</td>
<td>-0.208</td>
<td>-0.036</td>
<td>0.010</td>
<td>0.007</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCINQ</td>
<td>-0.71</td>
<td>-0.07</td>
<td>0.053</td>
<td>0.174</td>
<td>0.040</td>
<td>0.013</td>
<td>0.045</td>
<td>0.077</td>
<td>0.076</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLASSIZE</td>
<td>0.050</td>
<td>0.012</td>
<td>0.024</td>
<td>0.000</td>
<td>0.213</td>
<td>-0.066</td>
<td>0.008</td>
<td>-0.32</td>
<td>-0.17</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_SELF CON</td>
<td>0.045</td>
<td>-0.007</td>
<td>0.017</td>
<td>0.313</td>
<td>0.090</td>
<td>-0.027</td>
<td>-0.086</td>
<td>0.034</td>
<td>0.104</td>
<td>0.244</td>
<td>0.000</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M_HOMERES</td>
<td>0.288</td>
<td>0.030</td>
<td>0.383</td>
<td>0.014</td>
<td>0.054</td>
<td>0.021</td>
<td>0.100</td>
<td>-0.031</td>
<td>0.036</td>
<td>0.062</td>
<td>0.062</td>
<td>0.045</td>
<td>1</td>
</tr>
</tbody>
</table>
### Table A.5 Inter-Correlation Matrix Between Variables for Scotland: Results of the Bivariate Analysis (n = 2,665)

<table>
<thead>
<tr>
<th></th>
<th>Science PVs</th>
<th>GENDER</th>
<th>HOMERES</th>
<th>SELF CON</th>
<th>TEXP</th>
<th>TEDUC</th>
<th>TMAJOR</th>
<th>TSUPPOR</th>
<th>SCIMIN</th>
<th>SCINQ</th>
<th>CLAS SIZE</th>
<th>M_SELF CON</th>
<th>M_HOMERES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science PVs</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENDER</td>
<td></td>
<td>-0.088</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOMERES</td>
<td>0.196</td>
<td>0.038</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SELF CON</td>
<td>0.194</td>
<td>-0.022</td>
<td>0.076</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXP</td>
<td>0.049</td>
<td>0.033</td>
<td>0.013</td>
<td>0.010</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEDUC</td>
<td>0.084</td>
<td>-0.033</td>
<td>0.024</td>
<td>0.005</td>
<td>-0.155</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TMAJOR</td>
<td>0.008</td>
<td>-0.034</td>
<td>0.023</td>
<td>0.053</td>
<td>0.051</td>
<td>-0.012</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TSUPPOR</td>
<td>-0.011</td>
<td>0.138</td>
<td>-0.059</td>
<td>0.069</td>
<td>-0.057</td>
<td>-0.010</td>
<td>-0.017</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCIMIN</td>
<td>-0.010</td>
<td>0.017</td>
<td>0.000</td>
<td>-0.003</td>
<td>-0.187</td>
<td>0.086</td>
<td>-0.118</td>
<td>-0.016</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCINQ</td>
<td>-0.053</td>
<td>0.055</td>
<td>0.024</td>
<td>0.188</td>
<td>-0.077</td>
<td>0.032</td>
<td>-0.042</td>
<td>0.115</td>
<td>0.066</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLAS SIZE</td>
<td>0.060</td>
<td>-0.003</td>
<td>0.035</td>
<td>0.018</td>
<td>0.139</td>
<td>-0.060</td>
<td>0.093</td>
<td>-0.036</td>
<td>-0.072</td>
<td>-0.024</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M_SELF CON</td>
<td>0.109</td>
<td>-0.011</td>
<td>0.040</td>
<td>0.275</td>
<td>0.037</td>
<td>0.019</td>
<td>0.193</td>
<td>-0.002</td>
<td>-0.011</td>
<td>0.142</td>
<td>0.064</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M_HOMERES</td>
<td>0.272</td>
<td>0.005</td>
<td>0.337</td>
<td>0.033</td>
<td>0.039</td>
<td>0.072</td>
<td>0.067</td>
<td>-0.142</td>
<td>0.001</td>
<td>-0.034</td>
<td>0.105</td>
<td>0.120</td>
<td>1</td>
</tr>
</tbody>
</table>
APPENDIX B

TIMSS 2003 FOURTH GRADE EXAMPLE SCIENCE ITEMS
A strong magnet will separate a mixture of
(A) clear glass and green glass.
(B) paper caps and plastic caps.
(C) iron nails and aluminum nails.
(D) sand and salt.

TIMSS 2003
Content Domain
Physical Science
Main Topic
Classification and composition of matter
Cognitive Domain
Conceptual Understanding
Key
C

Look at this diagram.

The saltiest water is in the
(A) Great Ocean
(B) Mountain Pond
(C) Old River
(D) Small Streams

TIMSS 2003
Content Domain
Earth Science
Main Topic
Earth's structure and physical features
Cognitive Domain
Conceptual Understanding
Key
A
Humans interpret seeing, hearing, tasting and smelling in the

A  brain
B  spinal cord
C  receptors
D  skin

The picture shows three solid objects of the same size floating in water.

Which object weighs the most?

A  Object A
B  Object B
C  Object C
D  They all weigh the same.
TIMSS 2003
Content Domain
Earth Science

Main Topic
Earth's structure and physical features

Cognitive Domain
Reasoning and Analysis

Key
D

TIMSS 2003
Content Domain
Physical Science

Main Topic
Physical states and changes in matter

Cognitive Domain
Conceptual Understanding

Key
C
Draw an arrow on the Old River to show the direction that the water flows.

Note: To receive credit, responses must indicate that the flow is "To the Great Ocean". The response may be in words rather than on the diagram or the arrowhead may be parallel to the flow but farther away from the river. For example, arrowhead anywhere up by the feeder rivers and streams even if direction is correct. Code 19 should be used for these other types of correct responses.

<table>
<thead>
<tr>
<th>Code</th>
<th>Response</th>
<th>Item: S011032</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correct Response</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Draws an arrow on the diagram that is on river or along the river below the feeder rivers and streams with an arrowhead pointing towards the Great Ocean.</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Other correct</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Incorrect Response</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>Draws an arrow pointing the opposite way or a prose description of this direction.</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Draws an arrow that points in any other direction.</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Other incorrect (including crossed out/erased, stray marks, illegible, or off task)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nonresponse</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Blank</td>
<td></td>
</tr>
</tbody>
</table>
The properties of three materials are compared in the table below. One of the materials is wood, one is rock and one is iron.

<table>
<thead>
<tr>
<th>Property</th>
<th>Material 1</th>
<th>Material 2</th>
<th>Material 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sinks in water?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Burns easily?</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Attracted by a magnet?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Identify the three materials by filling in the spaces below.

Wood is material number: _____________

Rock is material number: _____________

Iron is material number: _____________

---

Note: To receive full credit, all three materials must be identified correctly. Partial credit is given if one or two of the substances are identified correctly. If two substances are identified with the same number, neither one can be considered correct. For example, a response of 2, 1, 1 should be given a Code 11. A response of 2, 2, 2 should be given a Code 79.

<table>
<thead>
<tr>
<th>Code</th>
<th>Response</th>
<th>Item: 5031053</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>All three materials identified correctly: wood = 2, rock = 3, iron = 1</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Two materials identified correctly (1 left blank)</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Only wood identified correctly (2); rock and iron are blank or reversed</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Only rock identified correctly (3); wood and iron are blank or reversed</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Only iron identified correctly (1); wood and rock are blank or reversed</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Other partially correct (with at least one material correct)</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Incorrect (including crossed out/erased, stray marks, illegible or off task)</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Blank</td>
<td></td>
</tr>
</tbody>
</table>
Some of the organisms shown below give birth to young that develop inside the mother. Some of the organisms have young that hatch from eggs that are laid outside the mother.

In the table below, write down the names of the organisms that belong to each group.

<table>
<thead>
<tr>
<th>Organisms that give birth</th>
<th>Organisms that lay eggs</th>
</tr>
</thead>
</table>

Note: To receive full credit, ALL organisms must be classified correctly. Partial credit is given for responses with only one or two organisms missing or misclassified. If more than two organisms are missing or incorrect, then code 79 should be given.

<table>
<thead>
<tr>
<th>Code</th>
<th>Response</th>
</tr>
</thead>
</table>
| 20   | Organisms that give birth: Human, Dog, Whale  
       Organisms that lay eggs: Frog, Butterfly, Bird |
| 10   | One organism omitted or misclassified. |
| 11   | Two organisms omitted or misclassified. |

Source: International Association for the Evaluation of Educational Achievement (2007)
REFERENCES


Beaton A.E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., & Kelly, L. D.


Ingersoll, R. M. (2004). *Why do high-poverty schools have difficulty staffing their classrooms with qualified teachers?* Center for American Progress and the Institute for America’s Future.


Rowan, B., Correnti, R., & Miller, R.J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record, 104*(8), 1525-1567.


Walberg, H. (1986). Synthesis of research on teaching. In M. C. Wittrock (Ed.) 
*Handbook of research on teaching* (pp. 214–229). Washington, DC: American 


Wayne, A. J., & Young, P. (2003). Teacher characteristics and student achievement 

inside the classroom: A study of K-12 mathematics and science education in the 
United States*. Chapel Hill, NC: Horizon Research.


competition among immigrants and native-born ethnic groups. *Social Science 

Woessman, L., & West, M. R. (2002). *Class-size effects in school systems around the 
world: evidence from between-grade variation in TIMSS*. Bonn: Institute for the 
Study of Labor (IZA).


achievement, and cognitive skills: A rule-space model. *Studies in Educational 

teacher–parent partnership on early childhood performance in public and private 
Sibel Kaya received her B.S. degree in Chemistry Education from the Middle East Technical University in Ankara, Turkey. She taught in public elementary and middle schools in Ankara for two years. In 2002, she was awarded a scholarship by the Turkish Ministry of Education for graduate study in the United States. She received her MEd degree in Elementary Education from the University of Pittsburgh in December 2003 and received her PhD degree in the same field from the Florida State University in December 2008. She has also received a certification from the Program of Measurement and Statistics in the Department of Educational Psychology and Learning Systems at the Florida State University.

During her doctoral study, she taught an undergraduate level course, Teaching Science in Elementary Schools. During the last three years of her doctoral study, she worked as a research assistant in Dr. Carol Connor’s the Individualizing Student Instruction Project at Florida Center for Reading Research. She is currently working as a full-time researcher in the Individualizing Science Instruction Project at the same center. Her research interests include science education in elementary schools, parental involvement in education, and international comparisons of education.