Turkish Students’ Scientific Literacy Scores: A Multilevel Analysis of
Data from Program for International Student Assessment

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of
Philosophy in the Graduate School of The Ohio State University

By
Haci Bayram Yilmaz, M.A.

Graduate Program in Education

The Ohio State University

2009

Dissertation Committee:
Dr. James W. Altschuld, Advisor
Dr. Jerome V. D’Agostino
Dr. Sebnem Cilesiz
ABSTRACT

A vast majority of the studies exploring the associations between student and school related factors and standardized test scores were conducted in developed countries. On the other hand, research suggests that the generalization of the findings of those studies to developing countries often leads to incorrect conclusions. The purpose of this study was to explore the effects of selected student- and school-level factors on 15-year old Turkish students’ scientific literacy achievement. The Program for International Student Assessment (PISA) 2006 database was utilized to explore a) if there were differences among schools in terms of their mean scientific literacy scores, b) which student-level factors can explain the differences in students’ scores within a particular school, and c) which school-level factors can explain the scientific literacy differences between schools.

Hierarchical linear modeling (HLM) method was selected as the analytic method due to its usefulness in exploring relationships between a dependent variable and sets of layered independent variables. While the dependent variable of the study was represented by five plausible scores, independent variables consisted of 25 variables. Among those predictors, 15 were measured at student-level (level 1) and grouped in 4 clusters
(background characteristics, teaching and learning factors, affective factors, and out-of
school science related activities), and the remaining variables were measured at level 2
and grouped under two clusters (school resources and school context).

The results of the study indicated that more than half of the variation in students’
scientific literacy scores occurred among schools. While eight student-level variables—
grade, economic, social and cultural status, general and personal value of science,
responsibility for sustainable development, science self-efficacy, in-school time spent on
science learning, and hands-on activities—explained about one-third of the variation at
the student-level, three school-level predictors—school sector, mean economic, social
and cultural status, and mean in-school time spent on learning—accounted for more than
70% of the variation at that level.

Several findings of the present study were different than what mainstream
research suggested. Instrumental motivation and interest in science were found to be
insignificant in predicting scientific literacy of Turkish students; while the students’
perceptions about the general value of science had a positive association with scientific
literacy, their perceptions on the personal value of science had a negative association.
Many factors included in the teaching and learning cluster had either no relationship
(investigations and interactions) or negative relationship (hands-on experiences) with the
scientific literacy of Turkish students. Teacher-student ratio, lack of instructional
resources, and science teacher shortage were among the non-significant predictors of
Turkish students’ PISA scores.
DEDICATION

Dedicated to my parents, Fatma and Mehmet Yilmaz

And

My wife Sahinur and my daughter Mehlika Reyyan
ACKNOWLEDGMENTS

I would never have been able to finish this work without help and support from several people. I am most fortunate to have had the opportunity to work with the members of my dissertation committee, Drs. James Altschuld, Jerome D’Agostino and Sebnem Cilesiz.

I will be forever indebted to my mentor, Dr. Altschuld, who kindly agreed to advise me after I received my M.A. degree. He devoted an incredible amount of time toward this study, gave me insightful feedback on numerous drafts and challenged me to become a better researcher and a scholar. Dr. Altschuld’s intellect and positive attitude are qualities that I hope to emulate through my life. You have been an outstanding adviser and I am very grateful for everything you have done for me.

I am also immensely grateful to Dr. D’Agostino who provided methodological guidance and expertise throughout the dissertation, and Dr. Cilesiz who accepted to serve as a committee member and reviewed the chapters of this dissertation.

Many thanks are due to my best friends, Sukru Kaya, Cemalettin Ayas, Ozgur Balsoy, Osman Topac and Kerem Dirlikli, for their continuous friendship and support throughout the project.
I also want to thank Turkish Ministry of National Education for their scholarships that have supported my education life since high school. In return, I’ll do my best when to serve my country as an educator and a researcher, as promised.

I owe tremendous amount of gratitude to my dear wife, Sahinur, who newer failed to give me support and time to pursue my dreams and aspirations; and my daughter Mehlika Reyyan for the joy and happiness she has brought to my life. I hope we can spend more time together from now on.

The great support, sacrifice, unconditional love and prayers of my dear parents, Fatma and Mehmet, enabled me to finish my long journey from elementary to graduate school. I wish I did not have to leave you for many years to follow my dreams and become who I am now. My appreciation and affection go also to my sister Sukran and brothers Ilhan, Hakan, Mehmet, Mustafa and Ahmet for their continued support and encouragement.

And, thank you God… for everything!
VITA

March 10th, 1975 ......................................... Born-Cicekdagi – Kirsehir, Turkey

1997 .......................................................... B.S. Elementary Education
Gazi University, Ankara, Turkey

1997-1999 ................................................ Elementary School Teacher
Istanbul, Turkey

2001-2003 ................................................ M.A., Educational Policy and Leadership
The Ohio State University, Columbus, Ohio

2004-2007 ................................................ Graduate Research Associate
The Ohio State University

2005 .......................................................... Faculty Lecturer
Ohio Dominican University, Columbus, Ohio

2007-2008 ................................................ Graduate Teaching Associate
The Ohio State University, Columbus, Ohio

2008-2009 ................................................ Social Sciences Research Specialist/Consultant
Ohio Department of Education, Columbus, Ohio

FIELD OF STUDY

Major Field: Education

Specialization: Quantitative Research, Evaluation and Measurement
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
<tr>
<td>CHAPTERS</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Turkish Republic</td>
<td>4</td>
</tr>
<tr>
<td>Turkish Educational System</td>
<td>4</td>
</tr>
<tr>
<td>Program for International Student Assessment (PISA)</td>
<td>8</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>9</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>11</td>
</tr>
<tr>
<td>Research Questions</td>
<td>13</td>
</tr>
<tr>
<td>Conceptual and Theoretical Framework</td>
<td>13</td>
</tr>
<tr>
<td>Methodology</td>
<td>19</td>
</tr>
<tr>
<td>Assumptions of the Study</td>
<td>20</td>
</tr>
<tr>
<td>Limitations and Delimitations of the Study</td>
<td>20</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>22</td>
</tr>
<tr>
<td>2. Literature Review</td>
<td>24</td>
</tr>
<tr>
<td>Scientific Literacy</td>
<td>25</td>
</tr>
</tbody>
</table>
Importance of Scientific Literacy ............................................................................... 32
PISA ......................................................................................................................... 37
How PISA Assesses Scientific Literacy .................................................................. 39
Studies Using PISA Data ....................................................................................... 41
Turkey Specific Studies using PISA data ............................................................... 52
Toward a Focus on Current Study ......................................................................... 57
3. Methodology ........................................................................................................ 60
   Introduction ........................................................................................................... 60
   Population and Sample ....................................................................................... 60
   Instruments and Data Collection ........................................................................ 62
      Cognitive Test Questionnaires .................................................................. 63
      Student and Administrator Surveys .......................................................... 65
   Variables ............................................................................................................... 65
      Student-level Variables ............................................................................. 66
      School-level Variables .......................................................................... 70
   Validity and Reliability of the Measures ........................................................... 74
      Inter-coder Reliability Study on the Sub-sample of Assessment Booklets ...
      Multiple Marking (Variance Components) Study ........................................ 75
   Data Analysis ....................................................................................................... 77
4. Results ................................................................................................................... 82
   Introduction ........................................................................................................... 82
   Data Characteristics ........................................................................................... 83
   Inter-correlations between Variables ................................................................ 86
   Research Findings ............................................................................................... 89
      Research Question 1 ................................................................................. 90
      Research Question 2 .................................................................................. 92
      Research Question 3 .................................................................................. 96
   Testing HLM Assumptions .............................................................................. 105
Summary .............................................................................................................................. 109

5. Conclusions and Recommendations ............................................................................. 110
   Conclusions about the Research Design ....................................................................... 111
   Discussion of the Findings ............................................................................................. 113
     Decomposition of Total Variance ............................................................................. 113
     Student-level Factors and Scientific Literacy ......................................................... 114
     School-level Factors and Scientific Literacy ......................................................... 119
   Recommendations for Policy-makers ......................................................................... 121
   Recommendations for Further Research .................................................................... 126

LIST OF REFERENCES ........................................................................................................ 129

APPENDICES
   Appendix A – Collinearity Statistics of Level-1 Variables ........................................... 145
   Appendix B – Multiple Regression Model for Collinearity Diagnostics at Level-2 .... 147
   Appendix C – Normality Check for Predictors .............................................................. 149
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1: Types and Indicators of Scientific Literacy</td>
<td>31</td>
</tr>
<tr>
<td>Table 2.2: Student and School-level Variables Having Effect on Literacy Scores as Included in Ma’s Study</td>
<td>47</td>
</tr>
<tr>
<td>Table 3.1: School Types and Number of Students Attending</td>
<td>61</td>
</tr>
<tr>
<td>Table 3.2: Distribution of Items by the Dimensions of the PISA-2006 Science</td>
<td>64</td>
</tr>
<tr>
<td>Table 3.3: Independent Variables Included in the Present Study</td>
<td>73</td>
</tr>
<tr>
<td>Table 3.4: Reliability Values of Latent Variables</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.1: Descriptive Statistics of Student-level Variables</td>
<td>84</td>
</tr>
<tr>
<td>Table 4.2: Descriptive Statistics of School-level Variables</td>
<td>85</td>
</tr>
<tr>
<td>Table 4.3: Correlations between Student-level Variables</td>
<td>88</td>
</tr>
<tr>
<td>Table 4.4: Correlations between School-level Variables</td>
<td>89</td>
</tr>
<tr>
<td>Table 4.5: One-way ANOVA Model Results</td>
<td>91</td>
</tr>
<tr>
<td>Table 4.6: Random Coefficient Model Results</td>
<td>93</td>
</tr>
<tr>
<td>Table 4.7: Means as Outcomes Model Results</td>
<td>98</td>
</tr>
<tr>
<td>Table 4.8: Intercepts as Outcome Model Results</td>
<td>102</td>
</tr>
<tr>
<td>Table 5.1: Effect Sizes</td>
<td>124</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Figure 1.1</td>
<td>System Approach to School Functioning</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>OECD’s Conceptual Map of Educational Indicators</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Conceptual Framework for the Study</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Characteristics of a Scientifically Literate Person</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Components of Scientific Literacy in PISA 2006 Framework</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Erbas’ (2005) Model to Explain the Variability in Scientific Literacy</td>
</tr>
<tr>
<td></td>
<td>Achievement and Attitude toward School</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Graphical Representation of the Relationships between Significant School-</td>
</tr>
<tr>
<td></td>
<td>level Variables and School Mean Achievement of Scientific Literacy</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Normal Q-Q Plot of Level 1 Residuals</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Plot of MDIST vs CHIPC.</td>
</tr>
</tbody>
</table>
INTRODUCTION

Knowledge and skills are commonly viewed as key success factors for economic development and social well-being (Shalberg, 2007). Investigating educational results as given above has been important for researchers and policy-makers for a long period of time. Countries continually look at outcomes as a means to: improve the quality of education, demonstrate cost-effectiveness, and show public accountability for results (Bonnet, 2002). Policy-makers monitor educational processes and achievement to provide the best available education for students.

One movement tied to outcomes is globalization (Mortimor, 2001), particularly as related to the success of elementary and high school students in science and mathematics, critical areas in education systems. The influence of such areas on everyday life amplifies their prominence in educational assessment.

Historically, an early and major attempt to examine effectiveness was conducted by Coleman et al. (1966) in “Equality of Educational Opportunity”. While the aim of the study was to show the relationship between students’ ethnic and social background and achievement, the potential effects of school factors on learning attainment was looked at
as well. Three clusters of school characteristics were measured: (a) teacher characteristics; (b) materials, facilities and curriculum; and (c) characteristics of the groups or classes in which the pupils were placed. The three clusters accounted for about 10 percent of the variance in student performance after controlling for ethnic background and socio-economic status (Scheerens, 1999). The Coleman Report created a wave of research about how schools affect achievement (Kyriakides, 2006; Aksit, 2007).

Studies on this topic, however, have reached conflicting conclusions. Edmunds (1979) and Rutter et al. (1979) concluded that schools have a small but significant effect on achievement. On the other hand, Heyneman and Ransom (1990) suggested that family socio-economic background is more salient for academic achievement in industrial countries whereas school settings are as important as out-of-school factors, especially in science and mathematics achievement in low-income countries. Other investigations indicated that student attainment is a function of a “school” variable (Sammons, Thomas, & Mortimore, 1997).

The main methodology for exploring effects has been achievement tests. Gerberich (1963) noted that the first standardized written examinations were used in China for the selection of public officials more than three thousand years ago. The University of Bologna was the first educational institution using tests in 1219. The concept of testing as an education reform tool dates back to 15th century Italy where teacher salaries were tied to student performance (Abrams & Madaus, 2003). Likewise, utilization of achievement tests for school comparisons started in 1845, when Horace Mann, secretary of the Massachusetts State Board of Education, examined inequalities
among Boston Schools. He administered a 154-item written test with questions from variety of subjects (history, arithmetic, and science) to a sample of 530 students selected from more than 7000 (Evans & Nguyen, 2006).

Reliance on testing for assessment and accountability has clearly gained favor with policy-makers since the mid-twentieth century. Educational tests have served as input for three main functions to: help students and teachers in classroom learning; inform decision makers about the selection and placement of individuals for schooling and employment; monitor system-wide educational outcomes to make changes and revisions as necessary (Kornhaber, 2004).

In the last 20 years with the press of globalization and competitive job markets, countries have started to use international comparative assessments to determine how well their students are doing in reading, science, and mathematics. These studies are conducted by the International Association for the Evaluation of Educational Achievement (IEA) and the Organization for Economic Co-operation and Development (OECD). Among the most widely used assessments are Trends in International Mathematics and Science Study (TIMSS), Program for International Student Assessment (PISA), and The Progress in International Reading Literacy Study (PIRLS).

Turning to the situation of one country, the results indicate that Turkish students are well below international averages on various subjects, which increased the level of discourse about elementary and high school education in the country. The study being proposed will investigate PISA test results in Turkey. Before explaining the need for this
research endeavor, it is necessary to give an overview about Turkey, its educational system, and PISA.

Turkish Republic

As a result of World War I, the 600-year-old Ottoman Empire collapsed and the Turkish Republic was founded under the leadership of Ataturk in 1923. Turkey is a country of about 75 million people, slightly larger than Texas. Located in both Europe and Asia, it is at the crossroads between the East and the West and has been a member of the North Atlantic Treaty Organization (NATO) since 1953. Turkey-European Union (EU) relations began in 1963 after the country became a member of European Economic Community (EEC). In 1999, the government’s application for a full EU membership was accepted. Since then Turkey has been taking actions in different sectors such as industry, agriculture, and education in accord with the requirements of membership.

Turkish Educational System

After the foundation of the new republic, the Ministry of National Education (MONE) was reconstructed and secularized; and all educational institutions were placed under its authority. In 1927, education was made compulsory for all citizens between the ages 7-12, and the following year an adapted Latin alphabet was accepted. In 1973, the Basic Law for National Education (No. 1739) mandated a compulsory eight years of schooling, but the government could not implement the law until 1997 as a consequence of limited personnel and facility resources (World Bank, 2004). In August 1997, the
parliament issued a new Basic Education Law (No. 4306) that stipulated the eight-year compulsory elementary education for all children between the ages 6-14.

Secondary education includes all educational institutions of a general or vocational and technical character following elementary education. MONE oversees all educational institutions excluding universities, which are governed by the Higher Education Council. Secondary school, not yet compulsory, was extended from 3 to 4 years in 2005.

Since 1963, the State Planning Organization has been publishing 5-year development plans which set goals for important sectors such as energy, health, justice, and education with road-maps for achieving them. The rationale behind the objectives stated in the education section of the plans was the need to close the gap between developed countries and Turkey via opportunities and equipment to enable students to compete with their European counterparts.

In recent years the Turkish government has carried out many initiatives to achieve this goal. Among them, Basic Education Program (BEP) and Support for Using Computers in Education (SUCE) are the most important ones. The first phase of BEP started in 1998 and continued for 4 and 1/2 years. Activities included building new schools, increasing enrollment rates, and establishing more than 3100 Information and Communication Technologies (ICT) classrooms in about 2800 elementary schools. Each classroom was equipped with computers, printers, scanners, software (office programs, software for different courses, and computer literacy), electronic references, video players, overhead projectors, TV, educational videocassettes, and transparencies
(Akkoyunlu & Orhan, 2001). Three hundred thousand teachers were trained on the most effective use of ICT classrooms and hardware and software products. Even so it was noted that the computers were not used effectively and intensively other than in computer courses. (World Bank, 2004).

The second phase of the BEP (establishing new ICT classrooms in 3000 additional schools, increasing enrollment rates, building new schools, and renovating the old ones) started in 2002 and is currently in effect. It supports in-service teacher training to develop basic computer skills; provide assistance for utilizing ICT to improve instruction; and design training courses for special education and preschool teachers, and early childhood parents (World Bank, 2006).

Support for Using Computers in Education (SUCE) project started in June 2005. Its goal was to establish at least one computer lab and provide the internet connection for every elementary and secondary school. Overall MONE intended to create 14 thousand computer labs with 300 thousand computers by 2005-2006. In the first 6 months of the SUCE, with the help of public and private companies, about 100,000 computers were purchased and more than 4400 laboratories were in place. In addition, MONE signed contracts with computer companies to buy 120,000 more computers by the end of 2005-2006 school year (MONE, 2005a). Before SUCE, the computer-student ratio was 1:40 and only a small number of students had opportunities to use computers at schools. The program will continue until this ratio reaches to 1:12, which is still far below the 1:4 ratio in the United States (Market Data Retrieval, 2003). Coupled with the additional equipment, it would be possible to connect all schools to the internet (MONE, 2005b).
Along with the infrastructure and technology, MONE has implemented two big scale curriculum development studies since 1990. The first was a part of a comprehensive National Education Development Project, which focused on restructuring the education process and increasing the quality of education and student achievement. MONE put great weight on disciplined and systematic research and development. In line with this, the Education Research and Development Directorate (EARGED) was formed and it has fostered new curricula for 22 subjects taught in elementary and secondary schools as well as a new teacher education curriculum. But the secondary school curricula were not approved and implemented by the Board of Education (BOE). Karip (2005) summarized the reasons for the failure as:

- lack of a clearly shared vision and ownership of the new curriculum,
- lack of an agreement between EARGED and BOE,
- lack of an effective dissemination strategy, and
- overlapping roles among different units in MONE (Karip, 2005).

A second curriculum development program started in January 2004 as a two-step process with the goal of shifting from a behaviorist approach to more of a constructivist one. In the first step, new curricula for elementary school subjects were developed and piloted. Starting in 2005-2006, they have been implemented in all elementary schools in the country (Babadogan & Olkun, 2006). In the second step, new curricula for grades 9-12 were developed and expected to be in effect starting in 2008-2009.

Many other projects in addition to the ones mentioned here have been conducted with the purpose of reforming educational system in Turkey. Yet, the country is still far
from closing the gap with developed countries in terms of social and economic indicators. Acknowledging the need for in-depth analyses of reasons for that gap, MONE announced that findings of international assessments will be incorporated into future educational reform efforts (MONE, 2005c, d).

Program for International Student Assessment (PISA)

As a study conducted by the Organization for Economic Co-operation and Development (OECD), PISA is a large scale assessment designed to measure 15-year-old students’ literacy in mathematics, science, and reading. Although each cycle has included questions regarding all three subjects, one of them was emphasized in each study respectively. An additional investigation on problem solving was included only in 2003. A representative sample, between 4500 and 10000 students coming from at least 150 schools per country, takes the test.

In the PISA context, literacy means one’s capacity to apply his/her knowledge and skills to meet the challenges of today’s competitive world, rather than the learning of specific curricula. It assesses if young people have the ability to analyze, rationalize, and communicate their ideas effectively when confronted with problem situations that might occur in real life (OECD, 2006). Assessment tasks include some texts, tables, and/or diagrams describing a situation followed by a series of related questions. PISA surveys literacy in reading, mathematics, and science every three years since 2000. Both multiple choice and open-ended questions are used to measure relevant skills.

The main purpose of the PISA is to provide policy-oriented indicators of skills and knowledge. The testing program enables countries to monitor the outcomes of their
educational systems on a regular basis in terms of student achievement. Similar to most comprehensive international programs assessing student performance, PISA collects data about social, cultural, economic, and educational factors that might have a relationship with student performance (OECD, 2007). According to a report by Thomson, Cresswell, and Bortoli (2004), PISA provides data to answer the following questions:

- “How well are young adults prepared to meet the challenges of the future? What skills do they possess that will facilitate their capacity to adapt to rapid societal change?
- Are some ways of organizing schools and school learning more effective than others?
- What influence does the quality of school resources have on student outcomes?
- What educational structures and practices maximize the opportunities of students from disadvantaged backgrounds? How equitable is education provision for students from all backgrounds?” (p. vi.).

To find answers to these questions, 32 countries participated in PISA 2000 with the number increasing to 41 in 2003 and 57 in 2006. Although Turkey did not participate in the first cycle, it has participated in the other two.

Statement of the Problem

In order to maximize student learning, policy-makers allocate resources, time, and money to increase the quality of education. Sommons (1999) indicated that over the last three decades much academic interest has been focused on school improvement and
school effects. In that regard, recall that previously mentioned research has reached contradictory conclusions about the relative importance of different school variables on student learning.

One reason for this conflict is that most studies on the topic have used Ordinary Least Squares (OLS) regression analysis in order to control out-of-school factors such as gender and socio-economic status (SES). Nested data lack independence of observations, an important assumption of OLS. To illustrate, students in a particular school are more similar to each other than to students randomly selected from a city. Students in the same school tend to come from a less diverse community in terms of many characteristics like SES, family background, and ethnicity (Osborne, 2000; Raudenbush and Bryk 2001). Those students share the same school resources. Failure to incorporate the hierarchical nature of the data into statistical analyses leads to increasing Type I error probabilities and aggregation bias when making statistical inference (Creemers & Kyriakides, 2006; Raudenbush & Bryk, 1986).

Another reason for the lack of agreement regarding school and student factors comes from the models used to investigate them. Although different research paradigms have been introduced to examine variables, their infrequent and isolated use isn’t helpful in regard to better understanding what underlies school success and student achievement (Scheerens & Bosker, 1997; Teodorovic, 2005; Creemers & Kyriakides, 2006). According to Raudenbush and Bryk (1986) the reason for this observation is closely associated with methodological advances in that “the available analytic models tend(ed) to limit conceptualization as to what can be empirically tested through such models.
There is a natural hesitancy to form conceptualization when it remains unclear how to test the fruits of that conceptualization” (p. 15).

Furthermore, the majority of the studies on the topic have been done in developed countries such as the Unites States and Europe. Despite the obvious limitation, the results are often generalized to other areas that are still developing and/or have contextual differences. Research suggests that there is lack of scientific evidence indicating that school, student, and background variables have the same association with student outcome in diverse settings (Yilmaz, 2008).

For instance, Heyneman (1986) reported that the relationship between socio-economic status and educational achievement in less-developed countries is weaker compared to developed ones. To stress the importance of the differences among countries, Jenkins (2004) said “I am bothered that the researcher so readily moves from a particular context in Israel or New Zealand to claim some strong generalization” (p.117). Utilizing the appropriate statistical techniques, this study will explore the effect of different variables on learning to find what school and student-level factors have an impact on science literacy in Turkey, a developing country.

Significance of the Study

Since the sixties, Turkey has been working on adjustments in its educational system to become a part of the developed world and a full member of EU. Still, a World Bank report (2005) maintains that it is far from reaching this goal. As Turkey’s prospects grow more reliant on the economies of its European neighbors and other countries, and participation in the global labor market, higher levels of schooling and achievement will
be required of entry level workers. The World Bank posited that Turkey needs to raise educational qualifications of its population to those of the EU and upgrade the competencies of students to prepare them for national and global job markets.

Recognizing the imperative for educational reform, MONE initiated a new series of changes, such as the massive curriculum development, using new technologies in classrooms, and modifications in teacher education programs, to create more effective schools. With that in mind, Guven (2007) indicated, “The question for policy-makers to contemplate is: how can the national government enhance the effectiveness of the system and consolidate efforts to support school improvement endeavors?” (p.364). Furthermore, the Education Research and Development Directorate (EARGED) of MONE is planning to initiate investigations of school and student factors in Turkey using international databases (EARGED, 2008).

Examining different student and school level variables measured by PISA 2006, the proposed study will respond to the need and has a potential to help MONE answer its concerns and make future decisions about the educational system. This research is also important in that, researchers, teachers, and administrators could potentially learn from its findings. As Altschuld and Kumar (2002) concluded after analyzing the chapters of a book, *Evaluation of science and technology education at the dawn of a new millennium*, that there is a pressing need for in-depth research on science education, and it is important to know more about goings-on in science classrooms.

This study will focus on a number of factors related to in-class activities, school characteristics, and educational resources. The findings will potentially help in
understanding the effect of those factors and to see if particular instructional activities (e.g., hands-on applications) are related to scientific literacy. Such an outcome would have immediate application for teachers and administrators.

This research is also important in that it will address the problems resulting from applications of inappropriate statistical methods through the application of better theory-based models. Utilizing hierarchical linear modeling and benefiting from the educational effectiveness models, this study will examine student and school-level variables to answer the following specific questions.

Research Questions

1. Are there differences in the mean achievement scores of scientific literacy among Turkish schools?
2. Which student-level measures explain the differences (if any) in Turkish students’ achievement of scientific literacy?
3. Which school-level measures are associated with differences in mean school achievement?

Conceptual and Theoretical Framework

Historically, educational effectiveness research has been guided by several conceptual frameworks (e.g., Carroll, 1963; Frank, 1998; Heyneman, & Loxley, 1983; Rumberger, 1995; Scheerens, 1999) that can be divided into input-output studies and process-output studies (Rumberger & Palardy, 2005; D’Agostino, 1997). Rooted in the theory of human capital (Becker, 1964), the input-output (economic or education
production) approach mainly concentrates on the “supply of selected purchased schooling inputs and educational outcomes controlling for the influence of various background features” (Monk, 1992, p. 308). This paradigm describes schools as a productive entity in which inputs are transferred into outputs. The assumption is that an increase in inputs like per-pupil expenditure, or reducing teacher/student ratio leads to increments in outcomes such as attainment or job placement. While on the surface this seems reasonable studies conducted using economic models revealed that the relation between input and outcomes is more complex than was assumed, and maximizing the inputs does not necessarily result in higher student achievement (Hanushek, 1986; Monk, 1992; Trippollini, 2008).

The process-output studies of educational effectiveness research emphasize the issues related to two features of schools: organizational (school climate, interactions between teachers, leadership, and school-environment) and teaching and learning (student self belief, motivation and interest, and in-class practices). Studies based on this approach have revealed that there is a significant relationship between process factors and student learning (Carroll, 1963; D’Agostino, 1997; Dowson and McInerney, 2001; Lumsden, 1994; Marsh and Byrne, 1999; Pajares, 1996).

Starting from the 1990s input-output and process-output approaches have been combined in multilevel models and testing propositions through more sophisticated statistical methods. Contemporary international studies like TIMSS and PISA provide data on both input and process level, and enable researchers to look at the hierarchical relations between school, classroom, and student level units. For example, OECD, utilizing PISA data, employs a system approach (OECD, 2007) presented in Figure 1 to
investigate how contextual factors, input characteristics, and classroom and school experiences affect students’ academic development (Sammons, 2006; Schreens, 2000).

Figure 1.1: System Approach to School Functioning

Input indicators are structural (resources) and contextual (demographics, school type, and size) of the schools (Hanushek, 1994; Levin, 1997), and process indicators consist of school policies and practices (participation in decision making process, in-class activities) (Rumberger, 2005; Teddlie & Reynolds, 2000).

OECD developed a set of indicators with the purpose of comparing different aspects of schooling among countries. From its three-year International Indicators of Educational Systems (INES) project from 1988 to 1991, a conceptual map was created (Figure 1.2), which permits countries to see how their schools are doing, compared to
others (Schreens, 1995; Wendel, 2000). OECD’s framework has been utilized by PISA surveys to investigate three areas of schooling: effective teaching and instruction, school administrative characteristics, and school resources (OECD, 2007).

Figure 1.2: OECD’s Conceptual Map of Educational Indicators
The conceptual model given in Figure 1.3, a focused synthesis of the prior configurations, will guide this study. However, testing all indicators of educational effectiveness is not feasible. Therefore, a selection of variables based on a literature review will be examined. Following the Raudenbush and Willms’ (1995) analogy, between and within school variation in science literacy achievement will be explained by student and school level factors.
While student background characteristics (demographic and family information), teaching and learning factors (student-teacher interaction and learning activities), affective factors (student self-efficacy and motivation in science), and out-of school activities related to science (watching TV programs about science and reading books on scientific topics) are grouped under student-level variables; school resources (teacher shortage and material/artifact availability) and school context (school type and school mean socio-economic status) will be incorporated under school-level variables.

Although research on school effectiveness underlies the importance of school processes such as staff participation in decision-making and shaping school policies, the above framework excludes that type of variables due to the nature of Turkish educational system. In Turkey, the decision making process is so centralized that a school principal is required to communicate directly with an official in the ministry about non-routine issues requiring a decision (Fretwell and Wheler, 2001). As a result the variation in school processes is very limited and it is hard to test if they have an effect on educational outcomes.

Contemporary researchers identified three major strands of school effectiveness research: school effect research evolves from input-output studies, and focuses on schools’ ability to affect the outcomes, effective schools research examines the processes of successful schools, and school improvement research is concerned with processes involve in school-change (Reynolds, Tedlie, Creemers, Schreens, and Townsend (2000);
and Tedlie, Reynolds and Sammons (2000). Because this study is more focused on selected school inputs than processes, it can be considered as a school effect study.

Methodology

Publicly available datasets including school principal surveys, student surveys, and student assessment files will be taken from OECD’s website (www.pisa.oecd.org). The datasets include information collected in 57 countries; for this study the Turkish PISA 2006 principal survey file and a student file containing both survey and achievement data will be extracted and used.

Independent variables of the study come from survey responses and grouped under six categories in two levels as shown in Figure 1.3. The dependent variable consists of five plausible science literacy achievement scores. Those values represent five samplings from theoretical distributions of each student’s measures on cognitive items. Specific details about independent and dependent variables and procedures are provided in Chapter 3.

Hierarchical linear modeling (HLM) has been selected as the analytic method to explore relationships between a dependent variable and sets of layered independent variables. HLM has advantages over traditional statistical methods, such as multiple regression or analysis of variance because it provides statistically efficient estimates of regression coefficients, correct standard errors, confidence intervals, and significant tests. Since HLM can incorporate covariates at any level of a hierarchy, it is possible to examine the differences in between-school mean achievement scores and the relative
variation in student scores within particular schools is accounted for by predictors in a model (Goldstein, 1999).

Assumptions of the Study

Several assumptions underlie the proposed study:

1. The participating schools and the students are a representative sample of high schools and 15-year-old students from across Turkey.
2. Participants’ responses to the survey questions are honest and reflect the real situation.
3. The latent variables used in this study are estimated appropriately by PISA data analysis teams.
4. The questionnaires and cognitive items used in data collection procedures were translated and adapted from English to Turkish. It is assumed that translations have been done correctly.

Limitations and Delimitations of the Study

1. This study uses a database to conduct secondary data analysis. There is a lack of control over the variables and constructs of the data and the base does not deal with all factors exhibited in the literature. To illustrate, neither survey includes sufficient information about school climate which may well impact student learning.
2. The constructs were developed based on self-reports from principals’ and students’ answers to PISA 2006 surveys. There are no other data sources to verify
accurate reporting, which may create concerns about validity. For example the
constructs related to in-class instructions were created from student responses and
teachers might have differed if the same questions had been asked of them.

3. Because this study is exploratory in nature and the dataset used in this study is
cross-sectional rather than longitudinal, drawing causal inferences should be done
with caution. The purpose of testing statistical models is to predict science
literacy achievement rather than investigating causal relationships among the
variables.

4. PISA studies do not provide one estimate for each student’s literacy score;
instead, they use five different values as estimations of the test scores. Since the
values have variance, there is a degree of uncertainty in measurement.

5. This study adopts the definition of PISA 2003 for the term “scientific literacy”. In
PISA 2006, different components like interest in science and support for scientific
inquiry are considered as part of it, hence, plausible values for those variables are
also provided; but this study only uses the scores on the cognitive items as
dependent variables.

6. The sampling frame of this study does not allow detecting classroom level factors,
which is usually done in school effectiveness research. There is no information if
the students from a particular school are coming from same or different
classrooms, thus classroom level factors such as instructional factors are
confounded in student level factors.
Definition of Terms

*Affective factor* refers to a wide range of beliefs and feelings such as self-efficacy and motivation that exist beyond cognition (McLeod, 1992).

*Anatolian high schools* are schools providing education in a foreign language, often English, in subjects such as mathematics and science. Admission to Anatolian high schools is based on a nation-wide competitive placement test results.

*Background variable* is “an explanatory variable that can affect other (dependent) variables but cannot be affected by them. For example, one's schooling may affect one's subsequent career, but the reverse is unlikely to be true” (Upton & Cook, 2006).

*Economic, social and cultural status* is a composite factor created by using information about student’s highest parental education, parental occupation and home possessions (OECD, 2007).

*Latent variables* are theoretical constructs that are not measured directly, but inferred from directly measured variables (Byrne, 2001).

*Plausible values* are random numbers drawn from the estimated proficiency distribution of students with similar item response patterns and background characteristics. Item responses and background characteristics are used to predict those values (OECD, 2002; OECD, 2009).

*Science high schools* are schools with a concentrated curriculum in mathematics and science courses. Admission to science high schools is based on nation-wide placement test results.
*Scientific literacy* is the capacity to use scientific knowledge to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity (OECD, 2003).

*Self efficacy* is “people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances” (Bandura, 1994).

*Socioeconomic status* is an individual’s or family’s ranking on a scale based on their ability to access or control over products or services (Mueller & Parcel, 1981).
CHAPTER 2

LITERATURE REVIEW

The world has entered an era of rapid change over the last several centuries, due to the industrial revolution and advancements in science. These events have led to a vast expansion in knowledge and new artifacts, techniques, and procedures. Products of scientific research have helped us understand and change the world. Scientists acting upon what has been learned have developed things ranging from painkillers to virtual communities, and transformed our lives as compared with those of past generations (Millar & Osborne, 1998).

The growth in knowledge and the rising interdependence of financial markets call for a workforce with solid scientific literacy skills. Scientific literacy is essential to an individual’s participation in modern society. It contributes to the personal, social, professional and cultural lives of all people. A large proportion of problems and issues encountered by individuals in their daily lives require an understanding of science before they can be fully addressed. Science and technology confront individuals at personal, community, national and even global levels. Hence it is important for policy-makers to
ask about the degree to which all individuals in society are prepared to in relevant content areas and have understanding of them (OECD, 2006).

As United States National Assessment Governing Board puts it the “future depends on scientifically literate citizens who can participate as informed members of society and as a highly skilled scientific workforce, well prepared to address challenging issues at the local, national, and global levels” (NAGB, 2008, p. v.). For this reason, scientific knowledge is one of the most important goals for countries.

The focus of the hapter is first on what scientific literacy is and why it is important for individuals and countries in general, and Turkey in particular. Next it looks at how PISA assesses scientific literacy and then it turns to reviews of studies conducted using data from previous PISA data collection efforts. Lastly, information regarding the Turkish students’ science achievement and scientific literacy is discussed.

Scientific Literacy

Since the 1990s the term “scientific literacy” has become progressively more prominent in discussions of the purposes of science education (Millar, 2006). Two major reports published in the last decade of the 20th century have concentrated on the importance of scientific literacy. Project 2061: Science for All Americans (Rutherford & Ahlgren, 1990), released by the American Association for the Advancement of Science, advocated that the schools need to emphasize scientific literacy and to teach it more effectively rather than teaching more and more content.

The second publication was a summary report of a seminar series took place in England in 1997: Beyond 2000: Science Education for the Future (Millar & Osborne,
Science educators discussed four principal areas regarding the science education in that country: a) successes and failures of science education to date, b) characteristics of the science education needed by young people today, c) the content and structure of an appropriate model for a science curriculum, and, d) potential issues that would be raised by the implementation of such a curriculum, and how these might be handled. Although the scope of the seminars was local, recommendation for improving the nature and quality of the science education drew attention from the world. The main suggestion of the Beyond 2000 report was that enhancing scientific literacy should be the chief purpose of the science curriculum from 5 to 16.

The decision of the OECD to survey students’ scientific literacy rather than their knowledge of school science in the PISA project is perceived as the clearest indication of the importance of scientific literacy (Harlen, 2001; Millar, 2006). Yet, when it comes to defining the term, the discussion becomes more complex. (Millar, 2006; Miller, 2006; Scarse, 2007). According to Bybee (1997) scientific literacy has defied precise definition not only because is it a broad concept encompassing many educational themes, but also it has been used as a slogan to support more and better science teaching.

Literature indicates that scientific literacy was first used after the launch of Sputnik which started a discussion on the state of the science education in United States. A report published by Rockefeller Brothers Fund in 1958 explored the ways in which United States should respond the rapid scientific and technological changes in areas like nuclear energy and aerospace engineering. It argued that although such fields needed highly-skilled personnel, a technically trained workforce was particularly important to
catch up with rapid changes in science (DeBoer, 2000). An article published in the same year by Hurd (1958) stressed attention to the importance of providing education to young people to live and work in a quickly changing world. He stressed that children’s learning experiences at schools should foster “the development of an appreciation of science as an intellectual achievement, [and] as a procedure for exploration and discovery” (p.16). In Hurd’s definition, scientific literacy means an understanding of science and its impacts on social life, values, politics and economic problems.

In 1963 Robert Carleton, executive secretary of the National Science Teachers Association (NSTA), asked a number of prominent scientists and science educators to define scientific literacy. The majority of the answers indicated that one who has scientific literacy should have some content knowledge in a broad range of science fields, and few other responses mentioned science-society relationships and the process of conducting a scientific study (Garfield, 1988). A literature review by Pella, O’Hearn and Gale (1966) examined 100 articles to determine the components meaning of the term. They concluded that the most important “referents” of scientific literacy were a) interrelations between science and society, b) ethics of scientific study, and c) nature of science.

In the same vein NSTA declared that main characteristics of a scientifically literate person are using science concepts, process skills, and values in making everyday decisions as he/she interacts with other people and with environment; and understanding the interrelationships between science, technology and other facets of society, such as social and economic development (NSTA, 1971). Although the idea that science
education should be tied with social life was well established, it gained popularity when the NSTA announced it as the most important purpose of science education in its position statement: *School Science Education for the 1970s* (DeBoer, 2000).

Based on previous studies Shen (1975) suggested that scientific literacy has three categories: practical, civic, and cultural. *Practical scientific literacy* consists of the “possession of the kind of scientific knowledge that can be used to help solve practical problems” (p.46), and knowledge addressing the basic human needs like food and shelter. *Civic scientific literacy* was considered as having enough knowledge to become aware of science-related public issues in order to be involved in the decision making process related to social issues such as health and natural resources. *Cultural scientific literacy* referred to having knowledge about the scientific ideas that lead major achievements.

John Miller (1989) noted that scientific literacy does not imply an ideal or acceptable level of understanding, and proposed that scientific literacy should be functional. An individual with functional scientific literacy has at least a level of understanding of science and technology to function minimally as a citizen and consumer in society. Miller argued that functional scientific literacy has three dimensions:

1. Understanding the process of science. He believed that to be classified as having a minimal level of understanding of the process of science, one should be able to give a satisfactory explanation of what it means to conduct a scientific study and to distinguish between science and pseudo science;
2. Understanding basic scientific terms and concepts. Miller (1989) added that one should be able to understand essential terms like atom, molecule, gravity, or
radiation so that he can comprehend much of the public debates on scientific results or policy issues concerning science and technology; and

3. Understanding the impact of science in society and on individuals in coping with public policy issues. This dimension of functional scientific literacy requires comprehension of general knowledge on variety of scientific and technological areas like pros and cons of using computers in K-12 education, or the effects of using hormones to grow vegetables and subsequent impacts on health.

Hurd (1998) summarized the 350-year history of efforts to tie academic science with its applications to real life, and concluded that scientific literacy is “a process of acquiring, analyzing, synthesizing, coding, evaluating, and utilizing achievements in science and technology in human and social contexts” (p.413). A scientifically literate person, he argued, is capable of using science knowledge in forming opinion, resolving problems, and taking action (Hurd, 1998). Important characteristics of a person who has scientific literacy can be summarized in Figure 2.1 based on Hurd’s comprehensive study and interpretation by the writer of this dissertation.
Figure 2.1: Characteristics of a Scientifically Literate Person

- Symbiotic relationship between science, technology and human affairs
- Scientific knowledge has limits and risks
- Science produces cumulative and falsifiable knowledge

- Evidence from propaganda
- Theory from dogma
- Data from myth and folklore
- Knowledge vs. opinion

- when more data is needed to make valid judgments
- when causal relationship cannot be drawn
- the importance of collaboration to solve science–social problems.
- that a synthesis of knowledge from various social and scientific fields is needed to solve social problems

- science problems in social contexts may have more than one “right” answer
- science in social contexts is multidimensional (political, judicial, ethical, and sometimes moral).

- how to analyze and process information to generate knowledge and go beyond facts
In his book *Achieving scientific literacy: From purposes to practices*, Bybee (1997) identified four types of scientific literacy: nominal, functional, conceptual and procedural, and multidimensional. He proposed that the aim of science curriculum should be to increase scientific literacy level of students from illiteracy or nominal literacy to multidimensional literacy. Types and indicators of scientific literacy in Bybee’s view are summarized in table below.

<table>
<thead>
<tr>
<th>Types of Scientific Literacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
</tr>
<tr>
<td>Understanding if a question or topic is scientific</td>
</tr>
<tr>
<td>Providing naive explanations to phenomena</td>
</tr>
<tr>
<td>Exhibiting misconceptions</td>
</tr>
</tbody>
</table>

Table 2.1: Types and Indicators of Scientific Literacy Based on Bybee’s Classification
Importance of Scientific Literacy

Despite the debate on its meaning, there has been a consensus upon the importance of scientific literacy (Shen, 1975; Garfield, 1988; Bybee, 1997; Millar & Osborne, 1998; Dos Santos, 2007; Murcia, 2009). Laugksch (2000) summarized previous arguments at two levels: macro and micro. The former relates to benefits to the country, society, and science; the latter corresponds to direct effects in improving an individual’s life.

As indicated in the beginning of this chapter, a reason for promoting scientific literacy is its relationship with economic development of nations, the macro level (Laugksch, 2000). He argued that competing successfully in international markets is crucial to increase national wealth and to participate in the international competition for producing and distributing new technologies. Becoming successful in this race relies upon a strong national research, development, and production programs and ultimately creating and disseminating new artifacts helps countries develop and amplify their assets. A steady supply of scientists, engineers and a well trained workforce with an appropriate level of scientific literacy is vital.

The second macro level advantage of greater scientific literacy concerns public expectations of science. Laugksch stated that as the public understands the objectives, processes, and capabilities of science, it develops realistic and realizable expectations. He discussed that unrealistic expectations may lead to loss of confidence in, and withdrawal of support for science. The solution for this problem, he thought, would be in increasing the level of scientific literacy among public.
Another macro level reason why scientific literacy should be promoted is the relationship between science and culture. Citing from Thomas and Duran (1988), Laugksch stressed that effective integration of science in the culture is crucial for the health of a society. If science is isolated from the wider culture, the public not only fails to understand its role and importance but also may see it more as only an area of specialization, thus creating a mixture of fear and adulation.

For Laugksch (2000) the last macro level is related to science itself. He believed that scientifically literate individuals are more likely to value science and support spending public funds for scientific studies. On the other hand, Schibeci (1989) found that expectation unrealistic, and noticed that individuals with some level of scientific understanding are “likely to approve of some aspects of science, but not of others. If understanding is empowering, it is naive to suppose that there is an automatic link between understanding and approval.” (p. 246).

When it comes to the micro level benefits, Laugksch (2000) stated that as one’s scientific literacy increases he feels more confident and competent to deal with science-and technology-related issues in daily life. More knowledgeable individuals can make informed decisions about personal choices such as smoking, diet, and vaccination. Scientifically literate individuals are more likely to take full advantage of technical developments in their jobs and exploit new career opportunities (Royal Society, 1985).

The second major benefit at the micro level deals with one’s intellectual capacity and aesthetic view. According to Laugksch (2000) knowledge of science is an important element of being an educated person in the 20th century. Scientific literacy is believed to enable one to express feelings in creative and impressive ways through fine arts such as
paintings and photography (Shortland, 1988). As Trefil (2007) illustrated, a photographer can take exciting pictures of space or an artist can reflect his inspirations from molecular science in his drawings.

Many researchers have proposed that science curriculum in schools should focus on scientific literacy (Bybee, 1997; Cross & Pierce, 1999; Roth & Lee, 2004), and some argued that school curriculum should even go beyond that and create a discussion culture around social and political issues in relation to science (Freire, 1994; Delizoicov et al., 2002; Dos Santos, et al, 2006). For example, students should be able to think on and discuss about global and political problems and ask questions like “Why do rich nations produce a lot of waste, whereas in other nations there is not enough food for their inhabitants? Why are the natural resources and the workers from Third World nations exploited by multinational industries…?” (Dos Santos (2007, p.379).

Due to the benefits of scientific literacy, understanding, communicating and using scientific knowledge has been one of the most important educational goals for countries. Turning to the focus of the current study, Turkey, that educational goal is more challenging than it is for most of the other countries. As a member candidate Turkey has to narrow, if not close the gap between the EU countries in many areas including economic development, labor productivity and education. Therefore it is imperative to pinpoint EU’s holistic aim on these issues to understand the challenge Turkey is facing.

Recognizing that industrialization is depended upon education, the EU has emphasized since early 1990s the importance of investment in developing people’s skills (Jones, 2005). Its Lisbon meeting in 2000 was a turning point in formalizing EU’s aspiration. In the European Council’s Lisbon report, political leaders declared the new
strategic aim for the EU as “becoming the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion.” (p. 3).

A subsequent report by the Education Council of the EU identified a number of priorities for the future and the contribution which the education and training systems must make if the above goal is to be achieved. EU needs to increase the general levels of scientific culture so that its citizens would have a basic understanding of mathematics and science to understand issues and make informed choices to contribute to the public debate, decision making and the legislative process (Education Council, 2001).

A recent publication of the European Commission (2007) indicated that although there have been numerous projects and actions to realize the target set in 2001, the sign of improvement is not of the scale to bring about substantial impact. The report concluded that science in Europe’s schools is essential for the development of individual European countries; furthermore, it is also crucial if the members of the EU are to collectively make progress towards meeting the Lisbon objectives.

While EU is making progress on improving science teaching in classrooms and scientific literacy in the general population, comparative studies have revealed that Turkey is behind in both areas. A comprehensive research study on adult scientific literacy conducted by the EU compared scientific literacy in European countries and United States in 2005. More than 32 thousand individuals from 33 countries participated in face-to-face interviews and answered questions about their a) interest, level of information about scientific discoveries, and involvement in science and technology issues, b) image and knowledge of science and technology, c) attitudes towards science
and technology, d) responsibilities of science and policy-makers, and e) perceptions of European scientific research (European Commission, 2005). Only 2% of Turkish adults were qualified as civic scientifically literate, which put the country in last place. The countries with highest literary rates were Sweden with 35%, USA with 28%, and the Netherlands with 24% (Miller, 2006).

International studies comparing student achievement in science and scientific literacy produced similar results. Turkey has joined in 2 of the 4 Trends in Mathematics and Science (TIMSS) investigations done by the International Association for the Evaluation of Educational Achievement (IEA) which assessed 4th and 8th graders’ mathematics, science and reading achievement since 1995. Turkish students’ mean achievement scores on TIMSS are lower than that of their European counterparts. All 17 of the EU member countries (except Cyprus) participating in TIMSS-2007 had better mean science achievement scores than Turkey (IEA, 2008). Although the scope of TIMSS somewhat differs from that of PISA, it is important to note that country rankings are very similar in both studies.

Turkey has also taken part in PISA-2003, and PISA-2006 and had a low level of literacy in mathematics, reading and science. In the last PISA study, Turkish students’ scientific literacy scores were higher than only one EU member country, Romania; out of twenty four involved EU members (OECD, 2006).

As in the previous chapter, the data set analyzed in the current study is derived from PISA-2006. What is the history of PISA, the nature of questions in PISA tests, and previous studies conducted employing the program?
Due to international demand, the OECD decided to implement an ongoing assessment system with the help of participating countries and started developing PISA in 1997. In OECD’s initial publication about the project, *Measuring Student Knowledge and Skills: a New Framework for Assessment*, it was argued that PISA produces policy-oriented and internationally comparable indicators of student achievement on a regular and timely basis (OECD, 1999). The assessments are administered to 15 year-olds at the end of their compulsory schooling in many countries and conducted every three years. Building an understanding of the degree to which education systems in nations are preparing their young citizens to become lifelong learners and to play productive roles in society is the main goal of PISA.

PISA provides countries with measures to monitor the outcomes of their education systems in terms of student achievement, within a common, internationally agreed framework. OECD announced that the primary reason for developing and conducting PISA is to provide sound information along with other indicators of education systems such as drop-out rates and measures of teachers’ job satisfaction to inform policy. The results of this large-scale international assessment are expected to help guide educational reform and school improvement programs, especially where schools or education systems with similar inputs achieve different results (OECD, 1999).

Every cycle allocated about two-third of the test items to one area in 2000, 2003, and 2006. Although the three domains correspond to school subjects, PISA does not examine how well young people have mastered specific curriculum content; but it assesses the extent to which students have acquired the wider knowledge and skills in that
they will need in the future (OECD, 1999). Beside test items, students respond to a background questionnaire. School principals also fill-out an instrument to supply information about school context and resources. As an ongoing system PISA is way to monitor trends in student learning and skills in participating countries.

The first OECD assessment, PISA-2000 was administered to about 265 000 students from 29 OECD members and 14 partner countries. While all 30 OECD member countries participated, the number of partner countries decreased to 11 in the second cycle. Despite this fact the number of students tested increased to about 275 000 in 2003 assessment. The last PISA was administered in 57 countries including all OECD members and 27 partners. More than 400 000 students were tested in 2006. As of March of 2009, 67 countries have signed up to participate in PISA 2009 (OECD, 2009). Although number of students tested varies, most countries have selected between 4 500 and 10 000 students for the assessments.

Students spend two hours to answer test items. The total set of items is divided into different clusters. Each cluster is taken by a sufficient number of students for appropriate estimates to be made of the achievement levels on all questions by students in each country and in relevant sub-groups (i.e., gender, SES) within a country. Test takers also spend 20 minutes answering the context questionnaire (OECD, 1999, 2004). Based on students’ and school administrators’ responses to PISA questions OECD produces various indicators:

1. a baseline profile of students’ knowledge and skills;
2. contextual indicators, showing how such skills relate to important demographic, social, economic and educational variables;
3. trends from the cyclical nature of the data collection and that show changes in outcome levels and distributions; and

4. trends in relationships between student-level and school-level background variables and outcomes over time (OECD, 1999).

How PISA Assesses Scientific Literacy

PISA examines whether 15 year-old students have functional literacy skills as explained by Miller (1989). The main question that PISA 2006 aimed to answer was: “What is it important for citizens to know, value, and be able to do in situations involving science and technology?” (OECD, 2007, p. 20). That question requires an assessment of cognitive and affective aspects of scientific literacy (OECD, 2007). Cognitive aspects take account of students’ knowledge and effective use of that knowledge as they answer personal, social, and global scientific enquiries. The cognitive part of PISA attends to students’ scientific knowledge and their ability to evaluate scientific information. Affective aspects look at students’ awareness of science and technology’s roles in shaping material, intellectual, and cultural environments, and their interest, support and motivation to engage in science-related issues.

OECD (2007) characterized PISA-2006’s perception of scientific literacy as consisting of four interrelated aspects: Context, Knowledge, Competencies, and Attitudes.

- **Context**: recognizing life situations involving science and technology.
• **Knowledge**: understanding the natural world on the basis of scientific knowledge that includes both knowledge of the natural world, and knowledge about science itself.

• **Competencies**: demonstrating competencies that include identifying scientific issues, explaining phenomena scientifically, and drawing conclusions based on evidence.

• **Attitudes**: indicating an interest in science, support for scientific enquiry, and motivation to act responsibly towards, for example, natural resources and environments (p.25).

Test items were framed in real-life situations to assess scientific literacy competencies which are influenced by students’ attitudes and knowledge. The following figure illustrates that framework.
Studies Using PISA Data

Many publications about PISA have focused on the nature of the assessments, procedures of implementation, and basic descriptive statistics such as differences in country means and subgroups of populations. Most research conducted on PISA data has employed reading or mathematics literacy because they were primary emphases of the PISA 2000 and PISA 2003 respectively, only a small number of studies incorporated science literacy.

Baker and his colleagues (2004) examined the relationship between instructional time and achievement in different subjects including mathematics and science using data
from three international student assessments: PISA-2000, Third International Math and Science Survey (TIMSS-1999), and the International Study of Civic Education (CIVICS-1999). They analyzed data from more than 40 countries and found generally weak association between in-class time spent on math and science and students’ scores on those subjects at the cross-national level. The majority of countries showed no significant correlation between school time spent on math learning; however, while the association was positive in countries such as Greece, Japan, Poland and the Republic of Korea (r up to .29); it was negative in others including Luxemburg, Netherlands, and Switzerland (r as low as -.52) suggesting more classroom instruction is associated with lower achievement.

The association between instructional time and achievement in science was found to be greater. The percentage of variance in achievement explained by the number of hours spent in science instruction was 5% on average. In all countries, the relationship between school time on science and student achievement in science was positively correlated, with the majority of indices being significant.

In a similar study, Is-guzel and Berberoglu (2005) examined the effects of attitudes towards reading and math, student-teacher relations, classroom climate, communication with parents, and use of technology on mathematics and reading literacy. The PISA 2000 database was employed to compare Brazil, Japan, and Norway. Results of the structural equation models (SEM) indicated that the use of technology in Brazil, communication with parents in Japan, and attitudes towards reading in Norway had the strongest effects on mathematical literacy. Also, in all three countries reading literacy had a strong impact on mathematical literacy.
Other findings of Is-guzel and Berberoglu were that positive attitudes towards reading had a negative relationship with mathematical literacy but positive relationship with reading literacy measures. Communication with parents had a positive relationship with reading literacy skills in participating countries. A disciplined classroom environment fostered more success in PISA tests in Japan; however, it caused a reverse result in Brazil. Finally, the use of technology had a strong effect on reading skills in Brazil; however, it had no or negative effects in Norway and Japan respectively.

A comprehensive study by Ma (2003) focused on the effects of student and school characteristics on achievement of immigrant and non-immigrant students using a Canadian sample from the PISA-2000 data. He reported that non-immigrant students achieved significantly higher scores than immigrant students in reading and science literacy with no statistically significant differences in mathematics achievement between the two groups. His study included 21 student-level and 21 school-level independent variables to capture a broad range of factors with the potential to influence literacy of reading, science and mathematics. Two-level hierarchical models were tested to analyze the data coming from 1718 immigrant students from 438 schools and 27196 non-immigrant students from 1090 schools.

Ma provided detailed information on the factors related to literacy achievement on three areas of the PISA assessments and many other important variables. Females scored significantly higher than males on the reading literacy test in the immigrant group. Students with higher father's SES, living with both parents, having fewer siblings, better home educational resources, more social communication, less family educational support, more enjoyment in reading, an home language as either English or French scored
significantly higher than their counterparts in immigrant group. Three of the 21 school-level variables, attending schools with lower academic pressure, higher teacher morale and better teacher participation in decision-making were statistically significant predictors of reading achievement among immigrant students at the school level. This two-level model accounted for 22% of the student-level variance, and 62% of the school level variance in reading achievement when immigrant students were the focus.

Among the nonimmigrant students, female students scored statistically significantly higher in reading literacy. Other variables associated with higher reading literacy were higher parental education, higher SES, living with both parents, having fewer siblings, speaking either English or French at home, better home educational resources, more home cultural communication (i.e., discussing political or social issues, books, films or TV programs), more social communication (i.e., discussing how well they do at school, eating the main meal together, spending time talking with them), more home activities related to classical culture (i.e., visiting a museum, attending an opera, ballet or classical symphony concert), with less family educational support (a family member helping the student on school work), more reading enjoyment, reading more diversely, and spending more time on homework. The school-level variables associated with higher reading literacy in the non-immigrant group were attending larger schools, higher percentage of teachers in language arts with a university degree, better disciplinary climate, better student behavior, and more school autonomy. This hierarchical model accounted for 28% of the reading literacy variance among non-immigrant students and 60% of the variance among schools attended by these non-immigrant students.
Six of the 21 student-level variables, higher father’s SES, fewer siblings, better home educational resources, more home cultural communication, less family educational support, and more reading enjoyment were statistically significant predictors of higher mathematics literacy achievement of immigrant students. Better disciplinary climate, less academic pressure, higher student-teacher ratio, and attending a school in a metropolitan area with a population under 100,000 were associated with higher mathematic literacy among the immigrant population. Ma’s model accounted for 10% of the student level variation, and 59% of the school-level variation in immigrant groups’ mathematical literacy scores.

In the non-immigrant students’ mathematics literacy achievement, males outperformed females. Most of the student-level variables were statistically significant predictors of mathematics literacy of non-immigrants. Higher parental SES, and parental education, living with both parents, fewer siblings, having better educational resources at home, more cultural communication, more cultural activities related to classical culture, reading more diversely, receiving less family educational support, and spending more time on homework were associated with higher mathematics literacy. Being at a large school, having a better disciplinary climate, less academic pressure, and better student behavior were among the characteristics of high achieving schools in terms of mathematical literacy of non-immigrants. That model is associated with 17% of the student level variation and 52% of the school-level variation in mathematics literacy achievement.

When it comes to the scientific literacy of the students, the pictures stays about the same in both groups. Immigrant students with higher parental SES, and higher
mother’s education level scored significantly higher. In addition, fewer siblings, having a non-immigrant mother, less family educational support, and more reading enjoyment were among the significant predictors of scientific literacy of immigrant students.

Schools with more autonomy, higher teacher morale, and higher student-teacher ratio were more successful than others at the PISA 2000 scientific literacy test. This hierarchical model accounted for 17% of the student level variation and 55% of the school level variation in scientific literacy of non-immigrant group.

As it was in the mathematical literacy, male students outperformed the females on scientific literacy in the non-immigrant group. Higher parental SES and parental education, fewer siblings, speaking English or French at home, better home educational resources, more home cultural communication, more home activities related to classical culture, less educational support, more reading enjoyment, and spending more time on homework were the factors increased scientific literacy of that group. School factors associated with higher student scores were attending larger schools, better disciplinary climate, lower academic pressure, more teacher participation in decision making process. This model accounted for 23% of the student level variation and 57% of the school-level variation for non-immigrant group. Ma’s findings on the variables having an effect on students’ scores are summarized in table 2.2 below. Note that “+” represent a positive effect, and “-” represents a negative effect.
<table>
<thead>
<tr>
<th>Levels</th>
<th>Variables</th>
<th>Groups and Assessment Areas</th>
<th>Immigrant</th>
<th>Nonimmigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Math</td>
<td>Reading</td>
<td>Science</td>
</tr>
<tr>
<td>Student Level</td>
<td>Being a male student</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Higher Mother’s SES</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Higher Father’s SES</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Higher Mother’s Education</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Higher Father’s Education</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Nonimmigrant mother</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Member of a wealthier family</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Living with both parents</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Having fewer siblings</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>French or English as home language</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Higher self-expect SES</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Better home educational resources</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Higher home cultural communication</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Higher home social communication</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>More home activities related to classical</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>culture</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Less family educational support</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>More reading enjoyment</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>More diversity in reading</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>More time on homework</td>
<td>+</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2: Student and School-level Variables Having Effect on Literacy Scores as Included in Ma’s Study

(Continued)
<table>
<thead>
<tr>
<th>Levels</th>
<th>Variables</th>
<th>Immigrant</th>
<th>Nonimmigrant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Math</td>
<td>Reading</td>
</tr>
<tr>
<td>School Level</td>
<td>Larger school size</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Higher teacher-student ratio</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>More teacher participation in decision making</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Higher teacher morale</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Low academic pressure</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>More teachers with a university degree in assessment area</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disciplinary climate</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Student behavior</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>School autonomy</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>Being in the Atlantic region</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Located in a metropolis with fewer than 100,000 people</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>
Utilizing the PISA 2000 data Turmo (2004) compared the relationship between SES represented by cultural, social and economic capital, and level of scientific literacy of students coming from Norway, Sweden, Denmark, Finland, and Iceland. Results of the multiple regression analysis showed that a comprehensive SES measure explained between 10% and 21% of the variation in scientific literacy scores of the students. Turmo found a weak relationship between the social capital (social ties used for different purposes like career choice) and scientific literacy ($R^2$ less than 3%), and economic capital (financial resources) and scientific literacy ($R^2$ from 10% to 20%) in all the Nordic countries. He reported stronger relationship between the cultural capital (engaging high status cultural practices like attending theatre) and the level of scientific literacy in several of these countries ($R^2$ between 10% and 20%).

In her study Nonoyama (2005) focused on the effect of family background and school resources on students’ reading literacy achievement. She used PISA 2000 and PISA 2003 data from more than 40 countries Based on multilevel analyses Nonoyama concluded that across the countries a multidimensional SES measure representing parental education, job status, and family possessions has stronger effects on achievement across countries than the simpler measures of SES like mother’s education; and effects of variables related to family background are consistently larger than the effect of school-level variables (quality of school infrastructure, instructional resources, information resources, teacher credentials, and student-teacher ratio) across countries.

Marks (2006) was another researcher that analyzed PISA 2000 results. He compared students from 30 countries in terms of their performance on all three assessment areas to ascertain the extent that between and within-school differences in
literacy scores can be attributed to students’ socio-economic background and their home resources. Before focusing on this study, it is important to provide information about three measures that are not directly given in the PISA database but included in his study.

- **School program**: Marks classified school programs into four groups in terms their primary function: lower-non-academic programs (middle or secondary schools prepare students to either vocational highs schools or the job market); higher non-academic schools (higher secondary schools preparing students for higher level vocational programs, or work preparation programs), lower academic schools (middle schools preparing students for academic programs), and higher academic schools (higher secondary schools preparing students for academic programs).

- **Academic location**: Using the program information above and student grades, he created a continuous composite variable, academic location, and

- **Global measure of the socio-economic status**: Marks believed that the variable representing socio-economic background of the students was limited because it was constructed based on father and mother’s education and occupation status. Mark developed the new variable combining by measures of socio-economic background, material resources (i.e., washing machine), and cultural resources (i.e., number of books at home).

Marks (2006) demonstrated that grade, school program and academic location had higher bivariate associations with reading, mathematics, and science literacy in many countries. In addition, both between and within-school differences in student performance in all three assessment areas are largely accounted for by those variables, rather than socio-economic background especially in countries with highly tracked school systems.
He thought that a school which prepares its students for universities or a school having an academic program would be more responsible for between school performance differences than would socio-economic background variables. He also drew attention to the possible influence of general academic environment that the school creates. Recognizing the previous findings on the effect of socio-economic variables on student achievement, Marks elaborated that higher-status parents have more resources to choose academic locations than lower-status parents; “once in different academic locations, differences in student performance arise and increase because students from higher socio-economic backgrounds are in pedagogically richer and academically more challenging environments” (p.38).

Fuchs and Wobmann (2006) used the PISA-2000 database to investigate the association between student characteristics, family backgrounds, home inputs, and institutional variables with math, science and reading literacy achievement in 31 countries. Decomposition of the total variance of student-level test scores pointed out that in each subject area more than half of the total variance occurred at student level; between 37% (in science) and 43.8 (in math) of the variance occurred at school-level, and between 9.6% (in reading) and 16.1% (in math) of the total variance occurred at country level. These results suggest that variables related to students (e.g. SES) are more important than variables related to schools (e.g. school resources), and variables related to countries (e.g. GDP). A more detailed discussion of their findings is given below.

Fuchs and Wobmann found that many variables related to students’ characteristics, family-background data and school-related variables of resource use and institutional settings (significant associations with student performance). Major findings
are as follows: the effects of family background measured by parental education, parental occupation, or the number of books at home are considerably stronger in reading than in math and science. Additional findings were that boys outperform girls in math and science, but the opposite is true in reading.

Better equipment with instructional material and better-educated teachers had a positive effect on achievement in all areas, but class size did not. Students attending schools with external exit exams outperformed others in math and science and performance was positively associated with school autonomy in personnel-management and process decisions (deciding budget allocations within schools, textbook choices). In addition, students in privately managed schools performed better than students in public ones. The findings on institutional factors were consistent across the three subject areas.

Turkey Specific Studies using PISA data

EARGED publishes reports on the performance of Turkish students based on basic descriptive statistics: country ranks, mean scores, standard deviations, ranges, and bivariate relationships on selected variables. The institution encourages and invites researchers to conduct more detailed analyses to explore the associations between potential exploratory variables and performance measures. Unfortunately, only a few such studies have been found.

In a study of mathematic achievement across school types and geographical regions Berberoglu and Kalender (2005) employed PISA-2003 dataset. Results revealed that students attending Science High Schools, Anatolian High Schools, and Police College had better scores than students attending other type of schools. The researchers
found that science high schools performed about two standard deviations above the international PISA mean, general high schools performed one standard deviation below, and vocational schools performed one and a half standard deviation below. From Multivariate Analysis of Variance (MANOVA) results in both assessments, students living in Eastern and South-eastern part of the country did significantly worse than the rest; however, this difference was not practically significant.

Erbas (2005) used PISA 2003 data in his study and via structural equation modeling (SEM) he explored the effect of student-teacher relations (relation), feeling of loneliness (lonely), attending out-of-school courses (coaching), and remedial study and homework (remedial) on scientific literacy achievement (scielit) and attitude towards school (atschool). Erbas’ model shown in Figure 2.3 below.
Erbas’ proposed model explained 34% of the variation in attitudes towards school and 14% of the variation in scientific literacy scores of the students. The goodness of fit index (GFI) and adjusted goodness of fit index (AGFI) were 0.97 and 0.96 respectively; the Root Mean Square Error of Approximation (RMSEA) and standard root mean residual (RMR) were equal to 0.04. This model explains 0.34 percent of variance on attitudes towards school, and 0.14 percent variance of scientific literacy skills.

In the second step of the analysis, he included a bi-polar latent variable (isced) which consisted of information about the number of books at home and preschool attendance (no preschool education, one year or less, and more than one year), and removed the path between “coaching” and “scielt”. The new model increased the explained variance in attitude towards school and scientific literacy increased to 36% and
49% respectively. The standardized path coefficients (gammas) of the final model revealed that better student-teacher relation has a positive impact on both dependent variables; number of books and preschool attendance had a negative impact on attitude toward school, and positive impact on scientific literacy; on the contrary, receiving remedial work had a positive effect on students’ attitude, but a negative effect on their literacy; loneliness had a negative impact on the literacy scores; and attending out-of-school science courses had a negative effect on the attitude variable. Finally, students’ attitudes toward school had a negative effect on their literacy scores, which suggests the better the students’ attitudes, the lower their scores on scientific literacy.

The most comprehensive study about the Turkish PISA data was conducted by Is-guzel (2006). She aimed to analyze the impact of numerous student and school-level variables on mathematical literacy of students across Turkey, EU member and candidate countries. Here, only the results about Turkey are summarized.

Is-guzel included more than twenty variables in each level, almost all available variables on student and school questionnaires. The ANOVA model of HLM indicated that 45% of the variance in mathematical literacy was among schools. The Means as Outcomes Model, which provided information about school characteristics associated with the differences in the mathematical literacy skills in Turkey showed that while mathematics self-efficacy, school size, proportion of females enrolled at school, academic selectivity, quality of school’s physical infrastructure, teacher morale and commitment, and student-related factors affecting school climate had a positive association with the outcome variable, student-teacher ratio, student-mathematics teacher ratio, student morale and commitment, and teacher-related factors affecting school
climate had a negative association. This indicated that 75.9% of the between-school variance in mathematical literacy is accounted for by these factors.

The variables included in Random Coefficient Model focusing on student-level factors’ influence on mathematical literacy accounted for 37.5% of the student-level variance in the variable. Males performed statistically higher than females. There was a positive effect of home educational resources, sense of belonging at school, mathematics self-efficacy, self-concept, preference of control strategies, and disciplinary climate in mathematics lessons. But better student-teacher relations, higher mathematics anxiety, preference of elaboration, and memorization strategies led to significantly lower achievement scores.

Two studies were conducted by Anil in 2008 and 2009 focused on PISA-2006. In the first one, Anil (2008) used time spent on mathematic learning, cultural possessions at home (availability of art work, literature, and poetry at home), home environment (availability of desk, computer, study room, software, internet, and book to study math), and education (education level of parents, and number of books at home) to explain the variability in scores of mathematics literacy in PISA-2006. All variables had a significant positive association with achievement and a total of 35% of the variability in the outcome measure was explained.

In the second study, Anil (2009) focused on the scientific literacy achievement. She analyzed how the scientific literacy scores’ change in terms of students’ parents education, availability of computers at home, cultural possessions at home and attitudes toward science. Based on the outcomes of stepwise multiple regression analysis with scale scores obtained through factor analysis, 20% of the variability in the literacy scores
was explained by the model. Although all variables entered in the equation had statistically significant contributions to the explained variability, father’s education was the most important predictor.

Demir, Kilic, and Depren (2009) also utilized multiple regression analysis in their paper discussing the effect of students’ background, affective factors, learning strategies, and school climate on their mathematics achievement on PISA 2003. Independent variables explained about 34% of the variation in the mathematics literacy. Students’ economic, social, and cultural background was the most important predictor of their literacy scores. Students relating learning material to their past experiences, and students with less anxiety and more positive attitudes toward mathematics performed better than others. Furthermore, better school climate was associated with higher achievement.

The last study on Turkish students’ PISA achievement was by Balim and his colleagues in 2009. The researchers examined the relationship between students’ use of information communication technologies and scientific literacy achievement. Specifically did having a computer at home to study lessons and do homework and the availability of educational software and an internet connection at home have an association with the outcome variable? Balim et al. (2009) found that students answering “yes” to any one of the three questions had significantly better scores than their counterparts.

Toward a Focus on Current Study

Previous literature indicates the importance of scientific literacy for advancement of individuals and societies. Sadly, comparative studies showed that there is a significant gap between citizens of Turkey and European countries in this regard and parallel results
in reading and mathematics. At the same time, the literature on factors related to Turkish students’ performance on PISA is extremely limited. Apart from the general reports issued by OECD and EARGED, just a few publications could be found. Among those, only three looked at scientific literacy achievement (Erbas, 2005; Anil, 2009, and Balim et al, 2009) and there are shortcomings in those studies. To illustrate, Erbas (2005) did not provide in depth explanations of his models and findings. He did not discuss why he included a bi-polar variable consisting of only two-items, mother’s education and preschool attendance, nor did he elaborate on interpretation of the significant effect of that factor. More importantly, he did not follow the commonly accepted rule that each factor should include at least three observed items (Guilford, 1956; Schumacker & Lomax, 1996). In his second model, Erbas introduced other predictive variables to increase the explained variability in students’ attitudes toward school and scientific literacy; on the other hand, he omitted a significant path between attending out-of school courses (coaching) and literacy, potentially resulting in model misspecification. No rationale was given for doing so.

Compared to Erbas, Anil (2009) seemed to have used appropriate methods and estimation procedures but did not explain if she used all plausible values (PVs) in her analyses, choosing one or taking an average of all five. As indicated by OECD (2000; 2005) PVs are not point estimates, they are randomly selected values of estimated ability score distribution for individuals; therefore all five should be taken into account for statistical analyses. Neither Anil (2009) nor Erbas (2005) utilized school-level variables in their research. Therefore, their studies have a narrow scope.
Likewise, Demir, Kilic, and Depren (2009) used only student-level variables in their study. What is more, instead of using all five PVs, they used the mean of them as dependent variable. Although Demir and his colleagues utilized the same dataset used by Is-guzel (2003), some of their conclusions contradicted what Is-guzel reported. While Is-guzel pointed a negative association between preference of elaboration and memorization strategies, and mathematics literacy, Demir, Kilic, and Depren claimed a positive association.

Very few studies like those of Berberoglu and Kalender (2005), and Is-guzel (2006) provided comprehensive and statistically sound results, and they did not look at scientific literacy achievement. By focusing on the potential variables associated with scientific literacy of 15-year old Turkish students at the individual and school-levels, this study will fill the gap in research and provide insight about the relative importance of those factors.
CHAPTER 3

METHODOLOGY

Introduction

In this study the focus is on the relationship between students’ scientific literacy and selected student and school-level factors. The dataset for this study was derived from the OECD’s PISA 2006 International Database. This chapter contains the methodology to answer the research questions of Chapter 1: (a) Are there differences in the mean achievement scores of scientific literacy among Turkish schools? (b) Which student level measures explain the differences (if any) in Turkish students’ achievement of scientific literacy? and (c) Which school level measures are associated with differences in mean school achievement? The text is organized as follows: target population, sample, instruments, procedures, variables, validity and reliability of measures, and hierarchical models tested.

Population and Sample

The target population in PISA was defined as 15-year olds enrolled in any form of schools in participating countries. Students who were aged between 15 years and 3
months and 16 years and 2 months at the beginning of the assessment period enrolled in an educational institution and attending grade 7 or higher had a chance to be included in this study (OECD, 2007). To ensure that sample covered the population, a two-stage sampling design was implemented. In Turkey, first, 160 schools from seven geographic regions were sampled based on the probability proportional to size (PPS) where the measure of size was a function of the estimated number of eligible 15-year-old students, and type of the schools in which they were enrolled. With schools selected, lists of 15-year-old students at each school were prepared. From them 35 students were sampled with equal probability; if a school had fewer than 35 students, all students participated in PISA (OECD, 2002). The Turkish sample had 4942 students (2290 females, 2652 males) in 160 schools. Table 3.1 provides information about sample characteristics.

<table>
<thead>
<tr>
<th>School Type</th>
<th>Number of schools</th>
<th>Number of students</th>
<th>Student Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anatolian high school</td>
<td>16</td>
<td>547</td>
<td>11.1</td>
</tr>
<tr>
<td>Anatolian vocational high School</td>
<td>5</td>
<td>170</td>
<td>3.4</td>
</tr>
<tr>
<td>General high school</td>
<td>70</td>
<td>2224</td>
<td>45.4</td>
</tr>
<tr>
<td>Multi-program school</td>
<td>2</td>
<td>67</td>
<td>1.4</td>
</tr>
<tr>
<td>Primary school</td>
<td>10</td>
<td>116</td>
<td>2.4</td>
</tr>
<tr>
<td>Science high school</td>
<td>1</td>
<td>35</td>
<td>.7</td>
</tr>
<tr>
<td>Secondary and vocational high school</td>
<td>10</td>
<td>278</td>
<td>5.6</td>
</tr>
<tr>
<td>Vocational high school</td>
<td>46</td>
<td>1485</td>
<td>30.1</td>
</tr>
<tr>
<td>Total</td>
<td>160</td>
<td>4942</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.1: School Types and Number of Students Attending
Instruments and Data Collection

PISA 2006 project was initiated in December 2002 with the development of the project framework by the OECD PISA Governing Board (PGB) which consisted of participating country representatives at senior policy levels. Via the help of international experts in assessment areas the PGB decided on priorities and standards for developing literacy indicators, establishing assessment instruments, and reporting results.

The design and implementation of PISA data collections has been the responsibility of an international consortium led by the Australian Council for Educational Research (ACER). The consortium refined the framework, developed assessment items and context questionnaires with the help of subject matter experts from involved countries. It also collaborated with local education departments in conducting field trials, and translation procedures.

Correct translation of the instruments is a major challenge in cross-national efforts (Van de Vijver & Hambleton, 1996). PISA studies employ double translation, which is considered to be superior to back translation in decreasing the likelihood of linguistic and cultural bias. In this technique the text is translated from the source language (English or French) into targeted languages, back into the source language, and then again into the translated language. The two translated versions are then compared to ensure appropriate translation.

Data collection took place between March 2006 and December 2006 (OECD, 2009). In Turkey, this occurred between May, 2 and May, 31 (MOE, 2006). The focus was students’ cognitive scientific literacy scores and two contextual survey instruments completed by students and school administrators.
Cognitive Test Questionnaires

Based on the PISA framework, a pool of assessment items measuring students’ knowledge of science and knowledge about science were developed and piloted in a field trial in all participating countries. Each item was rated for a) potential cultural, gender or other bias; b) relevance to 15-year-olds in school and non-school contexts; and c) familiarity and level of interest. Test developers and expert groups made the final selection decision in October 2005.

The Scientific literacy part included 37 units with 108 test items (open and closed constructed response, and multiple-choice) and 32 attitude questions. Minor assessment areas, mathematics literacy and reading literacy had fewer items. Each assessment area was divided into clusters and distributed into thirteen booklets. A total of 7 science, 4 mathematics, and 2 reading clusters appeared in the booklets. Each booklet had 4 clusters of questions. While science was assessed in all of them, reading and mathematics were assessed in selected ones. Booklets were randomly assigned to students who were given 120 minutes to complete the test. Table 3.2 contains the distribution of 2006 assessment items according to the various dimensions of the PISA frameworks.
Table 3.2: Distribution of Items by the Dimensions of the PISA-2006 Science Framework

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Number of multiple-choice items</th>
<th>Number of constructed response items</th>
<th>Item total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content area</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge of science &quot;Physical systems&quot;</td>
<td>11</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>Knowledge of science &quot;Living systems&quot;</td>
<td>16</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>Knowledge of science &quot;Earth and space&quot;</td>
<td>7</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Knowledge of science &quot;Technology systems&quot;</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Knowledge about science &quot;Scientific enquiry&quot;</td>
<td>19</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>Knowledge about science &quot;Scientific explanations&quot;</td>
<td>9</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>41</td>
<td>108</td>
</tr>
<tr>
<td><strong>Competency cluster</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identifying scientific issues</td>
<td>19</td>
<td>5</td>
<td>24</td>
</tr>
<tr>
<td>Explaining phenomena scientifically</td>
<td>33</td>
<td>20</td>
<td>53</td>
</tr>
<tr>
<td>Using scientific evidence</td>
<td>15</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>67</td>
<td>41</td>
<td>108</td>
</tr>
</tbody>
</table>

An extension of the Rasch modeling, Mixed-coefficients Multinomial Logit Model (MCMLM) and Plausible Values (PV) methodology is used in PISA to estimate item and student parameters. As a categorical response model, MCMLM models the responses to items through a logistic regression where the predictors are the difficulty of the items and ability of students. MCMLM assumes that student ability is random and unknown parameters are fixed. However, the PISA version of the MCMLM uses extra information such as student background information and their answers to mathematics and reading literacy questions to model student ability level at scientific literacy domain.
To reduce the error in student ability estimates (5 PVs), random values for each student, are drawn from an estimated ability distribution of students with similar background and response model (OECD, 2009; Adams & Wu, 2007).

**Student and Administrator Surveys**

PISA 2006 surveys consisted of numerous items to collect information about students and schools. A 30-minute student survey asking for information on student characteristics, perceptions of science and school, and family background was given after the literacy assessment. In addition, a 20-minute survey about school characteristics and perceptions of school principals were answered by the administrators. Although a parent survey was used in 16 countries, including Turkey, it was not incorporated into the current study because much of the information it provides is on the student questionnaire.

**Variables**

This study looked at 15 student-level and 10 school-level variables. While some of them, such as student gender and grade were derived from answers to individual questions, many of the variables like science self efficacy are latent constructs, and measured by more than one question. Dichotomous and Polytomous Item Response Theory (IRT) scaling techniques were used to estimate the scale scores for these measures. Each student and school-level variable included in the present study and how they were obtained are explained below. Note that answer options for the items are provided in parentheses.
Student-level Variables

A) Background Characteristics

Gender (GENDER). Based on students’ answers to the question on gender (Female, Male).

Grade (GRADE). Students’ answers to the question on their grade were used for this variable.

Economical, social and cultural status (ESCS). This index variable was calculated from many questions about occupational status (job type and title) and educational level (years of education) of parents, approximate number of books at home, and availability of home possessions such as computers, dishwashers, etc.

B) Affective Factors

Perceptions of the general value of science (GENSCIE). This was a derived variable coming from students’ answers to five statements: “How much do you agree with the statements below? a) Advances in science and technology usually improve people’s life, b) Science is important helping us to understand the natural world, c) Advances in science and technology usually improve the economy, d) Science is valuable to society, e) Advances in science and technology usually bring social benefits.” (Strongly agree, Agree, Disagree, Strongly disagree).

Perceptions of the personal value of science (PERSCIE). Another five statements were used here: “How much do you agree with the statements below? a) Some concepts in science help me see how I relate to other people, b) I will use science in many ways when I am an adult, c) Science is very relevant to me, d) I find that science helps me to understand the things around me, e) When I leave school there will be many
opportunities for me to use science.” (Strongly agree, Agree, Disagree, Strongly disagree).

*Instrumental motivation (INSTSCIE)*. Five Likert type items measuring this construct were included in the questionnaire: “How much do you agree with the statements below? a) Making an effort in my science subject(s) is worth it because this will help me in the work I want to do later, b) What I learn in my science subject(s) is important for me because I need this for what I want to study later on, c) I study science because I know it is useful for me, d) Studying my science subject(s) is worthwhile for me because what I learn will improve my career prospects, e) I will learn many things in my science subject(s) that will help me get a job.” (Strongly agree, Agree, Disagree, Strongly disagree).

*Interest in science learning (INTSCIE)*. Students’ answers to eight Likert type items contributed to this variable: “How much interest do you have in learning about the following science topics: a) Topics in physics, b) Topics in chemistry, c) The biology of plants d) Human biology, e) Topics in astronomy, f) Topics in geology, g) Ways scientists design experiments, and h) What is required for scientific explanations”. (High Interest, Medium Interest, Low Interest, No Interest).

*Responsibility for sustainable development (RESPDEV)*. Seven items comprised this measure: “How much do you agree with the statements below? a) It is important to carry out regular checks on the emissions from cars as a condition of their use, b) It disturbs me when energy is wasted through the unnecessary use of electrical appliances, c) I am in favor of having laws that regulate factory emissions even if this would increase the price of products, d) To reduce waste, the use of plastic packing should be kept to a
minimum, e) Industries should be required to prove that they safely dispose of dangerous waste materials, f) I am in favor of having laws that protect the habitats of endangered species, g) electricity should be produced from renewable sources as much as possible, even if this increases the cost.” (Strongly agree, Agree, Disagree, Strongly disagree).

Science self-efficacy (SCIEEFF). Eight items measuring students’ confidence in performing science-related tasks made up this index: “How easy do you think it would be for you to perform the following tasks on your own? a) Recognize the science question that underlies a newspaper report on a health issue, b) Explain why earthquakes occur more frequently in some areas than in others, c) Describe the role of antibiotics in the treatment of disease, d) Identify the science question associated with the disposal of garbage, e) Predict how changes to an environment will affect the survival of certain species, f) Interpret the scientific information provided on the labeling of food items, g) Discuss how new evidence can lead you to change your understanding about the possibility of life on Mars, h) Identify the better of two explanations for the information of acid rain.” (I could do this easily, I could do this with a bit of effort, I would struggle to do this on my own, I couldn’t do this).

C) Teaching and Learning Factors

In-school time spent on science learning (INTIME). Students were asked to indicate how much time they typically spend per week on regular lessons in science at schools (No time, Less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, 6 or more hours a week).

Hands on activities in science class (SCHANDS). Four items measuring this latent variable were included in the student questionnaire: “When learning science topics at
school, how often do the following activities occur? A) Students spend time in the laboratory doing practical experiments, b) Students are required to design how a science question could be investigated in the laboratory, c) Students are asked to draw conclusions from an experiment they have conducted, d) Students do experiments by following the instructions of the teacher.” (In all lessons, In most lessons, In some lessons, Never or hardly ever).

*Interactive teaching in science class (SCINTACT).* Four statements were rated by the students for this variable: “When learning science topics at school, how often do the following occur? a) Students are given opportunities to explain their ideas, b) The lessons involve students’ opinions about the topics, c) There is class debate or discussion, d) The students have discussion about the topics.” (In all lessons, In most lessons, In some lessons, Never or hardly ever).

*Student investigations (SCINVEST).* Three statements were rated by the students to measure the frequency of interactive teaching: “When learning science topics at school, how often do the following occur? a) Students are allowed to design their own experiments, b) Students are given the chance to choose their own investigations, c) Students are asked to do an investigation to test out their own ideas”. (In all lessons, In most lessons, In some lessons, Never or hardly ever).

**D) Science Related out of School Activities**

*Out of school time spent on studying science (OUTIME).* Students were asked how much time out of school they typically spend per week to study science and do homework (No time, Less than 2 hours a week, 2 or more but less than 4 hours a week, 4 or more but less than 6 hours a week, 6 or more hours a week).
Science related activities (SCIEACT). Six items were included to measure this construct: “How often do you do these things? a) Watch TV programs about science, b) Borrow or buy books on science topics, c) Visit web sites about science topics, d) Listen to radio programs about advances in science, e) Read science magazines or science articles in newspapers, f) Attend a science club”. (Very often, Regularly, Sometimes, Never or hardly ever). The items were reverse coded for IRT scaling when computing the entire index variables stated above. So that positive scores on those indices indicate higher levels on the variables.

School-level Variables

A) School Resources

Quality of educational resources (SCMATEDU). This index was computed from seven items measuring the school principal’s perceptions of potential factors hindering instruction at school: Is your school’s capacity to provide instruction hindered by any of the following? a) Shortage or inadequacy of science laboratory equipment, b) Shortage or inadequacy of instructional materials (e.g. textbooks), c) Shortage or inadequacy of computers for instruction, d) Lack or inadequacy of computer software for instruction, e) Shortage or inadequacy of library materials, f) Shortage or inadequacy of audio-visual resources”. (Not at all, Very little, to some extent, A lot). The items were reverse coded on this index as it was noted before.

Student-teacher ratio (STRATIO). This variable was obtained by dividing the school size by the total number of teachers with part-time teachers being weighted 0.5 and full-time teachers 1.0.
Science teacher shortage (STSHORT). School administrators were asked to indicate if their schools’ capacity to provide instruction is hindered by a lack of qualified science teachers (Not at all, Very little, To some extent, A lot).

B) School Context

School activities to promote science learning (SCIPROM). Administrators reported on activities to promote students’ learning of science at their school: a) Science clubs, b) Science fairs, c) Science competitions, d) Extracurricular science projects (including research), e) Excursions and field trips. (Yes=1, No=0)

School sector (SECTOR). Although eight different types of schools are included in the sample, two of them, Anatolian high schools and Science high schools are highly selective ones. They accept the brightest students in the country as determined by a centralized entrance exam. Those schools were considered as one group in this study. A total of 17 schools (1 science and 16 Anatolian) with 582 students fall into this category.

School mean of economic, social and cultural status (MESCS). This variable is mean of the ECSC variable for each school.

School mean of in-school time spent on science learning (MINTIME). This variable is mean of the INTIME variable for each school.

School mean of Hands on activities in science class (MSCHANDS). This variable is mean of the SCHANDS variable for each school.

School mean of Interactive teaching in science class (MSCINTACT). This variable is mean of the SCINTACT variable for each school.

School mean of Student investigations (MSCINVEST). This variable is mean of the SCINVEST variable for each school.
In addition to the explanatory variables mentioned above, current study also employed student and school weight variables, $W_{\text{FSTUWT}}$ and $W_{\text{FSCHWT}}$ respectively. Although the students included in the PISA sample were chosen randomly, the selection probabilities vary. That is, students and schools represent many other students and schools that have similar characteristics in the sampling frame. Weights are incorporated into the analysis to ensure that each sampled student and school represent the correct number of students and schools in the population. The predictive variables included in this study are included in table 3.3 below.
### Student-level Variables (Level 1)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Variable</th>
<th>Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student background</td>
<td>Gender</td>
<td>GENDER</td>
</tr>
<tr>
<td></td>
<td>Grade</td>
<td>GRADE</td>
</tr>
<tr>
<td></td>
<td>Economic, social and cultural status</td>
<td>ESCS</td>
</tr>
<tr>
<td>Affective factors</td>
<td>Perceptions of the general value of science</td>
<td>GENSCIE</td>
</tr>
<tr>
<td></td>
<td>Perceptions of the personal value of science</td>
<td>PERSCIE</td>
</tr>
<tr>
<td></td>
<td>Instrumental motivation</td>
<td>INSTSCIE</td>
</tr>
<tr>
<td></td>
<td>Interest in science learning</td>
<td>INTSCIE</td>
</tr>
<tr>
<td></td>
<td>Responsibility for sustainable development</td>
<td>RESPDEV</td>
</tr>
<tr>
<td></td>
<td>Science self efficacy</td>
<td>SCIEEFF</td>
</tr>
<tr>
<td>Teaching and learning</td>
<td>In-school time spent on science learning</td>
<td>INTIME</td>
</tr>
<tr>
<td></td>
<td>Hands on activities in science classes</td>
<td>SCHANDS</td>
</tr>
<tr>
<td></td>
<td>Interactive teaching in science classes</td>
<td>SCINTACT</td>
</tr>
<tr>
<td></td>
<td>Student investigations</td>
<td>SCINVEST</td>
</tr>
<tr>
<td>Science related out of school activities</td>
<td>Out of school time spent on studying science</td>
<td>OUTIME</td>
</tr>
<tr>
<td></td>
<td>Science-related out-of school activities</td>
<td>SCIEACT</td>
</tr>
</tbody>
</table>

### School-level Variables (Level 2)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Variable</th>
<th>Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>School resources</td>
<td>Quality of educational resources</td>
<td>SCMATEDU</td>
</tr>
<tr>
<td></td>
<td>Student-teacher ratio</td>
<td>STRATIO</td>
</tr>
<tr>
<td></td>
<td>Science teacher shortage</td>
<td>STSHORT</td>
</tr>
<tr>
<td>School context</td>
<td>School activities to promote science learning</td>
<td>SCIPROM</td>
</tr>
<tr>
<td></td>
<td>School sector</td>
<td>SECTOR</td>
</tr>
<tr>
<td></td>
<td>School mean of economic, social and cultural status</td>
<td>MESCS</td>
</tr>
<tr>
<td></td>
<td>School mean of in-school time spent on science learning</td>
<td>MINTIME</td>
</tr>
</tbody>
</table>

Table 3.3: Independent Variables Included in the Present Study

(Continued)
Table 3.3 (Continued)

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Variable</th>
<th>Variable Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>School context</td>
<td>School mean of hands on activities in science class</td>
<td>MSCHANDS</td>
</tr>
<tr>
<td></td>
<td>School mean of interactive teaching in science class</td>
<td>MSCINTACT</td>
</tr>
<tr>
<td></td>
<td>School mean of student investigations</td>
<td>MSCINVEST</td>
</tr>
<tr>
<td></td>
<td>School mean of hands on activities in science class</td>
<td>MSCHANDS</td>
</tr>
</tbody>
</table>

Validity and Reliability of the Measures

PISA 2006 is one of the biggest international student assessments. Validity and reliability of the constructs are determined by international experts. In order to develop valid and comparable latent measures Exploratory and Confirmatory Factor Analysis (EFA, CFA) were employed by PISA data analysis teams to validate indices of Likert type items. When dichotomous items are included in the validation analyses (i.e., validating ESCS scale), weighted least squares estimation with polychoric correlations was used (OECD, 2005, 2007). Validation of latent constructs was checked based on different fit indices such as root-mean square error of approximation (RMSEA), and comparative fit index (CFI). All variables in this study meet construct validity requirements.

To ensure the reliability of the coding of responses to open-ended items, the PISA consortium prepared guidelines which consisted of coding manuals, training materials for
recruiting coders, and workshop materials to train them. Based on the train the trainer approach facilitators from participating countries were taught by PISA consortium members and then the former provided training to the coders in their countries. After all the open-ended questions were coded, a sub-sample of assessment booklets from each country was examined by the PISA Consortium in order to inspect the consistency of the coding process. The inspection process included two analyses:

*Inter-coder Reliability Study on the Sub-sample of Assessment Booklets*

For investigating the inter-coder agreement, a homogeneity index, by item and country was computed. Homogeneity index coefficient theoretically can range from zero to one; one meaning perfect agreement between coders (OECD, 2002). Country mean index scores indicated that all science assessment items, except one had an index score above .8 which is high inter-rater agreement. Item index scores in Turkish sample ranged between .83 and 1; median coder agreement score was .98 (OECD, 2009).

*Multiple Marking (Variance Components) Study*

A selection of open-ended item clusters from 600 students was coded by same four coders and an estimate of the between-coder variability calculated within each country. Variability in item codes were partitioned into student, item, rater components and interaction among them. The results of the study indicated that in Turkish sample less than 1 percent of the variability in item ratings was explained by variability between raters’, suggesting no significant bias in coding of open-ended items.

To make sure the coding consistency at the between-country level, a coding review was implemented. First, a random sample of coded responses selected and re-coded by an independent multilingual reviewer. Then, a statistical analysis was
conducted to test if there was a difference between the scores reported by countries and the independent reviewer. When that difference was statistically significant for an item, it was identified as being potentially discrepant. At the third stage a senior member of the Consortium who is the leader of the related domain reviewed a random sample of 19 student responses coded by a country team and an independent reviewer and decided the final codes. When the difference between the final codes and rater codes from a country is statistically significant, it is reflected in the students’ final PVs. After ensuring the reliability of the open-response item coding, reliability of the science scale as a whole is obtained at international level as \( r = .84 \). Each construct used in this study has also shown to be reliable. Table 3.4 includes the reliability values of the latent variables. Those values were obtained using only the Turkish sample as provided by PISA 2006 technical report (OECD, 2007).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Reliability*</th>
<th>Variable</th>
<th>Reliability*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCS</td>
<td>.72</td>
<td>SCHANDS</td>
<td>.81</td>
</tr>
<tr>
<td>GENSCIE</td>
<td>.79</td>
<td>SCINTACT</td>
<td>.76</td>
</tr>
<tr>
<td>PERSCIE</td>
<td>.81</td>
<td>SCINVEST</td>
<td>.76</td>
</tr>
<tr>
<td>INSTSCIE</td>
<td>.91</td>
<td>SCIEACT</td>
<td>.82</td>
</tr>
<tr>
<td>INTSCIE</td>
<td>.83</td>
<td>SCMATEDU</td>
<td>.85</td>
</tr>
<tr>
<td>RESPDEV</td>
<td>.84</td>
<td>SCIPROM</td>
<td>.67</td>
</tr>
<tr>
<td>SCIEEFF</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Chronbach’s alpha

Table 3.4: Reliability Values of Latent Variables.
Data Analysis

The statistical software HLM developed by Raudenbush, Bryk and Congdon (2004) was utilized to answer the study questions. HLM has the capability of attending PISA’s complex sampling design with five plausible values. The hierarchical linear models utilized in this study are explained below.

To answer the first research question, *Are there differences in the mean achievement scores of scientific literacy among Turkish schools?*, the amount of between-school variance in scientific literacy will be investigated by partitioning the total variance in scientific literacy into it’s within-group and between-group components. This is conceptually equivalent to a One-way Analysis of Variance (ANOVA). The following equations will be estimated to respond the first research question:

Level 1: \( y_{ij} = \beta_{0j} + r_{ij} \)

Level 2: \( \beta_{0j} = \gamma_{00} + u_{0j} \)

Where

- \( y_{ij} \) is the dependent variable (scientific literacy) for each student at a school
- \( \beta_{0j} \) is the intercept or mean achievement of jth school
- \( r_{ij} \) is unique effect associated with student I in school j, or level-1 error
- \( \gamma_{00} \) is the average of the school means, or grand mean scientific literacy achievement
- \( u_{0j} \) is the random effect associated with school j.

To answer the research question 2, *Which student level measures explain the differences (if any) in Turkish students’ achievement of scientific literacy?*, the random
coefficients model will be estimated. In this model, the differences in the student achievement scores are estimated by the level-1 independent variables as follows:

**Level 1:**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(GENDER) + \beta_{2j}(GRADE) + \beta_{3j}(ESCS) + \beta_{4j}(GENSCIE) + \beta_{5j}(PERSCIE) + \beta_{6j}(INSTSCIE) + \beta_{7j}(INTSCIE) + \beta_{8j}(RESPDEV) + \beta_{9j}(SCIEFF) + \beta_{10j}(INTIME) + \beta_{11j}(SCHANDS) + \beta_{12j}(SCINTACT) + \beta_{13j}(SCINVEST) + \beta_{14j}(OUTIME) + \beta_{15j}(SCIEACT) + r_{ij} \]

**Level 2:**

\[ \beta_{0j} = \gamma_{00} + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + u_{1j} \]
\[ \beta_{2j} = \gamma_{20} + u_{2j} \]
\[ \beta_{3j} = \gamma_{30} + u_{3j} \]
\[ \beta_{4j} = \gamma_{40} + u_{4j} \]
\[ \beta_{5j} = \gamma_{50} + u_{5j} \]
\[ \beta_{6j} = \gamma_{60} + u_{6j} \]
\[ \beta_{7j} = \gamma_{70} + u_{7j} \]
\[ \beta_{8j} = \gamma_{80} + u_{8j} \]
\[ \beta_{9j} = \gamma_{90} + u_{9j} \]
\[ \beta_{10j} = \gamma_{100} + u_{10j} \]
\[ \beta_{11j} = \gamma_{110} + u_{11j} \]
\[ \beta_{12j} = \gamma_{120} + u_{12j} \]
\[ \beta_{13j} = \gamma_{130} + u_{13j} \]
\[ \beta_{14j} = \gamma_{140} + u_{14j} \]
\[ \beta_{15j} = \gamma_{150} + u_{15j} \]

where,

\( \beta_{0j} \) is the mean scientific literacy achievement in school j

\( \beta_{1j} \) is the mean difference between the scientific literacy achievement of male and female students

\( \beta_{2j} \) is the differentiating effect of grade in school j

\( \beta_{3j} \) is the differentiating effect of economic, social and cultural status in school j
\( \beta_{s,j} \) is the differentiating effect of perceptions of the general value of science in school \( j \).

\( \beta_{p,j} \) is the differentiating effect of perceptions of the personal value of science in school \( j \).

\( \beta_{e,j} \) is the differentiating effect of instrumental motivation in school \( j \).

\( \beta_{i,j} \) is the differentiating effect of interest in science learning in school \( j \).

\( \beta_{r,j} \) is the differentiating effect of responsibility for sustainable development in school \( j \).

\( \beta_{s,j} \) is the differentiating effect of science self-efficacy in school \( j \).

\( \beta_{t,0,j} \) is the differentiating effect of in-school time spent on science learning in school \( j \).

\( \beta_{t,1,j} \) is the differentiating effect of hands on activities in science classes in school \( j \).

\( \beta_{t,2,j} \) is the differentiating effect of interactive teaching in science classes in school \( j \).

\( \beta_{t,3,j} \) is the differentiating effect of student investigations in school \( j \).

\( \beta_{t,4,j} \) is the differentiating effect of out-of-school time spent on studying science in school \( j \).

\( \beta_{t,5,j} \) is the differentiating effect of science-related out-of-school activities in school \( j \).

To answer the final research question, *Which school level measures are associated with differences in mean school achievement?*, two models will be estimated.

The first one is means as outcomes model including only student-level (level 1) variables. After determining the significant school-level predictors, intercept as outcome model,
which combines means as outcomes and random coefficients models, will be developed.

Below is the equation for means as outcomes model.

Level 1: \( Y_{ij} = \beta_{0j} + r_{ij} \)

Level 2: \( \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SCMATEDU}) + \gamma_{02}(\text{STRATIO}) + \gamma_{03}(\text{STSHORT}) + \gamma_{04}(\text{SCIPROM}) + \gamma_{05}(\text{SECTOR}) + \gamma_{06}(\text{MESCS}) + \gamma_{07}(\text{MINTIME}) + \gamma_{08}(\text{MSCHAND}) + \gamma_{09}(\text{MSCINTACT}) + \gamma_{10}(\text{MSCINVEST}) + u_{0j} \)

where,

\( \beta_{0j} \) is the mean scientific literacy achievement for school \( j \)

\( \gamma_{00} \) is the grand mean outcome in the population

\( \gamma_{01} \) is the differentiating effect (slope) of quality of educational resources on the school mean scientific literacy achievement

\( \gamma_{02} \) is the differentiating effect (slope) of student-teacher ratio on the school mean scientific literacy achievement

\( \gamma_{03} \) is the differentiating effect (slope) of teacher shortage on the school mean scientific literacy achievement

\( \gamma_{04} \) is the differentiating effect (slope) of quality of school activities to promote science learning on the school mean scientific literacy achievement

\( \gamma_{05} \) is the differentiating effect (slope) of school sector on the school mean scientific literacy achievement

\( \gamma_{06} \) is the differentiating effect (slope) of mean of school economic, social, and cultural status on the school mean scientific literacy achievement.
\( \gamma_{07} \) is the differentiating effect (slope) of the mean of in-school time spent on science learning.

\( \gamma_{08} \) is the differentiating effect (slope) of the school mean of hands on activities in science class.

\( \gamma_{09} \) is the differentiating effect (slope) of the school mean of interactive teaching in science class.

\( \gamma_{10} \) is the differentiating effect (slope) of the school mean of student investigations.

The intercepts as outcome model will allow determining if any of the school-level variables can predict the school means controlling for the student-level-variables. That model will include only significant predictors based on the previous ones. Therefore, intercept as outcome model’s equation will be developed in chapter four.
CHAPTER 4

RESULTS

Introduction

This study represents an investigation of the effects of student and school-level factors on 15-year old Turkish students’ scientific literacy achievement. Publicly available PISA-2006 database was employed to answer the three research questions in Chapter 3. This chapter is divided into two parts. The first provides information about data characteristics including descriptive statistics, replacement of missing values, and bivariate correlations among the variables. The second part includes HLM analyses results to answer the research questions. The analysis of variance (ANOVA) model is employed to explore the differences in the mean achievement scores of scientific literacy among the schools and to determine the amount of variation occurring at student- and school- levels. Then, the random coefficient model is used to investigate the relationships between the student-level (level 1) variables and scientific literacy scores. Finally, the means as outcomes and intercepts as outcomes models are utilized to explore the variables associated with between-school (level 2) variation.
Data Characteristics

As indicated previously, the data set used for this study included 4942 15-year old students coming from 160 schools. Prior research suggested that when doing multilevel analysis at least thirty clusters with ten or more observations in each are necessary to obtain sound estimates of standard errors (Porter, 2005; Raudenbush & Bryk, 2002). Thus, five schools having less than ten observations were removed from the data before doing analyses. The final sample included 2282 female (46.5%) and 2630 male (53.5%) students coming from 17 highly selective (11%) and 138 not highly selective (89%) schools.

The number of students varied from 10 to 35 with an average of 31.7 students per school. Among the 4912 students included in the analyses, 582 (11.8%) attended highly selective (Anatolian or Science) high schools. Turkish students participating in PISA 2006 had a mean scientific literacy achievement score of 428.18 with a standard deviation of 82.4. Compared to OECD mean of 500 with a standard deviation of 100, these figures represent lower achievement (EARGED, 2007; OECD, 2007) and less variability.

All student-level predictor variables except for economic, social and cultural status (ESCS) and GRADE had missing values. Two, hands on activities in science classes (SCHANDS), and in-school time spent on science learning (INTIME) had more than 1% missing (1.5% and 2.28% respectively). Considering that there was not a pattern in missing responses by school and gender, it was assumed that the missing values at the student-level were completely random. Therefore list-wise deletion –dropping a case with missing data on any variable– was chosen when doing statistical analyses. Four school-
level variables, quality of educational resources (SCMATEDU), student-teacher ratio (STRATIO), science teacher shortage (STSHORT) and school activities to promote science learning (SCIPROM) had missing values. The highest rate of the missing data at that level was 2.58. Because the amount of missing data was not substantial, and excluding such schools would have resulted in eliminating a portion of the students from the subsequent analyses, the mean replacement method was employed at the school-level.

Descriptive statistics of the variables used in this study are provided in Table 4.1, and 4.2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S. D.</th>
<th>Min.</th>
<th>Max.</th>
<th>% Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACHIEVEMENT</td>
<td>428.18</td>
<td>82.4</td>
<td>115.57*</td>
<td>733.06*</td>
<td>0</td>
</tr>
<tr>
<td>GENDER</td>
<td>.54</td>
<td>.5</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>GRADE</td>
<td>9.58</td>
<td>.6</td>
<td>7</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>ESCS</td>
<td>-1.23</td>
<td>1.08</td>
<td>-4.42</td>
<td>2</td>
<td>.16</td>
</tr>
<tr>
<td>GENSCIE</td>
<td>.52</td>
<td>1.07</td>
<td>-3.66</td>
<td>2.19</td>
<td>.29</td>
</tr>
<tr>
<td>PERSCIE</td>
<td>.35</td>
<td>1.02</td>
<td>-3.08</td>
<td>2.53</td>
<td>.36</td>
</tr>
<tr>
<td>INSTSCIE</td>
<td>.36</td>
<td>.97</td>
<td>-2.1</td>
<td>1.82</td>
<td>2</td>
</tr>
<tr>
<td>INTSCIE</td>
<td>.26</td>
<td>.95</td>
<td>-3.14</td>
<td>3.29</td>
<td>.26</td>
</tr>
<tr>
<td>RESPDEV</td>
<td>.78</td>
<td>1.08</td>
<td>-4</td>
<td>2.3</td>
<td>.28</td>
</tr>
<tr>
<td>SCIEFF</td>
<td>.05</td>
<td>.99</td>
<td>-3.77</td>
<td>3.22</td>
<td>.3</td>
</tr>
<tr>
<td>INTIME</td>
<td>2.85</td>
<td>1.45</td>
<td>1</td>
<td>5</td>
<td>2.28</td>
</tr>
<tr>
<td>SCHANDS</td>
<td>.04</td>
<td>1.12</td>
<td>-2.1</td>
<td>2.91</td>
<td>1.5</td>
</tr>
<tr>
<td>SCINTACT</td>
<td>.45</td>
<td>.92</td>
<td>-2.5</td>
<td>2.47</td>
<td>1</td>
</tr>
<tr>
<td>SCINVEST</td>
<td>.8</td>
<td>1</td>
<td>-1.26</td>
<td>3.03</td>
<td>2</td>
</tr>
<tr>
<td>OUTIME</td>
<td>1.97</td>
<td>1.14</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>SCIEACT</td>
<td>.59</td>
<td>.91</td>
<td>-1.69</td>
<td>3.38</td>
<td>.2</td>
</tr>
</tbody>
</table>

* Minimum and maximum values show the most extreme estimates among the plausible values.

Table 4.1: Descriptive Statistics of Student-level Variables
As seen in Table 4.1, Turkish students’ responses to index variables had a mean slightly higher than the international average of 0 except for the ESCS variable. That is, Turkish sample had better attitudes toward science than the international test takers; also the former perceived that teaching and learning activities in classrooms involved more hands-on activities, interaction, and student investigations than the international average. On the other hand, Turkish students’ socio-economic and cultural status was more than one standard deviation below the international average.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>S. D.</th>
<th>Min.</th>
<th>Max.</th>
<th>% Miss</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCMATEDU</td>
<td>-.79</td>
<td>.99</td>
<td>-3.43</td>
<td>2.14</td>
<td>1.94</td>
</tr>
<tr>
<td>STRATIO</td>
<td>18.7</td>
<td>8.31</td>
<td>1.89</td>
<td>48.33</td>
<td>2.58</td>
</tr>
<tr>
<td>STSHORT</td>
<td>2.69</td>
<td>.99</td>
<td>1</td>
<td>4</td>
<td>2.58</td>
</tr>
<tr>
<td>SCIPROM</td>
<td>-.16</td>
<td>1.14</td>
<td>-2.27</td>
<td>1.64</td>
<td>.65</td>
</tr>
<tr>
<td>SECTOR</td>
<td>.11</td>
<td>.31</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>MESCS</td>
<td>-1.3</td>
<td>.64</td>
<td>-2.81</td>
<td>.75</td>
<td>0</td>
</tr>
<tr>
<td>MINTIME</td>
<td>2.82</td>
<td>.73</td>
<td>1.41</td>
<td>4.89</td>
<td>0</td>
</tr>
<tr>
<td>MSCHANDS</td>
<td>.05</td>
<td>.43</td>
<td>-1.05</td>
<td>.85</td>
<td>0</td>
</tr>
<tr>
<td>MSCINTACT</td>
<td>.46</td>
<td>.24</td>
<td>-.19</td>
<td>1.6</td>
<td>0</td>
</tr>
<tr>
<td>MSCINVEST</td>
<td>.82</td>
<td>.32</td>
<td>-.01</td>
<td>1.66</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.2: Descriptive Statistics of School-level Variables

Analogously, in Table 4.2 the school means of teaching and learning activity indicators had higher figures. The mean economic, social and cultural status index (MESCS) was well below the international average. School principals noted that student-teacher ratio (STRATIO) average 18.7, and the quality of education was affected by the shortage of science teachers (STSHORT). They also reported that quality of educational
resources (SCMATEDU) was well below the international standard. Note that the mean
for the dummy variable GENDER in Table 1 represents the percentage of males, and the
mean for the SECTOR in Table 2 shows the percentage of highly selective schools in the
dataset.

Inter-correlations between Variables

Preliminary bivariate relationships among student and school-level variables were
determined using SPSS (V.17) and are shown in Tables 4.3 and 4.4. Note that when zero-
order correlations between achievement scores and independent variables were
calculated, five bivariate correlations for plausible values and predictors were obtained
and the average of them were used to reflect the most unbiased estimates. Based on inter-
correlations (Pearson’s correlation and Kendall’s tau b) at the lower level all the
predictors except interactive teaching in science class (SCINTACT) (r = 0) had
significant relationships with scientific literacy scores. Because no correlation between
SCINTACT and achievement was observed, this variable was not included in the
multilevel modeling process. Other variables were used as predictors of scientific literacy
with the magnitudes of associations ranging from weak to moderate. Gender had a very
low correlation (r = -.06) with the dependent variable suggesting that no statistically
significant difference between female and male students’ scores on dependent variable
was expected. Achievement’s relationships between grade and student investigations in
science classes (SCINVEST) were low and negative (r = -.15, and r = -.17 respectively)
meaning, on average, lower graders and students having less hands-on experiences during
the class time received better scores on PISA-2006. Other correlations between achievement scores and independent variables were positive.

Inter-correlations among student-level independent variables ranged from -.17 to .72. While the affective variables were moderately correlated, moderate and high correlations among teaching and learning factors were observed. Since high correlations between variables produce unreliable and unstable estimates (Devaus, 2002; Raudenbush & Bryk, 2004), a multiple regression analysis at the student level was conducted to check multicollinearity. Variance Inflation Factors (VIF) for independent variables indicated no multicollinearity; namely, no VIF values were greater than 10 (relevant statistics are provided in Appendix A).

At the school level most of the significant associations were also low or moderate. The only high correlation was between mean student investigations and mean hands-on experience ($r = .84$). A series of multiple regression analysis using level-2 variables were conducted to detect if there was collinearity between those variables. Based on the results of the $R^2$ estimates it was decided to remove the school mean of the student investigations in science classes (MSCINVEST) from the models to be tested. About 80% of the variation in that variable was explained by the other level-2 predictors included in this study (see Appendix B for the results of the analyses).
<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-SCORE</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>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-GENDER</td>
<td>-.06*</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>
<td></td>
<td></td>
</tr>
<tr>
<td>3-GRADE</td>
<td>-.1**</td>
<td>0</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>
<td></td>
</tr>
<tr>
<td>4-ESCS</td>
<td>.39**</td>
<td>-.06**</td>
<td>-.03*</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>5-GENSCIE</td>
<td>.33**</td>
<td>-.06**</td>
<td>-.06*</td>
<td>.11**</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>6-PERSCIE</td>
<td>.23**</td>
<td>-.03*</td>
<td>-.11**</td>
<td>.1**</td>
<td>.62**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-INSTCIE</td>
<td>.18**</td>
<td>-.04**</td>
<td>-.17**</td>
<td>.03*</td>
<td>.39**</td>
<td>.59**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-INTSCIE</td>
<td>.22**</td>
<td>0</td>
<td>-.13**</td>
<td>.07**</td>
<td>.4**</td>
<td>.57**</td>
<td>.52**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-RESPDEV</td>
<td>.23**</td>
<td>-.05**</td>
<td>.0</td>
<td>.05**</td>
<td>.39**</td>
<td>.3**</td>
<td>.22**</td>
<td>.25**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10-SCIEFF</td>
<td>.36**</td>
<td>-.04*</td>
<td>-.08**</td>
<td>.19**</td>
<td>.4**</td>
<td>.48**</td>
<td>.35**</td>
<td>.5**</td>
<td>.28**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11-INTIME</td>
<td>.42**</td>
<td>0</td>
<td>-.33**</td>
<td>.2**</td>
<td>.22**</td>
<td>.29**</td>
<td>.36**</td>
<td>.26**</td>
<td>.09**</td>
<td>.24**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12-SCHANDS</td>
<td>-.15**</td>
<td>.11**</td>
<td>.01</td>
<td>-.02</td>
<td>.07**</td>
<td>.14**</td>
<td>.21**</td>
<td>.13**</td>
<td>.07**</td>
<td>.1**</td>
<td>-.04**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13-SCINTACT</td>
<td>0</td>
<td>-.05**</td>
<td>-.05**</td>
<td>.01</td>
<td>.2**</td>
<td>.18**</td>
<td>.28**</td>
<td>.19**</td>
<td>.17**</td>
<td>.17**</td>
<td>.05**</td>
<td>.54**</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14-SCINVEST</td>
<td>-.17**</td>
<td>.05**</td>
<td>0</td>
<td>-.04**</td>
<td>.09**</td>
<td>.14**</td>
<td>.25**</td>
<td>.15**</td>
<td>.09**</td>
<td>.11**</td>
<td>-.03*</td>
<td>.72**</td>
<td>.64**</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>15-OUTIME</td>
<td>.29**</td>
<td>-.01</td>
<td>-.2**</td>
<td>.2**</td>
<td>.17**</td>
<td>.32**</td>
<td>.35**</td>
<td>.3**</td>
<td>.04**</td>
<td>.23**</td>
<td>.5**</td>
<td>.01</td>
<td>.04**</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>16-SCIEACT</td>
<td>.21**</td>
<td>.08**</td>
<td>-.11**</td>
<td>.13**</td>
<td>.35**</td>
<td>.58**</td>
<td>.49**</td>
<td>.6**</td>
<td>.14**</td>
<td>.48**</td>
<td>.26**</td>
<td>.19**</td>
<td>.19**</td>
<td>.18**</td>
<td>.32**</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 4.3: Correlations between Student-level Variables
Table 4.4: Correlations between School-level Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-SCMATEDU</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-STRATIO</td>
<td>-.22**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-STSHORT</td>
<td>-.41** .24**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-SCIPROM</td>
<td>.07</td>
<td>.09</td>
<td>-.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-SECTOR</td>
<td>-.04 -.13*</td>
<td>.04</td>
<td>.22**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6- MESC</td>
<td>.12</td>
<td>-.14</td>
<td>-.05</td>
<td>.32**</td>
<td>.45**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-MINTIME</td>
<td>.01</td>
<td>.04</td>
<td>.14</td>
<td>.4**</td>
<td>.61**</td>
<td>.53**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8-MSCHANDS</td>
<td>-.06</td>
<td>-.07</td>
<td>-.1</td>
<td>.14*</td>
<td>-.19*</td>
<td>-.15</td>
<td>-.22**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9-MSCINTACT</td>
<td>-.1</td>
<td>.04</td>
<td>-.07</td>
<td>-.02</td>
<td>-.05</td>
<td>-.15</td>
<td>.04</td>
<td>.37**</td>
<td></td>
</tr>
<tr>
<td>10-MSCINVEST</td>
<td>-.11</td>
<td>.01</td>
<td>-.06</td>
<td>-.23**</td>
<td>-.32**</td>
<td>-.29**</td>
<td>-.29**</td>
<td>.84**</td>
<td>.57**</td>
</tr>
</tbody>
</table>

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Research Findings

Models to answer the questions established in the previous chapter were tested via HLM 6.0.7 (Raudenbush, Bryk & Congdon, 2004). Before going into the details of the procedures the centering approach chosen for multilevel modeling needs to be explained.

Centering is an important issue in hierarchical model specification. In general, variables in educational survey data do not have a real zero point. Therefore variables are often centered in multivariate statistics and HLM methods to make interpretations easier (De Leeuw, 2004). Two different centering methods are available: grand mean centering and group mean centering. The former is achieved by subtracting the sample mean of all observations from each value. In that case zero represents the mean of the whole sample.
The latter technique is done by subtracting the mean of a cluster from individual observations belong to that cluster; here, zero represents the cluster mean (De Leeuw, 2004; De Leeuw & Meijer, 2008).

Raudenbush and Bryk (2002) stated that group mean centering at level1 helps to reduce the multicollinearity among variables and bias in variance estimates. Considering the significant bivariate correlations among student-level variables, group mean centering was chosen at that level. The researchers also suggested grand mean centering at level-2 to make interpretations easier. So, that method was employed at school-level.

**Research Question 1**

Are there differences in the mean achievement scores of scientific literacy among Turkish schools?

To answer this research question, a null model including no variables at either level was tested. Because this model decomposes the total variance in achievement scores into within-school and between-school components, it is also called as one-way ANOVA model (Raudenbush & Bryk, 2002). Details of it are in Table 4.5 and it shows that there are significant differences among schools in terms of the outcome variable. Grand mean of the literacy achievement is 409.65 with a standard error of 8.7, producing a 95% confidence interval of 409.65 ± 1.96 (8.7) = (392.6, 426.7).
<table>
<thead>
<tr>
<th>Fixed Effect Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average school mean, $\gamma_{00}$</td>
<td>409.65</td>
<td>8.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect Variance Component df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, $u_{0j}$</td>
<td>3906.94</td>
<td>154</td>
</tr>
<tr>
<td>Level-1 effect, $r_{ij}$</td>
<td>3103.4</td>
<td></td>
</tr>
</tbody>
</table>

* $p < .001$.

Table 4.5: One-way ANOVA Model Results

The maximum likelihood estimate of the variance component at student-level ($\sigma^2$) is 3103.4, at the school-level it ($\tau_{oo}$) is 3906.94. A significant $\tau_{oo}$ ($\chi^2 = 7005.84$, df = 154) indicates that there is a considerable amount of variation among schools ($p < .001$), and HLM is appropriate for the dataset. The Intra-class Correlation (ICC), indicator of the proportion of the total variance at school-level (Kreft & De Leeuw, 1998; Raudenbush & Bryk, 2002), is calculated as:

$$ICC = \frac{\tau_{oo}}{(\tau_{oo} + \sigma^2)} = \frac{3906.94}{(3906.94 + 3103.4)} = .557.$$  

Thus, 55.7% of the total variance in scientific literacy achievement occurred between schools and 44.3% was within schools. In addition, an overall reliability estimate of school means ($\beta_{0j}$) is .98 suggesting that the sample means are very reliable estimates of the true school means. Following Raudenbush and Bryk’s (2002) guidelines, variation in achievement scores across schools is estimated by calculating plausible range for school means through the following formula:
\( \gamma_{oo} \pm 1.96 \left( \tau_{oo} \right)^{1/2} = 409.65 \pm 1.96 \left( 3906.94 \right)^{1/2} = (287.14, 532.16). \)

95% confidence interval for school means falls within a wide range between 287.14 and 532.16.

**Research Question 2**

Which student level measures explain the differences (if any) in Turkish students’ achievement of scientific literacy?

Literacy achievement scores were regressed on level 1 variables to determine which level 1 predictors are important in explaining the variation in student scores. Instead of building a saturated model, “where all potential predictors are included with random slopes” (Raudenbush & Bryk, p. 256) and then removing non-significant ones, a step-up strategy was used. It started with background variables, and then affective factors. In the next step, the teaching and learning cluster was introduced, and lastly, out-of-school activities related to science learning were entered. Non-significant predictors were excluded from the analyses as the model grew. Each student-level variable was treated as random assuming their effects are different across schools. The reason for not fixing the slopes of any level 1 variable was that all of them had reliability coefficients calculated by HLM model greater than .1 (Raudenbush, Bryk, Cheong, Congdon & Toit, 2004). Randomly varying level 1 slopes would allow to test if there are differentiated effects of level 1 predictors across schools, and to model those relationships, if they exist. Table 4.6 contains the results of the random coefficient model.
<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mean achievement, $\gamma_{00}$</td>
<td>410.59</td>
<td>8.76</td>
<td>46.83**</td>
</tr>
<tr>
<td>GRADE, $\gamma_{10}$</td>
<td>16.29</td>
<td>4.28</td>
<td>3.81**</td>
</tr>
<tr>
<td>ESCS, $\gamma_{20}$</td>
<td>5.57</td>
<td>2.08</td>
<td>2.68**</td>
</tr>
<tr>
<td>GENSCIE, $\gamma_{30}$</td>
<td>11.57</td>
<td>2.15</td>
<td>5.39**</td>
</tr>
<tr>
<td>PERSCIE, $\gamma_{40}$</td>
<td>-6.28</td>
<td>2.32</td>
<td>-2.71*</td>
</tr>
<tr>
<td>RESPDEV, $\gamma_{50}$</td>
<td>4.94</td>
<td>1.29</td>
<td>3.83**</td>
</tr>
<tr>
<td>SCIEEFF, $\gamma_{60}$</td>
<td>10.41</td>
<td>2.04</td>
<td>5.1**</td>
</tr>
<tr>
<td>INTIME, $\gamma_{70}$</td>
<td>6.94</td>
<td>1.89</td>
<td>3.69**</td>
</tr>
<tr>
<td>SCHANDS, $\gamma_{80}$</td>
<td>-7.55</td>
<td>1.7</td>
<td>-4.43**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df$^1$</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, $\gamma_{0j}$</td>
<td>3872.63</td>
<td>133</td>
<td>7183.78</td>
<td>.00</td>
</tr>
<tr>
<td>GRADE, $\gamma_{1j}$</td>
<td>198.12</td>
<td>133</td>
<td>169.62</td>
<td>.02</td>
</tr>
<tr>
<td>ESCS, $\gamma_{2j}$</td>
<td>38.19</td>
<td>133</td>
<td>311.99</td>
<td>.00</td>
</tr>
<tr>
<td>GENSCIE, $\gamma_{3j}$</td>
<td>33.6</td>
<td>133</td>
<td>143.86</td>
<td>.25</td>
</tr>
<tr>
<td>PERSCIE, $\gamma_{4j}$</td>
<td>58.41</td>
<td>133</td>
<td>220.55</td>
<td>.00</td>
</tr>
<tr>
<td>RESPDEV, $\gamma_{5j}$</td>
<td>19.16</td>
<td>133</td>
<td>164.63</td>
<td>.03</td>
</tr>
<tr>
<td>SCIEEFF, $\gamma_{6j}$</td>
<td>39.76</td>
<td>133</td>
<td>231.7</td>
<td>.00</td>
</tr>
<tr>
<td>INTIME, $\gamma_{7j}$</td>
<td>24.2</td>
<td>133</td>
<td>230.29</td>
<td>.00</td>
</tr>
<tr>
<td>SCHANDS, $\gamma_{8j}$</td>
<td>30.82</td>
<td>133</td>
<td>179.15</td>
<td>.01</td>
</tr>
</tbody>
</table>

Level-1 effect, $r_{ij}$ | 2272.58 |

$^1$ Chi-square statistics are based on only 134 of 155 units that had sufficient data for computation. Fixed effects and variance components are based on all of the data.

* p < .05.

** p < .01.

Table 4.6: Random Coefficient Model Results
Overall mean of scientific literacy achievement was 410.59 with a standard error of 8.76. Eight of the fourteen level 1 variables tested in a series of random coefficients model had significant effect on the outcome variable. The slope of student grade ($\gamma_{10}$) was 16.29 ($t = 3.81$). On average, one year increase in education was associated with about 16-point increase in the test scores when other student-level variables were held constant. Economic, social and cultural status of a student was also an important variable in predicting the literacy score. Its slope ($\gamma_{2j}$) was 5.57 ($t = 2.68$); indicating that about one standard deviation change in that variable was associated more than 5.5 change in the outcome. Four of the variables with significant effect were attitudinal in nature. The slope of student perceptions of the general value of science ($\gamma_{30}$) was 11.57 ($t = 5.4$), the slope of student perceptions of the personal value of science ($\gamma_{40}$) was -6.28 ($t = -2.7$), the slope of responsibility for sustainable development ($\gamma_{50}$) was 4.94 ($t = 3.83$), and the slope of science self-efficacy ($\gamma_{90}$) was 10.42 ($t = 5.1$). Remember that those attitudinal variables, like ESCS, had a mean very close to zero and a standard deviation of 1. Thus, these values represent the point change in scientific literacy scores associated with about a standard deviation change in the corresponding affective measure.

In addition, two variables related to teaching and learning activities in schools had significant effects on the outcome. The slope of in-school time spent on science learning ($\gamma_{100}$) was 6.94 ($t = 3.68$), and the slope of hands-on activities in science classes ($\gamma_{110}$) was -7.55. That is, students taking more science classes had better scores; and students involved in more hands-on activities got lower scores than their peers in same school.
Remaining independent measures at level 1—gender, instrumental motivation in science, interest in science, out-of-school time spent on science learning, and out-of-school activities related to science—turned out to have no significant effect on Turkish students’ scientific literacy scores.

The random effect part of the model revealed that school-level residual variance of the achievement (τ_oo) was 3872.14 with a χ^2 value of 7183 (p < .001) indicating the existence of significant differences among school means on the dependent variable. The 95% confidence interval range for the plausible values here was 410.59 ± 1.96\((3872.14)^{1/2}\) = (288.63, 532.55), which is close to the one from the unconditional (ANOVA) model.

Reliability estimate for the intercept was very high (.99); on the contrary, slopes had very low reliabilities ranging from .19 to .28. According to Raudenbush and Bryk (2002), the chief reason for low slope reliability is that the variances of the student-level predictors’ slopes are much smaller than the variances of the outcome means across schools. But for these authors the low reliability on slopes would not discredit analysis results; in fact, a slope reliability coefficient above .05 is acceptable for HLM.

Student-level residual variance was obtained as 2272.59. The proportion of the variance accounted for by the student-level variables can be calculated by comparing the level-1 variance estimates obtained from one-way ANOVA and random coefficient models:

\[
\left(\frac{\sigma^2 (ANOVA) - \sigma^2 (random\ coefficient)}{\sigma^2 (ANOVA)}\right)
\]
That means, 26.8% of the within-school variance in Turkish students’ scientific literacy scores is explained by the variables included in the model; and 73.2% of the variation remained unexplained. Furthermore, all variables except for GENSCIE had significant residual variances; thus, the relationship between those variables and PISA scientific literacy scores vary significantly across schools. Since the differentiated effects of those variables are beyond the scope of this study, no further analysis on them was done.

School-level residual variance, after including the student-level variables, was obtained as 3872.63. Reduction of the residual variance at level 2 was less than 1%:

\[
\frac{(\tau_{oo} \text{ (ANOVA)} - (\tau_{oo} \text{ (random coefficient)}) / (\tau_{oo} \text{ (ANOVA)})}
\]

\[
= \frac{(3906.94-3872.63)/3906.94 = .01.}
\]

This finding showed that the student-level variables did not have an effect on explaining between-school differences in scientific literacy.

**Research Question 3**

Which school-level measures are associated with differences in mean school achievement?

To answer the final research question two models were developed. The first model, means as outcomes (illustrated in Chapter 3) was utilized to examine the effects of school-level variables on the scientific literacy scores of schools. In other words, student-level part of the model did not differ from the one established in one-way ANOVA model, but school means were regressed on the predictive variables measured at level 2. The second, intercepts as outcomes, was developed based on the results of
previous models, and explored the school-level variables’ effects on mean achievement scores after controlling for the student-level variables.

Raudenbush and Bryk (2002) suggested that dividing the level 2 variables into conceptually distinct subsets and conducting HLM separately, then running a new analysis with the significant variables is the appropriate strategy when doing an exploratory study. Following that advice, the first means as outcomes model included the school resources cluster of SCMATEDU, STRATIO and STSHORT as predictors. Results revealed that none of those variables had a significant effect on the school literacy means. The second model contained contextual variables including SCIPROM, SECTOR, MESCS, MINTIME, MINTACT, and MSCINVEST. Table 4.7 has parameter estimates derived from that means as outcomes model.
<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Models for School Means</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>420.63</td>
<td>4.81</td>
<td>87.43**</td>
</tr>
<tr>
<td>SCIPROM, $\gamma_{04}$</td>
<td>3.11</td>
<td>4.3</td>
<td>.72</td>
</tr>
<tr>
<td>SECTOR, $\gamma_{05}$</td>
<td>44.67</td>
<td>18.01</td>
<td>2.48*</td>
</tr>
<tr>
<td>MESC, $\gamma_{06}$</td>
<td>42.24</td>
<td>8.92</td>
<td>4.73**</td>
</tr>
<tr>
<td>MINTIME, $\gamma_{07}$</td>
<td>22.91</td>
<td>10.2</td>
<td>2.25*</td>
</tr>
<tr>
<td>MSCHANDS, $\gamma_{08}$</td>
<td>-18.89</td>
<td>10.91</td>
<td>-1.73</td>
</tr>
<tr>
<td>MSCINTACT, $\gamma_{09}$</td>
<td>36</td>
<td>26.72</td>
<td>1.35</td>
</tr>
<tr>
<td><strong>Random Effect</strong></td>
<td>Variance Component</td>
<td>df</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>School mean, $u_{0j}$</td>
<td>1030.53</td>
<td>148</td>
<td>1819.73</td>
</tr>
<tr>
<td>Level-1 effect, $r_{ij}$</td>
<td>3104.55</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05.
** p < .01.

Table 4.7: Means as Outcomes Model Results

The second model revealed that only three predictors at the school-level – SECTOR, MESC, and MINTIME— had significant relationships with the school mean of scientific literacy. The fixed effect results can be summarized as follows: on average, being a highly selective school (either an Anatolian or Science high school) is associated with an increase of 44.67 points – more than half a standard deviation— in school means ($t = 2.48$) holding other level 2 variables constant; one point increase in the school-average of the economic, social and cultural status is associated with 42.24 points increase in the school outcome ($t = 4.73$); and, one point increase in mean in-school time
spent for science classes is associated with 22.91 points increase in a school’s scientific literacy score mean (t = 2.25). Other level 2 variables were all statistically non-significant in estimating school-level achievement.

Random effect part of the model showed that between-school residual variance was considerably lower than the one produced by ANOVA model. Although unexplained variance due to school differences declined to $\tau_{oo} = 1030.53$ from $\tau_{oo} 3609.94$, it was still statistically significant. The $\chi^2$ statistic testing the null hypothesis that level 2 variance is equal to zero, implying the school means on the test scores are the same, was significant ($\chi^2 = 1819.73, p < .001$) after including the predictors at level 2. Since the school-level variance dropped radically, 95% of the school means are distributed in a narrower range than they were estimated based on ANOVA model. Namely,

$$\gamma_{oo} \pm 1.96 (\tau_{oo})^{1/2} = 420.63 \pm 1.96 (1030.53)^{1/2} = (357.71, 483.55).$$

The variance accounted for by the school-level predictors is obtained by comparing the school-level residual estimates across the ANOVA and means as outcomes models. The proportion of the explained school-level variance is:

$$\frac{(\tau_{oo} (ANOVA) - \tau_{oo} (Means as Outcomes)) / \tau_{oo} (ANOVA)}{\tau_{oo} (ANOVA)} = \frac{(3906.94-1030.53)/3906.94 = .736.}$$

73.6% of the between-school variance in scientific literacy achievement is due to the level-2 predictors included in the model above. Reliability estimate of the achievement means was calculated .92. This reflects a small decrease from ANOVA model; school means were very reliable estimates of the population values.
After finding out which level 1 and level 2 predictors had significant effects on literacy achievement, the intercept as outcome model was created to test which school-level variables were significant predictors of the achievement controlling for student-level variables.

**Level 1:**

\[ Y_{ij} = \beta_{0j} + \beta_{1j}(\text{GRADE}) + \beta_{2j}(\text{ESCS}) + \beta_{3j}(\text{GENSCIE}) + \beta_{4j}(\text{PERSCIE}) + \beta_{5j}(\text{RESPDEV}) + \beta_{6j}(\text{SCIEEFF}) + \beta_{7j}(\text{INTIME}) + \beta_{8j}(\text{SCHANDS}) + r_{ij} \]

**Level 2:**

\[ \beta_{0j} = \gamma_{00} + \gamma_{01}(\text{SECTOR}) + \gamma_{02}(\text{MESCS}) + \gamma_{03}(\text{MINTIME}) + u_{0j} \]
\[ \beta_{1j} = \gamma_{10} + u_{1j} \]
\[ \beta_{2j} = \gamma_{20} + u_{2j} \]
\[ \beta_{3j} = \gamma_{30} \]
\[ \beta_{4j} = \gamma_{40} + u_{4j} \]
\[ \beta_{5j} = \gamma_{50} + u_{5j} \]
\[ \beta_{6j} = \gamma_{60} + u_{6j} \]
\[ \beta_{7j} = \gamma_{70} + u_{7j} \]
\[ \beta_{8j} = \gamma_{80} + u_{8j} \]

In the model above, each school’s mean on the outcome is represented by \( \beta_{0j} \), and the expected changes in the students’ scores associated with student-level predictors at a particular school, or level-1 slopes, are characterized by \( \beta_{1j} \) through \( \beta_{8j} \). Similarly, \( \gamma_{00} \) represents the average of the school means, and \( \gamma_{10} \) through \( \gamma_{80} \) indicate the grand means of the slopes. While \( r_{ij} \) is the difference between the observed and predicted value of the outcome variable for student \( i \) in school \( j \), \( u_{0j} \) is the difference between the observed and estimated intercept for school \( j \), and \( u_{1j} \) through \( u_{8j} \) are the differences between the observed and estimated slope coefficients. Because random coefficient model results
revealed that level 1 variable GENSCIE’s relationship with the test scores did not change from school to school, its effect was fixed in the final model.

As seen in Table 4.8, combined model’s results are similar to the results of ANOVA, random coefficients, and means as outcomes models. All variables entered in the final model retained their significance at both levels. At level 1, GRADE, INTIME, and SCHANDS had slightly greater effect on the dependent variable; however, GENSCIE, PERSCIE and SCIEEFF had lower effect than that measured by the random coefficient model. While there was about 6 points increase in effect of MINIME, slight decreases in the effects of SECTOR and MESCS were observed.
### Table 4.8: Intercepts as Outcome Model Results

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model for school means, $\beta_0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>422.06</td>
<td>5.31</td>
<td>79.49**</td>
</tr>
<tr>
<td>SECTOR, $\gamma_{01}$</td>
<td>41.06</td>
<td>18.6</td>
<td>2.21*</td>
</tr>
<tr>
<td>MESCS, $\gamma_{02}$</td>
<td>41.65</td>
<td>7.6</td>
<td>5.48**</td>
</tr>
<tr>
<td>MINTIME, $\gamma_{03}$</td>
<td>25.67</td>
<td>9.19</td>
<td>2.79**</td>
</tr>
<tr>
<td>Model for GRADE slope, $\beta_1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{10}$</td>
<td>17.18</td>
<td>4.39</td>
<td>3.92**</td>
</tr>
<tr>
<td>Model for ESCS slope, $\beta_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{20}$</td>
<td>5.32</td>
<td>2.11</td>
<td>2.52*</td>
</tr>
<tr>
<td>Model for GENSCIE slopes $\beta_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{30}$</td>
<td>10.61</td>
<td>1.92</td>
<td>5.52**</td>
</tr>
<tr>
<td>Model for PERSCIE slopes $\beta_4$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{40}$</td>
<td>-5.04</td>
<td>2.17</td>
<td>-2.33*</td>
</tr>
<tr>
<td>Model for RESPDEV slope, $\beta_5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{50}$</td>
<td>5.06</td>
<td>1.25</td>
<td>4.05**</td>
</tr>
<tr>
<td>Model for SCIEEFF slope, $\beta_6$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{60}$</td>
<td>10.52</td>
<td>2.04</td>
<td>5.15**</td>
</tr>
<tr>
<td>Model for INTIME slope, $\beta_7$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{70}$</td>
<td>7.04</td>
<td>1.85</td>
<td>3.8**</td>
</tr>
<tr>
<td>Model for SCHANDS slope, $\beta_8$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\gamma_{80}$</td>
<td>-7.71</td>
<td>1.69</td>
<td>-4.57**</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>School mean, $u_{0j}$</td>
<td>1068.34</td>
<td>130</td>
<td>1772.95</td>
<td>.00</td>
</tr>
<tr>
<td>GRADE, $u_{1j}$</td>
<td>197.18</td>
<td>133</td>
<td>170.56</td>
<td>.02</td>
</tr>
</tbody>
</table>

(Continued)
### Table 4.8 (Continued)

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Variance Component</th>
<th>df</th>
<th>( \chi^2 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCS, ( u_{2j} )</td>
<td>39.63</td>
<td>133</td>
<td>323.62</td>
<td>.00</td>
</tr>
<tr>
<td>PERSCIE, ( u_{4j} )</td>
<td>40.86</td>
<td>133</td>
<td>220.53</td>
<td>.00</td>
</tr>
<tr>
<td>RESPDEV, ( u_{5j} )</td>
<td>18.17</td>
<td>133</td>
<td>160.79</td>
<td>.05</td>
</tr>
<tr>
<td>SCIEEFF, ( u_{6j} )</td>
<td>38.7</td>
<td>133</td>
<td>235.43</td>
<td>.00</td>
</tr>
<tr>
<td>INTIME, ( u_{7j} )</td>
<td>23.21</td>
<td>133</td>
<td>231.99</td>
<td>.00</td>
</tr>
<tr>
<td>SCHANDS, ( u_{8j} )</td>
<td>30.67</td>
<td>133</td>
<td>180.92</td>
<td>.00</td>
</tr>
<tr>
<td>Level-1 effect, ( r_{ij} )</td>
<td></td>
<td></td>
<td>2303.6</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\) Chi-square statistics are based on only 134 of 155 units that had sufficient data for computation. Fixed effects and variance components are based on all of the data.

* p < .05.

** p < .01.

Intercepts as outcomes model’s results indicated that on average there was a 17-point gap (t = 3.92) between each consecutive grade. One unit increase in ESCS, GENSCIE, RESPDEV, SCIEEFF, and INTIME was associated with 5.32 (t = 2.52), 10.61 (t = 5.52), 5.06 (t = 4.05), 10.52 (t = 5.15), and 7.04 (t = 3.8) points increases, respectively, in a students’ literacy scores within a school keeping other variables constant. On the other hand, the same unit increase in PERSCIE and SCHANDS was associated with 5.04 (t = -2.33), and 7.71 (t = -4.57) point decrease in the dependent variable.

According to the final model, SECTOR’s slope had a value of 41.06 (t = 2.21), representing about half a standard deviation gap in mean school achievement between highly selective schools and others controlling for other variables included in the model.
In addition, on average, one unit increase in schools’ economic, social and cultural status was associated with more than 41 point increase in the mean of the school outcome ($t = 5.48$). The association between average in-school time allocated for science classes and school-level outcome was close to 26 points ($t = 2.79$). Figure 4.1 illustrates the relationships between level 2 variables and school mean achievement.

![Figure 4.1: Graphical Representation of the Relationships between Significant School-level Variables and School Mean Achievement of Scientific Literacy](image)

Furthermore, compared to the ANOVA model, intercept as outcome model produced lower between school residual variance; nonetheless, it was slightly higher than the one produced by means as outcomes model. The final model produced a significant level 2 variance ($\tau_{oo} = 1068.34, \chi^2 = 1772.95, p < .001$) an indication of significant unexplained variation among schools. After including the predictors at both levels 95% of the school means were distributed between 358 and 486.12, in a higher range than
previously estimated by the means as outcomes model. In addition, school-level variables examined in the intercepts as outcomes model explained 72.7% of the variation in that level. Student-level residual variance ($\sigma^2$) was 2303.6. The proportion of the level 1 variance accounted for by the predictors was 25.8%. About two third of the variation in achievement at student-level was still unexplained.

Testing HLM Assumptions

To determine the adequacy of the hierarchical model, several assumptions should be addressed. Raudenbush and Bryk (2002) stated that an HLM model should have the following characteristics:

1. Student-level errors are independent and normally distributed with a mean of zero and equal variances among schools.
2. Student-level predictors are independent of student-level residuals.
3. Random errors at school-level are multivariate normal with a mean of zero. The random error vectors are independent among schools.
4. School-level predictors are independent of school-level residuals.
5. Student-level errors and school-level errors are not correlated.

Assumptions 2, 4 and 5 are related to the associations between the variables included in the model and the ones excluded but consigned to the residual terms. Those assumptions are important for model specification. Misspecification is reported to result in biased parameter estimates and wrong conclusions about the intercept and slopes. To minimize the model misspecification, a number of steps suggested by researchers (Raudenbush & Bryk, 2002; Heck & Thomas, 2004; Bickel, 2007) were taken. First, as
mentioned earlier, bivariate correlations were examined and predictors were excluded from the model when necessary. Second, variance inflation factor (VIF) statistics were generated to ensure the absence of multicollinearity. Third, aggregates of student-level predictors were entered in the equations. Lastly, based on the reliability statistics of the slopes, they were included as randomly varying predictors at school-level.

In order to test assumptions 1 and 2, residual files were created and examined. Due to technical restrictions, assumptions checks could not be conducted using all the data in the data set. Since the HLM program could not generate statistics to test the equal level 1 residual variances using all five plausible values (PV) simultaneously, only the first PV was utilized as the dependent variable to test if errors are homogenous across schools. Before conducting this check of the assumption, the final intercepts as outcomes model was replicated using only the first PV as the dependent. After ensuring that the model parameters were almost identical, homogeneity of variance was examined. According to the test results, student-level variances were heterogeneous ($\chi^2 = 1241.42$, df = 133, p < .001). Several possible causes for heterogeneity of the level 1 variance have been reported. Using non-normal data, bad data coding, fixing the effects of level-1 predictor when it should be treated as random or non-randomly varying, and omission of important level 1 variables are among them (Raudenbush et al., 2004).

Skewness and kurtosis analyses on the parametric variables included in the final model indicated that they were normally distributed (See Appendix C); data set used for this study was provided by OECD and already checked for bad coding. In addition, effects of the all level 1 variables, except GENSCIE were allowed to vary randomly in
the final model. After allowing the effect of GENSCIE vary randomly across schools, homogeneity assumption was achieved ($\chi^2 = 145.04, \text{df} = 133, p = .22$). Unfortunately, HLM software does not have a utility to test the homogeneity of level 2 residuals; so the test could not be conducted at the school-level.

In order to see if student-level residuals are normally distributed Q-Q plot of expected and observed residuals was generated. When that assumption is not realized, fixed effects and standard errors could be biased (Raudenbush, et al., 2004). Since the following Q-Q plot resembles a 45 degree line, it can be concluded that the level-I errors are normally distributed (Raudenbush et al., 2004).

![Figure 4.2: Normal Q-Q Plot of Level 1 Residuals](image)
Multivariate normality of the school-level residuals was checked by examining the Q-Q plot of expected and observed Mahalonobis distance, a measure representing the distance of each observation from the mean of all variables. A 45 degree line is the evidence of the multivariate normality of the level-2 residuals. The results in Figure 4.3 indicate that this assumption is met. Note that MDIST represents the observed Mahalanobis distance and CHIPCT represents the order statistics, or expected Mahalonobis distance measure (Raudenbush, et al., 2004).

Figure 4.3: Plot of MDIST vs. CHIPCT
Summary

This chapter included a detailed analysis of 15-year old Turkish students’ scientific literacy achievement at PISA 2006. While the dependent variable of the study was represented by five plausible scientific literacy scores, independent variables were consisted of 25 variables. Among those predictors, 15 were measured at student-level and grouped in 4 clusters; remaining was at the school-level and divided into two groups. Two predictors were excluded from the subsequent HLM analyses based on the descriptive statistics.

A series of models indicated that more than half of the variation in the achievement occurred among schools. While eight student-level variables in the study (GRADE, ESCS, GENSCIE, PERSCIE, RESPDEV, SCIEFF, INTIME, and SCHANDS) explained about one-third of the variation at that level; three school-level predictors (SECTOR, MESCS, and MINTIME) were accounted for more than 70% of the variation at level 2. The results of the analyses were further discussed in Chapter 5 in relation to the factors measured at both levels.
CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The purpose of this study was to explore the effects of selected factors on 15-year-old Turkish students’ scientific literacy achievement in PISA 2006 through multilevel modeling. Although the literature examining such factors in terms of student performance on standardized tests is considerable, it mainly covers developed countries and generally does not apply to underdeveloped or developing countries (Heyneman, 1986; Jenkins, 2004).

In this study, four models were used to explore; a) if there were differences among schools in terms of their mean scientific literacy scores, b) student-level factors that might explain the differences in students’ scores within a particular school, and c) school-level factors that could account for scientific literacy score differences between schools. First, an empty model including no variables at any levels was examined to determine the proportion of the variance in scientific literacy occurring within school and between schools. Second, variables related to student background, in-class teaching and learning practices, affective variables, and out-of school activities were entered into the
first model to predict within school differences. Third, only the variables related to school resources and school context were used to predict between school differences. Finally, the second and the third model were combined to examine the effects of significant factors on the outcome measure.

This chapter is divided into three major sections. The first deals with conclusions about the research design, the second is a discussion of the findings with an emphasis on interpretation, and the third focuses on policy implications and recommendations for further research.

Conclusions about the Research Design

Three conclusions are drawn about the general features of PISA data.

1. Both student and principal surveys had very high response rates, no less than 96% for any variables included in the study. That was probably due to the plan and implementation of the PISA study with participants doing the surveys during the school day. Given this, a probable conclusion is that non-responses are not affecting the results.

2. The samples were believed to be representative of the population of 15-year old students in Turkey and the schools they attend since two-stage stratified sampling was part of the design of PISA.

3. A rigorous translation process ensured that all students were asked almost the same questions in their respective languages. On the other hand, that translation process might have produced some items difficult to understand for some students in specific cultural contexts (Salzman & Lowell, 2008). Because the
original surveys and cognitive items were written in English and French, students talking other languages might have had a disadvantage in understanding some of the questions. The researcher is a native of Turkey and has reviewed the language of the surveys and sample items and does not perceive this type of issue to be in play.

Regarding the nature of the present study the following conclusions are offered.

1. This study was designed based on input-output and process-output studies of educational effectiveness as well as OECD’s conceptual map of educational indicators. The model is comprehensive in terms of the variables it utilized. On the other hand this study was not intended to include every possible predictor to explain the variability in the outcome. Instead, it aimed to explain as much variability as possible with a small number of predictors. Additional or alternative predictors might have explained more variability in the outcome.

2. The results of the study are seen as reliable from a statistical standpoint since the analytic method employed is considered a suitable one for the nested data structure (Raudenbush & Bryk, 2002; Bickel, 2007).

3. As mentioned in Chapter 3, different variables like socio-economic background of the students were used to create the dependent variable manifested in five plausible values (PV). It should be noted that, although PVs can provide better achievement prediction than point estimates, they are not free from error.
Discussion of the Findings

Decomposition of Total Variance

While research on student and school factors’ effects on student performance have failed to reveal a clear picture as to their relative importance; most studies relative to less developed countries reported that school-related variables are more salient for achievement (Heyneman, 1986; Heyneman & Ransom, 1990; Sammons et al., 1997). Conversely, more recent investigations, especially the ones using TIMSS and PISA scores suggested varying relationships between student and school-level variables and scores on international assessments.

Initial inferential analysis through the first model, one-way ANOVA, indicated that not only do schools vary in terms of their mean achievement scores, most of the variation (55.7%) in scientific literacy comes from the schools. That result supports the findings of some studies and contradicts others. For example, it challenges Fuchs and Wobmann’s (2006) conclusion that effects of variables related to student characteristics and family background are larger than those of school-level variables from comparative analyses of PISA 2000 results of 31 countries. Then again, it agrees with Marks’ (2006) finding that background variables like socio-economic status have significant relationships with assessment results and school variables such as school program have greater association. The ANOVA model’s results are additionally in line with Mokhsein’s (1999) findings in which 1995 TIMSS data indicated that 58% of the variability in science achievement of Malaysian 8th graders came from the between
school differences. Details of the variance portioning at student and school levels are given below.

*Student-level Factors and Scientific Literacy*

*Background Factors and Scientific Literacy*

Earlier studies on science achievement found gender differences favoring males, recent ones proposed that the gap has narrowed substantially, if not disappeared (Harker, 2000; Lauzon, 2001; Ulkins, 2007; Brotman & Moore, 2008). Yet studies utilizing PISA data show that gender gap still exists. For example, Ma (2003) found that boys outperformed girls in scientific literacy among both immigrant and nonimmigrant groups on PISA 2000 in France. Examining 28 OECD countries’ PISA 2000 data Langen, Bosker, and Dekkers (2006) showed that boys did better than girls in all countries.

In the present case, students’ gender, grade, and economic, social and cultural status (ESCS) information were grouped under background characteristics. Unlike the previous studies, this study found that with a small margin, girls did better than boys controlling for other variables in the student model but the difference was not statistically significant. That would be a sign of a disappearing gender gap in Turkey.

Student grade was included as a control variable because it was expected that students attending higher grades would do better than the lower graders. So, its relationship with the outcome measure was not of primary interest; but its use as a covariate in a statistical model improves the accuracy of estimates (Bloom, Richburg-Hayes & Black, 2007). Thus it is concluded that the models in this specific study are more reliable than in previous ones (Erbas, 2005) that did not include the variable.
The relationship between achievement and socio-economic status has been well established in the literature. Researchers have consistently reported that socio-economic indicators (family income, parental education) have a positive impact on student test scores (Coleman, et al., 1966; Fuller, 1987; Turmo, 2002; Sirin, 2005). Studies utilizing PISA database also show that finding (Nonoyama, 2005; Is-guzel, 2006; Marks, 2006).

Similarly, the present investigation indicated that students having higher socio-economic and cultural status index score performed significantly better than others, but unexpectedly, that effect was considerably smaller than anticipated. One possible explanation here is that the achievement gap is fading out within Turkish schools; but schools be segregated based on socio-economic and cultural status. This finding is further discussed in the policy implications section later in this chapter.

### Affective Factors and Scientific Literacy

A scientifically literate person is seen as being interested in scientific issues and learning science and aware of the value of science for personal and social development (NSTA, 1971; Shen, 1975; Hurd, 1998; Laugksch, 2000, OECD, 2007). Responsibility for sustainable development (OECD, 2007, p.122) is another characteristic of scientific literacy. Variables measuring those factors and science self-efficacy, a factor seen as contributing to achievement (Bandura, 1993; Pajares, 1996; Zimmerman, 2000) were looked at via the affective factors cluster.

Some findings in this case between affective factors and scientific literacy were quite different than what previous researchers suggested. Contrary to previous observations instrumental motivation and interest in science were insignificant in
predicting the scientific literacy of Turkish students. Furthermore, while the students’ perceptions about the general value of science had a positive effect on scientific literacy; their views of the personal value of science had a negative effect. That is, students with high scientific literacy achievement believed that advances in science and technology improve the economy and bring social benefits; however, they did not find science relevant to them, and there would not be many opportunities for them to use science in the future. One explanation for that negative association would be that highly scientific literate students might have career plans in social sciences. Cengiz, Titrek, and Akgun (2007) indicated that a common conclusion of the studies on career preferences of high school students is that Turkish students’ career choice is mainly driven by parent expectations and preferences rather than their area of strength. Another possibility is that because most students with high scientific literacy also had high scores in mathematics and reading literacy (Kjaernsli & Molander, 2003), they could think math and reading are more relevant to them. Students with high literacy scores in more than one area could have highly varied interests that affect the associations noted here.

_Teaching and Learning Factors and Scientific Literacy_

the importance of the amount of time students engage in learning (quantity) and classroom experiences related to instruction (quality).

A vast majority of science education literature on teacher and student experiences in the classroom emphasized the positive effect of hands-on experiences, interactive teaching, and letting students construct their own knowledge through investigations and practical experiences on student attitudes towards, and understanding of science (Tobin, 1990; Freedman, 1997; Kipnis & Hofstein, 2005; Treagust, 2007). Studies with TIMSS data also supported that view (House, 2002; House, 2006). The present research found that classroom time devoted to science learning, or quantity of instruction has a positive effect on scientific literacy; however, other factors included in the teaching and learning cluster and seen as indicators of effective instruction in the science education literature had no effect (investigations and interactions) or a negative effect (hands-on experiences) on the scientific literacy of Turkish students. There was no significant difference between students as active learners and students receiving teacher-centered instruction (passive learners). Moreover, hands-on experiences (doing, designing and drawing conclusion from experiments) had a negative impact on scientific literacy. A common practice in Turkish schools is teacher-centered instruction which helps teachers save time and be more efficient (Worldbank, 2005). Considering that average class size in Turkish high schools is 34 (Incekara, 2006) teachers would have difficulties with exercising student-centered instruction.
Science Related Out-of School Activities and Scientific Literacy

Voluntary out-of school activities that are not part of school curriculum such as reading books and magazines, or watching TV programs about science are called informal learning activities. They are thought to help students create real-world connections with what they learn at school, and foster knowledge construction (Lynch, 2000; McLure & McLure, 2000; Ramsey & Edwards, 2004; Rennie, 2007). Surprisingly, the current study did not support that statement with no effect of informal learning was observed on the dependent variable. This raises a question about availability of publications that are suitable for elementary and high school students. For example Sahin and Altinay (2009) found that numbers of science-related books helping elementary school students improve their thinking skills and creativity is very limited.

The second factor analyzed under the science related activities cluster was out-of school time spent on studying science including doing homework. Studies about the effect of homework on learning provide inconsistent results. Cooper (1989) stated that requiring students do homework has many positive and negative effects. Immediate achievement and learning, better retention of knowledge and improved attitude toward school are among the potential positive effects; and loss of interest in learning and fatigue are among the negative effects of doing homework. After reviewing 32 studies correlating time on homework and achievement, Cooper, Robinson and Patall (2004) concluded that majority of the studies reported positive significant associations with achievement. In the present study, time spent on learning course material out-of school has no effect on the literacy scores. A possible reason for this could be that students learn
science at school well enough and homework or extra study does not add more knowledge, which is unlikely based on the test scores. Another possible reason could be that homework and/or extra study materials might be focusing on facts rather than encouraging students to think about everyday applications of science which is the main focus of PISA studies.

School-level Factors and Scientific Literacy

School Resources and Scientific Literacy

School resources (student-teacher ratio, school infrastructure, and availability of instructional materials) have been viewed as having an important impact on achievement (Alexander, 1998; Verstegen & King, 1998; OECD, 2007). Yet, when Hanushek (2003) examined 90 individual publications published before 1995, only 14% had positive and statistically significant relationship between teacher-student ratios and test scores, 27% of the studies showed a positive and significant effect of increasing resources, and 7% had a negative trend.

According to OECD’s PISA 2006 report, the proportion of 15-year-olds in schools with vacant science teacher positions was 10% in Turkey, one of the highest among the participant countries. The vast majority of the principals including more than 30% of those schools with all science teaching positions were filled reported that instruction was hindered by the lack of qualified science teachers (OECD, 2007). As mentioned in the previous chapter, Turkey was among the countries whose science education suffered from lack of instructional materials. The three variables examined in this study, teacher-student ratio, lack of instructional resources, and science teacher
shortage all had no significant effect on PISA test scores of Turkish test takers. This result somewhat supported the empirical findings, and could be considered as a sign of a stronger need for increasing quality of teaching than quantity of the materials. Put another way, improving workforce quality is crucial for effective use of instruments.

School Context and Scientific Literacy

Besides the student-level variables of teaching and learning factors and school economic, social and cultural status, two other variables, school sector and amount of activities organized by schools to promote student learning, were in the context variables cluster. Among the teaching and learning variables only in-school time allocated for science learning was a significant predictor of school-level scientific literacy. About every two-hour increase in weekly time allocated for science lessons was associated with about 26 point increase in school’s mean score. School’s economic, social, and cultural status also produced an important change in the outcome - more than 41 points for one standard deviation change in the predictor variable.

Another school-level variable having significant effect on the outcome was school sector. Turkey has many types of high schools, regular, vocational, Anatolian, and Science high schools. Anatolian and Science high schools are highly selective; students can attend these only if they are among the top performers in nationwide selection exams taken upon finishing 8-year compulsory education. Graduates of these schools are more successful at university entrance exams. Because the literature indicated that prior achievement has an important role in a students’ subsequent test performance (Schreen & Bosker, 1998; Luyten, Tymms, Jones, 2009); school type was chosen as a control
variable. Doing so, associations of other predictors with scientific literacy were estimated with more precision and an estimate of the effect of school type on student achievement was obtained.

It is widely believed that “activities external to the classroom can enhance students’ learning in science, as they provide a motivation for students and help to place science in a real-life context” (OECD, 2007; p.259). But in this study school activities such as organizing science competitions, extracurricular science projects, and field trips did not make a difference on school mean of scientific literacy. This finding may not sound surprising because it is parallel to the findings about in-school activities’ effects; but, it is another issue pointing towards the need for professional development programs.

Recommendations for Policy-makers

As the first research endeavor using multilevel analysis techniques to examine the effective factors on scientific literacy of Turkish students, this study is a response to the Turkish Ministry of National Education (MONE)’s Education Research and Development Directorate’s (EARGED) call for research on international assessment database studies. It has implications that policy-makers can utilize to improve Turkey’s education system, particularly science education.

Since the study is a snapshot of the current state of the scientific literacy caution should be used in generalizing its results. More research is needed to initiate comprehensive changes. While variables like SES are not easy to manipulate, others like in-class teaching and learning strategies could be worked on to improve the effectiveness of the system. Before providing suggestions for policy-making it is important to provide
evidence of the practical nature of the variables found to be statistically significant in this study. For that reason, effect size (ES) statistics for the final HLM model were calculated using the formula provided by Kirk (1995).

$$\text{Effect Size (r)} = \left( \frac{t^2}{t^2 + df} \right)^{1/2}$$

The effect sizes are in Table 5.1. According to Cohen’s (1992) guidelines an Effect Size (ES) value less than .23 indicates a small effect, an ES between .24 and .36 corresponds to medium effect, and if it is larger than .36, the effect of the corresponding variable is large.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>SECTOR</td>
<td>0.24</td>
</tr>
<tr>
<td>MESCS</td>
<td>0.50</td>
</tr>
<tr>
<td>MINTIME</td>
<td>0.22</td>
</tr>
<tr>
<td>GRADE</td>
<td>0.58</td>
</tr>
<tr>
<td>ESCS</td>
<td>0.20</td>
</tr>
<tr>
<td>GENSCIE</td>
<td>0.80</td>
</tr>
<tr>
<td>PERSCIE</td>
<td>0.29</td>
</tr>
<tr>
<td>RESPDEV</td>
<td>0.31</td>
</tr>
<tr>
<td>SCIEEFF</td>
<td>0.75</td>
</tr>
<tr>
<td>INTIME</td>
<td>0.69</td>
</tr>
<tr>
<td>SCHANDS</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Table 5.1: Effect Sizes
School-level, mean economic, social and cultural status had a large effect, but the effect sizes of the other two variables were relatively small; on the other hand all the statistically significant variables measured at student-level have either medium or large effects on students’ literacy scores with the only exceptions being economic, social, and cultural status of the individual student. Because most of the predictors had moderate to large effect sizes, policy-makers are encouraged to ask questions like the ones presented below based on the findings of this study and initiate new ones.

The results showed that although within school differences in economic, social and cultural status has a significant effect on student’s PISA scores, school mean of that variable has a much stronger association with mean school achievement. In other words, schools with high SES populations are much more successful than others. Do high SES families have more involvement in their children’s education? Do high SES schools have better teachers? Do those schools have better disciplinary climate? Do those schools have higher achievement expectations than others? Answering those questions will help policy-makers understand the factors resulting in high effect of SES.

Students attending Anatolian and Science high schools not only have significantly better prior achievement scores; they also have better teachers who were chosen among the top performers at a series of nationwide teacher placement exams. New teacher training programs on the characteristics of highly qualified teachers would increase the workforce quality of other schools and probably reduce the between school achievement gap. Viewing schools as equalizers particularly in underdeveloped and developing countries, it is crucial to explore what portion of the difference between highly selective
schools and others is due to individual factors such as students’ IQ and study habits; what portion is due to social reproduction; and what portion is due to country’s educational policies, like placing the best teachers at best schools? The author does not propose that those teachers should be employed at other schools; rather he points out the necessity of nation-wide training programs for teachers. Otherwise Turkish schools probably reinforce the achievement gap between haves and have-nots.

One of the most important findings of this study is that an increase in students’ perceptions about the value of science for their own life associated with a decrease in their PISA scores (science is useful for humanity but not for them). As noted before this could be an indication of high achiever students choosing career paths in areas other than science or being successful in other assessment areas. New programs should be developed to encourage high achievers in science and technology areas. The number of Turkish students receiving high marks in international student assessments is small. It is crucial to channel high achievers to areas of strength, especially science.

Even though constructivist teaching and learning strategies are widely advocated and popular methods for science education (Stofflett & Stoddart, 1994; Omer, 2002), they had no effect on the scientific literacy of Turkish students. How can this situation be explained? Is this an indicator of insufficient and/or inefficient teacher education? While many changes have been done in K-12 educational system in recent years and the curriculum was aligned so that more constructivist teaching is required, pre-service and in-service teachers may lack skills to apply those methods, and new professional development programs may be needed.
There is conflicting evidence in the literature as to whether instructional material inputs translate into better learning. This study did not show a difference in the student performance in this regard. It is not easy to explain why providing more instructional material to the schools of a country whose majority of school administrators claim a lack of material hinders the quality of education, do not make a difference in student learning. Can it be because teachers have lack of knowledge and skills to use them? Is there a need for professional development programs to solve this problem? Complicating the issue further is that though the vast majority of the school principals claimed that lack of science teachers is an issue preventing schools from offering quality education, no difference in PISA scores between schools needing more teachers and others was detected.

Additionally an accountability system needs to be developed. Currently, teachers working at governmental schools are not held accountable for progress. In fact, schools are not accountable for improving learning. A set of school performance indicators and annual report card system could be established. Characteristics of effective schools in Turkey could be identified through research studies and school principals could be trained on them (OECD, 2007b).

Policy-makers and educators should also think about the reasons for lack of relationship between school activities to promote science learning (field trips, projects, and competitions) and scientific literacy. Participating in those activities are mostly voluntary, and provide contextual learning by exemplifying applications of science in real
life. This might be due to the lack of partnerships between schools and science centers and museums where most field trips take place (Rennie, 2007).

In this era of international competition achievement tests like PISA provide an early picture of future workforce quality. Unfortunately, that picture for Turkey is not that good. Many changes in the educational system might be necessary to improve Turkish students’ achievement results whether they assess literacy or content knowledge. To initiate change, policy-makers should have an open mind and be courageous in their actions. Kumar and Altschuld (2003) pointed out:

if policies are not there to foster and reinforce change, if resources in the form of time and training besides finances are not provided, if the environment does not afford the opportunity to try out ideas and to learn from failures, and if other aspects of a conducive, open atmosphere are not present, the probability of institutionalizing successful new programs will be extremely low (pp. 605-606).

Recommendations for Further Research

Although many variables at student-level were examined via this investigation the total explained within school variance in scientific literacy is relatively small (25.8%). The proportion of the explained total variance (between school and within school) by student-level variables was even smaller (11.5). However, 3 variables at the school-level (in-school time spent on science learning, economic, social and cultural status of school, school sector) explained a very large portion of the achievement differences among schools (72.7%). Those three factors could explain about 40.5% of the total variation – combined variance within and between schools– in PISA 2006 scores. Variables at
student- and school-levels together explained 52% of the variability. Many research avenues are therefore possible.

1. Alternative models could be developed and tested. This could result in more explained variance in assessment scores, and potentially eliminate the heterogeneity in level 1 error variance. An important variable that could be added is private tutoring in Turkey. Many high school students are registered with private institutions to prepare for university entrance exams. That variable might explain some of the student-level variability.

2. This study could be replicated via structural equation modeling (SEM) techniques to check if the associations between predictors and dependent variable are causal.

3. This study could be replicated using data from other developed and developing countries to examine differences and similarities in terms of the variable associations. What is more, a three-level study, countries being at the third level, could be conducted to explain the effects of country-level factors like expenditure per student or gross domestic product (GDP) on PISA scores. That analysis would let policy-makers and educators have a better understanding of the differences between countries.

4. Previous research suggested that country rankings based on TIMSS and PISA are very similar. A study with TIMSS data can be conducted to explore if similar constructs (i.e., school resources, in-class learning activities) have similar relationships with science achievement as measured by TIMSS. A study comparing the effect of school resources on science achievement and scientific
literacy could provide policy-makers and educators more information when planning changes in educational system.

5. This study included a comprehensive economic, social and cultural status variable which was found to have a significant relationship with the outcome. It is possible to estimate the effect of cultural, social, and economic factors using different variables available in the database (see Turmo’s 2004 framework mentioned in Chapter 2).

6. A random intercepts and random slopes model based on the final model in this study needs to be developed to test if the effects of the significant level 1 variables are the same across different school settings. A random intercepts and random slopes model could tell if the effect of hands-on experiences differs based on school selectivity (Raudenbush & Bryk, 2002).

7. Teacher and student responses may differ when they are surveyed about classroom teaching and learning styles (Kunter & Jurgen, 2004). A study comparing their perceptions would provide a better understanding of the associations between them and scientific achievement.
LIST OF REFERENCES


137


APPENDIX A

COLLINEARITY STATISTICS of LEVEL 1 VARIABLES
<table>
<thead>
<tr>
<th>Model</th>
<th>Tolerance</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Economic, social and cultural status</td>
<td>.904</td>
<td>1.106</td>
</tr>
<tr>
<td>General value of science</td>
<td>.549</td>
<td>1.820</td>
</tr>
<tr>
<td>Personal value of science</td>
<td>.388</td>
<td>2.575</td>
</tr>
<tr>
<td>Instrumental motivation in science</td>
<td>.533</td>
<td>1.876</td>
</tr>
<tr>
<td>General interest in learning science</td>
<td>.510</td>
<td>1.959</td>
</tr>
<tr>
<td>Responsibility for sustainable development</td>
<td>.813</td>
<td>1.230</td>
</tr>
<tr>
<td>Science self-efficacy</td>
<td>.634</td>
<td>1.577</td>
</tr>
<tr>
<td>Hands-on activities</td>
<td>.478</td>
<td>2.090</td>
</tr>
<tr>
<td>Student investigations</td>
<td>.470</td>
<td>2.127</td>
</tr>
<tr>
<td>In school time</td>
<td>.680</td>
<td>1.470</td>
</tr>
<tr>
<td>Out of school time</td>
<td>.667</td>
<td>1.499</td>
</tr>
<tr>
<td>Science activities</td>
<td>.512</td>
<td>1.953</td>
</tr>
</tbody>
</table>

Dependent variable: Plausible value in science
APPENDIX B

MULTIPLE REGRESSION MODEL FOR COLLINEARITY DIAGNOSTICS AT LEVEL 2
Model Summary

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>R²</th>
<th>Adj. R²</th>
<th>SE</th>
<th>F Change</th>
<th>df1</th>
<th>df2</th>
<th>Sig. F Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.9</td>
<td>.81</td>
<td>.8</td>
<td>.14</td>
<td>78.1</td>
<td>8</td>
<td>146</td>
<td>.000</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), School activities to promote the learning of science, Mean interactive teaching in science class, Shortage science teachers, Teacher-student ratio, Mean in-school time spent on science learning, Quality of educational resources, Mean hands-on activities in science class, Mean economic social and cultural status.

Predicted variable: Mean student investigations.
APPENDIX C

NORMALITY CHECK FOR PREDICTORS
<table>
<thead>
<tr>
<th>Item</th>
<th>Skewness</th>
<th>SE Skewness</th>
<th>Kurtosis</th>
<th>SE Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESCS</td>
<td>.373</td>
<td>.035</td>
<td>-.367</td>
<td>.070</td>
</tr>
<tr>
<td>GENSICE</td>
<td>-.086</td>
<td>.035</td>
<td>-.467</td>
<td>.070</td>
</tr>
<tr>
<td>PERSCIE</td>
<td>.052</td>
<td>.035</td>
<td>.125</td>
<td>.070</td>
</tr>
<tr>
<td>RESPDEV</td>
<td>-.194</td>
<td>.035</td>
<td>-.067</td>
<td>.070</td>
</tr>
<tr>
<td>SCIEEFF</td>
<td>-.026</td>
<td>.035</td>
<td>1.322</td>
<td>.070</td>
</tr>
<tr>
<td>INTIME</td>
<td>.140</td>
<td>.035</td>
<td>-1.283</td>
<td>.071</td>
</tr>
<tr>
<td>SCHANDS</td>
<td>.014</td>
<td>.035</td>
<td>-.064</td>
<td>.070</td>
</tr>
<tr>
<td>MESCS</td>
<td>.654</td>
<td>.195</td>
<td>.943</td>
<td>.387</td>
</tr>
<tr>
<td>MINTIME</td>
<td>.795</td>
<td>.195</td>
<td>.159</td>
<td>.387</td>
</tr>
<tr>
<td>PV1</td>
<td>.411</td>
<td>.035</td>
<td>-.059</td>
<td>.070</td>
</tr>
<tr>
<td>PV2</td>
<td>.369</td>
<td>.035</td>
<td>-.118</td>
<td>.070</td>
</tr>
<tr>
<td>PV3</td>
<td>.384</td>
<td>.035</td>
<td>-.164</td>
<td>.070</td>
</tr>
<tr>
<td>PV4</td>
<td>.356</td>
<td>.035</td>
<td>-.073</td>
<td>.070</td>
</tr>
<tr>
<td>PV5</td>
<td>.385</td>
<td>.035</td>
<td>-.155</td>
<td>.070</td>
</tr>
</tbody>
</table>