The Effect of Reform-Based Science Teaching on
SES-Associated Achievement Gap on PISA 2006:
A Comparative Study of the United States and Taiwan

A Dissertation
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
NAI-EN TANG

Dr. Lloyd H. Barrow, Dissertation Supervisor

DECEMBER 2015
The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

THE EFFECT OF REFORM-BASED SCIENCE TEACHING ON SES-ASSOCIATED ACHIEVEMENT GAP ON PISA 2006: A COMPARATIVE STUDY OF THE UNITED STATES AND TAIWAN

presented by Nai-En Tang,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

________________________________________
Professor Lloyd Barrow

________________________________________
Professor Marcelle Siegel

________________________________________
Professor Mark Volkmann

________________________________________
Professor Ze Wang

________________________________________
Professor William Miller
ACKNOWLEDGEMENTS

I would like to express my gratitude to all of those who have made this project possible. My sincere thank you goes to my advisor, Dr. Lloyd Barrow, for his thoughtful guidance during my years of study abroad. His encouragement, patience, supervision and support from the preliminary to the concluding level enabled me to complete this dissertation.

I want to express my appreciation to my dissertation committee for their responses and suggestions on the dissertation. I am very thankful to Dr. Mark Volkmann, Dr. Marcelle Siegel, Dr. Ze Wang, and Dr. William Miller for their helpful feedback in my proposal writing and dissertation revision process.

Furthermore, I would like to acknowledge and thank all the faculty, students, and staff at the University of Missouri who have provided valuable help to me, particularly those in the Science Education program of the Department of Learning, Teaching, and Curriculum.

Very special thanks to my parents who have encouraged me, supported me, and trusted me as I pursued the doctoral degree. Finally, I would like to thank my wife, Chia-Lin Tsai. She has always been there cheering me up and helping me through all of the dissertation process.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .............................................................................................................. ii

LIST OF TABLES .......................................................................................................................... v

LIST OF FIGURES ....................................................................................................................... vii

ABSTRACT ..................................................................................................................................... viii

Chapter

1. INTRODUCTION ...................................................................................................................... 1

   Introduction to Reform

   Significance of the Study

   Purpose of Study

   Definitions

   Assumptions

   Summary

2. LITERATURE REVIEW .......................................................................................................... 18

   Theoretical Framework: Social Cognitive Theory

   Implementation of Reform-Based Science Teaching and Learning

   Measures of Socio-Economic Status (SES)

   International Comparison

   Summary

3. METHODOLOGY ..................................................................................................................... 40

   Research Questions and Hypotheses

   Data Source

   Measurement and Variables
Data Analysis

Limitations

Summary

4. RESULTS .................................................................................................................64

Descriptive Statistics

Implementation of Reform-Based Science Learning

Availability of Reform-Based Science Learning

SES Achievement Gap and Reform-Based Science Learning

Summary

5. DISCUSSION ...........................................................................................................91

Summary of the Study

Conclusions

Discussions

Recommendations for Future Research

REFERENCE.............................................................................................................110

VITA.........................................................................................................................121
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student Composition by Gender and Race for the PISA 2006 U.S. and Taiwan Data</td>
<td>43</td>
</tr>
<tr>
<td>2. Means and Standard Deviations of SES and their Three Components for the PISA 2006 U.S. and Taiwan Data</td>
<td>47</td>
</tr>
<tr>
<td>3. Means and Standard Deviations of Science Assessment for the PISA 2006 U.S. and Taiwan Data</td>
<td>48</td>
</tr>
<tr>
<td>5. Means and Standard Deviations of Fifteen Teaching and Learning Activities for the PISA 2006 U.S. and Taiwan Data</td>
<td>65</td>
</tr>
<tr>
<td>7. Mean scores of outcome variables for the four science learning subgroups for U.S.</td>
<td>70</td>
</tr>
<tr>
<td>8. Goodness-of-Fit Criteria for Various Latent Class Models for the PISA 2006 Taiwan data</td>
<td>72</td>
</tr>
<tr>
<td>9. Average latent subgroup probabilities for most likely latent subgroup membership (row) by latent subgroup (column) for the 5-subgroup model for Taiwan</td>
<td>73</td>
</tr>
<tr>
<td>10. Average latent subgroup probabilities for most likely latent subgroup membership (row) by latent subgroup (column) for the 4-subgroup model for Taiwan</td>
<td>73</td>
</tr>
<tr>
<td>11. Mean scores of outcome variables for the four science learning subgroups for Taiwan</td>
<td>76</td>
</tr>
<tr>
<td>12. Logistic regression coefficients and odds ratio for 4-subgroup model with SES and gender as a covariate using the highest usage reform-based learning subgroup as the comparison subgroup for U.S.</td>
<td>77</td>
</tr>
<tr>
<td>13. Mean scores of covariate variables for the four science learning subgroups for U.S.</td>
<td>77</td>
</tr>
<tr>
<td>14. Logistic regression coefficients and odds ratio for 4-subgroup model with SES and gender as a covariate using the highest usage reform-based learning subgroup as the comparison subgroup for Taiwan</td>
<td>78</td>
</tr>
</tbody>
</table>
15. Mean scores of covariate variables for the four science learning subgroups for Taiwan .................................................................78

16. The unstandardized and standardized estimates of structural equation model for U.S. ........................................................................................................81

17. The unstandardized and standardized estimates of structural equation model for Taiwan ........................................................................................................83

18. Parameter Estimates and Standard Errors for Four-Subgroup Mixture Regression Model for U.S. ..................................................................................................................84

19. Parameter Estimates and Standard Errors for Four-Subgroup Mixture Regression Model for Taiwan ........................................................................................................85

20. The unstandardized and standardized estimates of interaction model for U.S. ....87

21. The unstandardized and standardized estimates of interaction model for Taiwan .........................................................................................................................88

22. Summary of decision for the null hypotheses (α= 0.05) ........................................90
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Model of Social-Self-Outcome Interaction in Achievement Settings</td>
<td>19</td>
</tr>
<tr>
<td>2.</td>
<td>Conceptual Model Proposed in this Study</td>
<td>22</td>
</tr>
<tr>
<td>3.</td>
<td>Latent profile model diagram of science learning</td>
<td>50</td>
</tr>
<tr>
<td>4.</td>
<td>Latent class regression model diagram with covariate of SES</td>
<td>55</td>
</tr>
<tr>
<td>5.</td>
<td>Proposed SEM models of SES, science achievement, self-efficacy, self-concept, and science learning (i.e. hands-on activities, student investigation, interaction, and application)</td>
<td>56</td>
</tr>
<tr>
<td>6.</td>
<td>Regression mixture model diagram</td>
<td>58</td>
</tr>
<tr>
<td>7.</td>
<td>Proposed SEM interaction models of SES, science achievement, science learning (i.e. hands-on activities, student investigation, interaction, and application), and the interaction between SES and science learning</td>
<td>59</td>
</tr>
<tr>
<td>8.</td>
<td>Latent profile of four identified science learning groups for the U.S.</td>
<td>69</td>
</tr>
<tr>
<td>9.</td>
<td>Latent profile of five-subgroup model for Taiwan</td>
<td>72</td>
</tr>
<tr>
<td>10.</td>
<td>Latent profile of four identified science learning subgroups for Taiwan</td>
<td>75</td>
</tr>
<tr>
<td>11.</td>
<td>SEM models of SES, science achievement, self-efficacy, self-concept, and science learning (i.e. hands-on activities, student investigation, interaction, and application)</td>
<td>79</td>
</tr>
<tr>
<td>12.</td>
<td>Interaction models of SES and science learning (i.e. hands-on activities, student investigation, interaction, and application) on science achievement</td>
<td>86</td>
</tr>
</tbody>
</table>
THE EFFECT OF REFORM-BASED SCIENCE TEACHING ON
SES-ASSOCIATED ACHIEVEMENT GAP ON PISA 2006:
A COMPARATIVE STUDY OF THE UNITED STATES AND TAIWAN

Nai-En Tang
Dr. Lloyd H. Barrow, Dissertation Supervisor

ABSTRACT

The goal of this study is to examine how reform-based science teaching has been implemented and whether reform-based science teaching has promoted education equity through being available and beneficial for students from different socioeconomic status (SES) family backgrounds in the U.S. and Taiwan. No existing study used large-scale assessment to investigate the implementation and outcomes of the science reform movement in the U.S. and Taiwan. This study was developed to fill this gap using the Program of International Student Assessment (PISA) 2006 data including 5,611 students in the United States and 5995 students in Taiwan.

A Latent Profile Analysis (LPA) was used to classify students into different science learning subgroups to understand how broadly reform-based science learning has been implemented in classrooms. The results showed that students in the U.S. had more opportunity to learn science through the reform-based learning activities than students in Taiwan. Latent Class Regression (LCR) and Structural Equation Modeling (SEM) were used for examining the availability of reform-based science teaching in both countries. The results showed that in the U.S., higher SES students had more opportunity to learn science reform-based learning activities. On the other hand, students’ SES had no association with reform-based science learning in Taiwan. Regression Mixture Modeling and SEM were used to examine whether there was an
association between reform-based science teaching and SES-associated achievement gaps. The results found no evidence to support the claim that reform-based science teaching helps to minimize SES-associated achievement gaps in both countries.
CHAPTER ONE
INTRODUCTION

In the United States (U.S.), science education reform has been a continuous effort in the past two decades. This effort has drawn much attention from researchers, practitioners, and administrators. The vision for science education reform was to promote education excellence and education equity. The philosophical rationale for these U.S. reform documents was based on constructivism, a learning theory that consists of contributions from Jean Piaget, Lev Vygotsky, and David Ausubel. Based on this philosophy, the advocates of science education reform encouraged teachers to use scientific inquiry in their instruction to advance students’ understanding of scientific concepts and procedures. The goals of science education reform was to promote students’ academic achievement and promote equal opportunity to learn for all students. While many studies investigated the effect of education reform on education excellence (e.g., academic success), relatively few studies investigated education equity, the achievement of which is another mission behind the reform.

Thus, purpose of the study was to investigate the effect of reform-based science teaching on education equity, by assessing the availability of reform-based instructional activities during students’ science learning and their effect on SES-associated achievement gaps. This study not only investigated the effect of science reform in the U.S. but also explored similar science reform issues in Taiwan. Although the two countries have very different cultures and history in science education reform, they share similar goals and visions of the outcomes. This study provided the opportunity to make comparisons between science education reform issues in the U.S. and in Taiwan.
Introduction to Reform

Science Education Reform in the United States

In the U.S., during the past two decades, the science education reform has been carried out by the National Science Foundation (NSF), the National Research Council (NRC), and the American Association for the Advancement of Science (AAAS). These organizations developed innovative K–12 policy documents, organized teacher professional development programs, and supported science teaching and learning at school, district, and state levels. Two policy documents--*Project 2061: Science for All Americans* (AAAS, 1989) and *National Science Education Standards* [NSES] (NRC, 1996)--provided new visions and directions for science education in the U.S. Two accompanying documents, *Benchmarks for Scientific Literacy* (AAAS, 1993) and *Inquiry and National Science Education Standards* (NRC, 2000), were published to serve as practical guides for K-12 science teachers, school administrators, and science teacher educators. These two documents also served as guidelines for the development or refinement of states’ own science frameworks and as the content basis for the improvement of schools’ curriculum and assessments.

The philosophical rationale for these U.S. reform documents was based on constructivism, the learning theory that consisted of contributions from Jean Piaget, Lev Vygotsky, and David Ausubel. The basic idea of constructivism is that students cannot learn science well if they only passively receive knowledge from their science teachers. Instead, students should actively participate in the process of constructing knowledge when learning science. Based on this philosophy, one common goal for science education reform was to encourage teachers to use scientific inquiry in their instruction to advance students’ understanding of scientific concepts and procedures. These reform documents encouraged
science teachers to change from traditional teacher-centered teaching to reform-based student-centered teaching, such as engaging students in scientific activities and conversations that related to students’ everyday life. These scientific inquiry approaches, suggested by *NSES* (NRC, 1996; 2000), encouraged students to learn science by asking and refining questions about the natural world; by designing and conducting investigations; by gathering and analyzing information and data; by making interpretations; by creating explanations; by drawing conclusions; and by reporting findings. The constructivism framework also stressed social perspectives on the learning process. *Inquiry and National Science Education Standards* (NRC, 2000) suggested that teachers should provide students a learning community, in which students can communicate and share their scientific ideas in writing or oral presentation and receive peers’ feedback and questions.

**Education Reform in Taiwan**

Since the late 1990s, a series of political reforms in Taiwan drove every aspect of society, moving Taiwan towards greater openness, including openness in education. In general, the Taiwanese public perceived their students as lacking creativity, over-emphasizing testing, focusing on memorization over application, and spending too much time on school work. In addition, the Taiwanese school curriculum was perceived to be disconnected from real-life situations (Zhao, 2005). On the other hand, the Taiwanese public perceived students in the U.S. as happy, creative, and socially responsible, and it believed that the U.S. education model made positive influences on students (Zhao, 2005). Therefore, Taiwanese researchers and educators initiated an education reform movement, and the Taiwanese policy makers began to learn education reform ideas from the U.S.
In particular, the Taiwanese government adapted the curriculum and educational practices of the U.S. and developed a new curriculum framework, *Nine-year Integrated Curriculum Guidelines* (Ministry of Education [MOE], 1998). Since 2001, all schools in Taiwan have implemented this new curriculum. The central theme of the curriculum reform was to integrate related individual subjects into seven major fields of study at elementary and middle-school levels. The goal for this reform was to connect school curriculum with students’ real life and to decrease students’ study burden. In the field of science and technology, the Ministry of Education in Taiwan published a *Grade 1–9 curriculum of junior high and primary school: Science and technology* (MOE, 2006) and required elementary and middle school teachers to focus more on building a sense of competence and creativity in students rather than on memorizing factual knowledge. This guideline encouraged educators to involve student-centered and inquiry-based pedagogies to ensure that students can learn science knowledge and skills through scientific inquiry activities. More importantly, students should be cultivated with the ability to apply science knowledge and skills to solve problems in the real world, as well as to communicate these solutions to others (MOE, 2006).

**Significance of Study**

This study will be a significant endeavor in investigating whether reform-based teaching helps to increase equity and bridge the achievement gap for students from disadvantaged backgrounds. This study will also be beneficial to the teachers, administrators, and researchers to understand the effect of reform-based curriculum in their classroom setting on students’ achievement, and how these effects would potentially address education equality issues. By analyzing the frequency of using reform-based activities, this study help readers to see the real classroom situations and the implementation of reform-based curriculum. Moreover, this
research will discuss the science reform issues in two different counties (the U.S. and Taiwan), and offer recommendations on future research, which will allow readers to gain insights of the same issues in different contexts.

This study will be helpful to the researchers who are interested in international comparative education studies and informing them in the area of science reform and education equity in the U.S. and Taiwan. It will also serve as a future reference for researchers on the subject of SES achievement gaps in science. And importantly, this research will educate readers in the history of education reform in science, challenges in the classroom, and the association between reform-based teaching and SES achievement gaps in science. The unique challenges and SES achievement gap issues in each country were summarized below.

**The United States**

Whereas the science education research community shared the ideas of these more effective teaching approaches, teachers in K-12 classroom settings had many barriers to implementing the reform-based science teaching. Anderson (2002) categorized these barriers in terms of three dilemmas: political dilemmas (e.g., parental resistance, conflicts among teachers, and differing judgments about justice and fairness); cultural dilemmas (e.g., different beliefs and values about learning and assessment); and technical dilemmas (e.g., limited abilities to teach, challenges of new teacher and student roles, and inadequate in-service training).

In addition to barriers in real classroom settings, education policy makers created more barriers for reform-based science teaching by implementing the No Child Left Behind Act [NCLB] (U.S. Department of Education, 2002), which led to an increased emphasis on high-stakes testing. As a consequence, states, districts, and schools were judged and student ability was measured by students’ achievement test scores in the core subjects. Whitford and Jones
(2000) indicated that NCLB policy led teachers back to more traditional teaching methods, such as covering large amounts of content and emphasizing factual knowledge, in order to pursue better achievement scores. Given the conflicts among science education research community, classroom settings, and policy makers, it is interesting to understand how teachers teach and how students learn science in the U.S. classrooms.

When U.S. was working toward reform-based science teaching, the NSES (NRC, 1996) set a goal for the science education reform:

> The intent of the Standards can be expressed in a single phrase: Science standards for all students. The phrase embodies both excellence and equity. The Standards apply to all students, regardless of age, gender, cultural or ethnic background, disabilities, aspirations, or interest and motivation in science. (p. 2)

In addition to the focus on excellence, the statement from NSES (NRC, 1996) raised the important issue of equity. The need to promote educational equity in science education reform came from long-standing situations of significant achievement gaps by gender, race/ethnicity, and socioeconomic status (SES) in the U.S. Among different students’ backgrounds, SES is an important factor that influences people in many aspects of life and thus attracted many education researchers’ attention. In education, many studies have found SES to be a strong predictor of students’ achievement. In Coleman et al.’s (1966) study on *Equality of Educational Opportunity*, the author concluded that SES had greater influence on achievement than any activity within schools. Furthermore, low-SES students' learning was at a disadvantage that could not be overcome by school environment. Thus, SES was the dominant factor in determining students' academic success, regardless of the school they attended.

The science achievement gap between students with high- and low-SES family background has been a long-standing issue. Several large-scale studies used different indicators to estimate individuals’ SES and arrived at the same conclusion. For example, studies by the
National Assessment of Educational Progress (NAEP) used student’s eligibility for free or reduced-price school lunch as an indicator of SES. The results of those studies showed that from 1996 to 2005, the achievement gaps between students from low-income and high-income families remained unchanged for fourth, eighth, and twelfth grade students (Grigg, Lauko, & Brockway, 2006). A similar pattern of these achievement gaps also existed from 2009 to 2011 (National Center for Education Statistics [NCES], 2012). Moreover, Ma and Wilkins (2002) investigated the growth rate of science achievement from seventh grade to twelfth grade using longitudinal data from the Longitudinal Study of American Youth (LSAY). They used parent-reported education and occupation and student-reported household possessions as an indicator of students’ SES. They found that SES was related to the rate of growth in science achievement. Thus, the achievement gap increased with each increasing grade. The result indicated that school science teaching did not successfully close the achievement gap but in fact increased the achievement gap.

In addition to direct impact, SES was indirectly linked to students’ achievement through students’ race/ethnicity (Brooks-Gunn & Duncan, 1997). In the U.S., students’ minority status was often linked with their SES and achievement. According to the U.S. Department of Education reports (2000; 2006), minority students on average had lower academic achievement than their white peers. The report attributed the lower achievement to three main reasons: 1) minority students were from low SES families or lived in a single-parent household; 2) their parents were less educated; and 3) they went to under-funded schools (NCES, 2000; 2006). The NAEP study had also monitored student science achievement for different ethnicity groups and reported the long-term trend from 1969 to 2011. These reports showed that from 1970 to 1999, the achievement gap between white and black students in science was generally minimum for 9-
and 13-year-olds, but greater for 17-year-olds. The gap between white and Hispanic students at any age in 1999 was not significantly different from 1977. Since 1992, it had widened somewhat among 13-year-olds (Campbell, Hombo, & Mazzeo, 2000). In the recent two decades, the achievement gaps between white and minority students remained unchanged for eighth-grade and twelfth-grade students from 1996 to 2005 (Grigg, Lauko, & Brockway, 2006) and from 2009 to 2011 (NCES, 2012).

Moreover, Muller, Stage, and Kinzie (2001) investigated the growth rate of science achievement from eighth grade to twelfth grade by different racial-ethnic groups using longitudinal data from the National Education Longitudinal Study (NEL: 88). They found that African-American and Hispanic students not only had lower initial achievement at Grade 8 but also had a lower growth rate in science achievement. The science education reform started before 1990 and continued to be advocated through 2000; thus, it was expected that some evidence would show the decrease of achievement gap for demographic groups. However, these large-scale and longitudinal studies showed that even though the education reform movement had the good intentions of increasing education equity and closing achievement gaps, the achievement gap in science remained.

When reviewing past research, Lynch (2000) described equity issues in terms of equality of inputs and equality of outputs. Equality of outputs means that all students should have the same level of success regardless of their demographic background, and equality of inputs means that students should have the same level of educational opportunity and resources regardless of their demographic background. The United States has a decentralized education system based upon the federal Constitution, which gives power over education to the states and local authorities. The idea of decentralization is based on giving more power to local leaders and
school officials, who apparently know more about local educational situations than national officials, and who have more ability to lobby for more resources from the local community. However, this decentralization caused education resources to be distributed unequally, and the school systems in low-SES communities were often under-resourced (Aikens & Barbarin, 2008). Given the fact that low-SES students come from low-SES communities and then go to under-resourced schools, it is not unexpected that under-resourced schools also contributed to the achievement gap.

Although research had provided evidence that reform-based teaching promoted student excellence (Wilson, Taylor, Kowalski & Carlson, 2010), not many studies investigated whether reformed-based teaching promoted education equity, which was helping low-SES students to learn as well as their counterparts. Given the commitment of the NSES in student excellence and education equity (NRC, 1996) and given the fact that science education reform seemed to fail to achieve its goal of creating equity, more research was needed to empirically investigate whether reform-based teaching helps increase equity and bridges the achievement gap for students from disadvantaged background in the U.S.

Taiwan

Taiwan’s education system faced a different obstacle when implementing reform-based science teaching. Examination tradition had been deeply rooted in the Taiwan education system and played a major role in teaching practice and school emphasis. The paper-and-pencil format of the entrance examinations for high schools and colleges focused exclusively on science content knowledge, which tended to direct teachers and schools’ attention to teaching science facts and concepts rather than science process and inquiry skills. Research by Chang, Chang, and Yang (2009) showed that science teachers believed that it was their duty to cover all the science
content outlined in textbooks to help students achieve high scores on high school and college entrance examinations. This belief pushed teachers to focus exclusively on traditional science teaching and less on reform-based science teaching.

The chronic achievement gap between high- and low-SES students not only existed in the U.S. but also in Taiwan. Although Taiwanese students ranked near the top in science performance for international assessments, such as the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA), the report showed a wide range of scores among students’ performance and showed that students’ SES played a major role in this gap (Mullis, Martin, Gonzalez, & Chrostowski, 2004; Mullis, Martin & Pierre, 2009). Taiwanese scholars also found that the relationship between Taiwanese students’ SES and achievement was positively correlated (Lin & Hwang, 2008). Cheng and Hung (2006), addressing equity issues in Taiwan’s education system, showed that students entering top universities in Taiwan were from high-SES families in disproportionately larger numbers than from low-SES families. Similarly, students in low-tier colleges had a higher percentage of economic constraint and had parents who did not attend college (Wu, 2009). Based on the fact that college entrance examination scores decided which university students would attend, such disproportion distribution of students’ SES backgrounds in top and low tier university had a direct impact on the SES achievement gap.

Like the U.S., Taiwan also had minority issues in education. The majority of Taiwanese belonged to the Han ethnic group, who came from China at different times. The minority of Taiwanese belonged to 12 aboriginal tribes who had similar ethnicity with other aboriginal island people in southwestern region of the Pacific Ocean. Many Taiwanese aboriginals resided in remote or high mountain areas, and most of them were culturally and economically
disadvantaged. The aboriginal students usually had fewer educational resources in their families and paid less attention to school learning (Chen, Crockett, Namikawa, Zilimu, & Lee, 2012).

In addition to aboriginal groups, there was another special minority group rapidly emerging in Taiwan, especially in schools, for the past two decades (Chen et al., 2012). They were children with foreign-born mothers. This minority group came from the fact that more and more Taiwanese women were achieving higher social and economic status. These women had a choice not to marry Taiwanese men with undesirable status, such as people with lower income, less education, physical disabilities, or who were beyond the age of 40. As a consequence, these “undesirable” Taiwanese men looked for their spouses in China or other Southeastern Asian countries, such as Indonesia or Vietnam (Hsueh, 2007). Since these foreign-born women were mostly from low-SES backgrounds in their home country and did not speak Taiwanese or Chinese, the children of such families were rapidly forming a culturally, economically, and linguistically disadvantaged group in schools (Chang, 2007).

Theoretically, current Taiwan education policies should have been helpful in terms of promoting educational equity, for several reasons (Chen et al., 2012). First, the educational system in Taiwan was centralized. What teachers taught and textbook content followed the national curriculum and guidelines. Every student received the same curriculum at the same grade regardless of where students live and what textbooks were used. Second, the government ensured equal distribution of resources across schools for learning and instruction and provided additional funding and grants for schools in remote or mountain areas or with large numbers of underperforming low-SES and minority students. Third, the education policy in Taiwan prohibited ability-grouping or the tracking of students’ achievement in school. Every student
received equal educational opportunities, and the curriculum and was taught by qualified
teachers regardless of students’ achievement.

Based on the educational situation described above, science teaching in Taiwan had two
problems: 1) the conflict between responding to science education reform and examination
tradition, and 2) the substantial achievement gap existing even under the policies that sought to
mitigate education equity. It was critical to understand these issues from the perspective of
science instruction. Specifically, more research was needed to examine to what extent the
reform-based science teaching was implemented in science classrooms in Taiwan and how
would the reform-based science teaching influenced Taiwanese students’ science-learning from
different SES backgrounds.

**International Comparison**

The Programme for International Student Assessment (PISA) is a triennial international
survey conducted by the Organization for Economic Co-operation and Development (OECD)
since 2000. The PISA 2006 was focusing on the evaluation of science education to understand
whether 15-year-old students are prepared to have scientific literacy for life in modern society.
There were 30 OECD countries (e.g., the U.S.) and 27 non-OECD countries (e.g., Taiwan) in the
PISA 2006.

The PISA 2006 results showed that Taiwan was the third best-performing country in
science in terms of students’ performance, with an average of 532 score points, while the United
States performed slightly below the OECD average with a score of 489 (OECD, 2007). The U.S.
students’ scores were 49 point difference associated with one standard deviation in SES
background, and SES explained 17.9% variance in student performance; Taiwan was 42 point
difference associated with one standard deviation in SES background, and SES explained 12.5%
variance in student performance (OECD, 2007). In terms of students’ science performance, Taiwan ranked among the top of the participating countries and the United States ranked among the middle of the participating countries. In terms of students’ SES-associated achievement gap, the United States ranked among the top of the participating countries and Taiwan ranked among the middle of the participating countries. Based on this coincidently opposite ranking, it would be interesting to compare these two countries to see how reform-based science teaching has been implemented and how reform-based science teaching influenced education equity.

**Purpose of Study**

The purpose of this study was to investigate whether reform-based science teaching promoted education equity through being available and beneficial for students from different SES backgrounds in the U.S. and Taiwan. The data were obtained from American and Taiwanese 15-year-old students who participated in the PISA 2006. This study was tri-folded, and each section was designed to answer a specific research question.

**Implementation of Reform-Based Science Learning**

The first section was to understand the subgroups that represented to what extent students learned science through reform-based learning activities in the U.S and Taiwan. A latent profile analysis (LPA) was used to classify students into different science learning subgroups based on frequency and types of the reform-based learning activities in science classroom. That is, different latent subgroups would represent different usage and foci of the learning activities that students experienced in the classroom.

**Availability of Reform-Based Science Learning**

The second section was to investigate whether reform-based science learning is equally available for all students regardless of their SES backgrounds in the U.S and Taiwan. Latent
Class Regression (LCR) and Structural Equation Modeling (SEM) were used to examine whether there was a relationship between students’ SES and their frequency to learn science through reform-based activities. The LCR and SEM were conducted in order to address the availability issues using different approaches.

**SES Achievement Gap and Reform-Based Science Learning**

The third section was to investigate whether reform-based science learning helped to minimize the SES-associated achievement gap among students in the U.S and Taiwan. Regression Mixture Modeling and SEM were used to examine whether there was a relationship between the frequency of students learning science through reform-based activities and SES-associated achievement gaps. Similarly, two approaches, regression mixture modeling and SEM, were used to address the SES achievement gap issue.

**Definitions**

In order to clarify the meaning of the terms used in this study, the following definitions are provided.

- "Apply different phenomena" refers to teachers explaining how a school science idea can be applied to a number of different phenomena.
- "Choose own investigations" refers to students being given the chance to choose their own investigations.
- "Class debate" refers to debate or discussion in the classroom.
- "Discussion topics" refers to students’ discussions about the topics.
- "Design for lab" refers to students being required to design how a school science question could be investigated in the laboratory.
• "Design own experiments" refers to students being allowed to design their own experiments.

• "Do experiments" refers to students spending time in the laboratory doing practical experiments.

• "Draw conclusions" refers to students being asked to draw conclusions from an experiment they have conducted.

• "Explain ideas" refers to students being given opportunities to explain their ideas.

• "Explain relevance" refers to teachers explaining the relevance of broad science concepts to our lives.

• "Follow instruction" refers to students doing experiments by following the instructions of the teacher.

• "Hands-on activities" refer to teaching and learning activities that provide students with the opportunity to carry out scientific tasks such as conducting experiments, collecting scientific data, making analyses, and drawing conclusions from their results.

• "Interactive science teaching" refers to students’ collaboration, student discussions, and teacher-student interactions.

• "Real-life applications" focuses on learning science in real-life contexts and applying scientific knowledge to solve real-life problems.

• "Reform-based Science Teaching" refers to four reform-based teaching dimensions that the PISA 2006, based on the basic aspects of science learning and teaching, considered to be important for students’ development of scientific literacy: hands-on activities, student investigations, interactive teaching, and real-life applications.
- "Science achievement" is the science assessment in the PISA 2006 that focused on assessing three scientific competencies: identifying scientific issues, explaining scientific phenomena, and using scientific evidence.

- "Science self-belief" is a term to include science self-efficacy and science self-concept.

- "Science self-concept" refers to students’ self-perceived general academic ability in science.

- "Science self-efficacy" refers to students’ self-perceived confidence to succeed in specific scientific tasks, science courses, or science-related activities.

- "Society relevance" refers to teachers using examples of applied technology to show how school science is relevant to society.

- "Socioeconomic status (SES)" is based on the definition used for PISA study that is a combination of economic capital (e.g., family wealth and possessions), and social and cultural capital (e.g., parents’ occupational status and parents’ educational status).

- "Student investigations" involve students in the broad process of scientific research in which students identify their own research question, test their own ideas, use science process skills, and design research processes for their own scientific research.

- "Student opinion" refers to lessons involving students’ opinions about the topics.

- "Test own ideas" refers to students being asked to do an investigation in order to test their own ideas.

- "Understand world outside" refers to teachers using science to help students understand the world outside of school.
Assumptions

There are three assumptions for this study. They are:

1. Students’ reported science teaching and learning activities reflect classroom science teaching accurately.
2. The results of PISA are reliable and valid and accurately represented for each country.
3. The data is multivariate normally distributed and the missing data are completely random.

Summary

This chapter describes what motivated the author to conduct this research. It highlights the significance of this study, supported by the challenges and issues faced in science classrooms. The purposes of this research are also described in this chapter and are elaborated on in further chapters and are used to organize the methods, results, and discussions. The definitions of the terminology used and the assumptions of this study are also introduced in this chapter. Chapter Two presents the review of the literature related to SES, reform-based science teaching, and science achievement. Chapter Three presents the research questions, hypotheses, descriptions of PISA 2006 data, the measure of each variable, the data analytic procedures, and the limitations of this study. Finally, Chapters Four and Five present the results and discussion respectively.
CHAPTER TWO
LITERATURE REVIEW

This chapter presents the theoretical framework: Social Cognitive Theory and describes the proposed conceptual model in this study. A thorough summary of the literature are presented to show relevant studies in the areas of 1) social cognitive theory (i.e., relationship between socio-economic status (SES) and personal variables, environmental variables, and behavior outcomes), 2) reform-based science teaching, and 3) the implementation of reform-based science teaching in the U.S. and Taiwan. This chapter also describes specific issues related to measures of SES and using Programme for International Student Assessment (PISA) to conduct internal comparison studies.

Theoretical Framework: Social Cognitive Theory

Social cognitive theory (Bandura, 1986, 1997) was developed to explain human functioning. The key concepts of the theory, such as outcome expectations, self-efficacy, reinforcement, and self-control, have significant implications for teaching and learning. Social cognitive theory takes the perspective that individuals’ learning is affected by the interaction of personal, environmental, and behavioral determinants (Bandura, 1989). Thus, in social cognitive theory, person, behavior, and environment are viewed as inseparable and interactive in creating learning behaviors, and the theory depicts these interactive and triadic reciprocal determinants that influence individuals’ learning.

Schunk (1999) adapted the social cognitive theory to develop a dynamic model of achievement (Figure 1), which emphasized the constant interaction among personal influences, social/environmental influences, and behavioral outcomes. According to Schunk, social factors are the primary force in individuals’ early stages of learning because learners take in abundant
information from the social environment. Once learners gradually internalize skills and strategies, they adapt to the environment in unique ways. In individuals’ later stages of learning, learners increasingly structure their social environments to make this environment more beneficial in advancing their learning and skills. Thus, the strength of association and primary directions of these factors vary as individuals develop across time. The model of social-self-outcome interaction allows researchers to understand how students from different SES backgrounds have different learning outcomes. The Model of Social-Self-Outcome Interaction in Achievement Settings (Schunk, 1999) in presented in Figure 1, and the concept model for this current study is presented in Figure 2.

Figure 1. Model of Social-Self-Outcome Interaction in Achievement Settings. Adapted from “Social-self interaction and achievement behavior” by D.H. Schunk (1999)
Figure 2. Conceptual Model Proposed for this Study

**SES, Learning Environment, and Learning Outcomes**

In social cognitive theory (Bandura, 1986), environment determinants refer to the external or situational factors that influence individuals’ behaviors. In the Model of Social-Self-Outcome Interaction in Achievement Settings (Schunk, 1999), the environment determinants include role-models, feedback, and classroom instruction. They also include the use of reward or punishment to modify behaviors and exposure to available tools, resources, and supports facilitating learning. All of these environment determinants are closely related to SES factors and have strong effects on students' learning. Family is the first environment to which children were exposed. Families differ in SES, which is determined by capital, such as financial or material resources (e.g., income), human or nonmaterial resources (e.g., education), and social resources (e.g., social networks and connections) (Bradley & Corwyn, 2002). In terms of financial and material resources, poor families have less access to educational resources than children coming from higher SES families, as differing access is thought to be the reason for differing student
performance. In terms of human or nonmaterial resources, Majoribank (1996) showed that high SES is associated with more parental involvement and that better parenting is positively associated with children’s learning and development (Bradley & Corwyn, 2002). Finally, from the social cognitive theory perspective, the peer modeling related to social resources is also critical to learning. Individuals from a low-SES environment lack successful role models. Muijs, Harris, Chapman, Stoll and Russ (2009) found that there tended to be a high level of unemployment in low-SES communities. Lacking successful role models in the environment is likely to shape individuals’ goals, motivation, and learning.

Morgan et al. (2009) found that children from low-SES households and communities developed academic skills more slowly as compared to children from higher SES groups (Morgan, Farkas, Hillemeier & Maczuga, 2009). Similarly, Aikens and Barbarin (2008) found that children from low-SES environments acquired language skills more slowly. These children tended to exhibit delayed letter recognition and phonological awareness and were at higher risk for reading difficulties. In terms of mathematics learning skills, Coley (2002) found that children with higher SES backgrounds were more likely to be proficient on tasks of addition, subtraction, ordinal sequencing, and mathematics word problems than children from lower SES backgrounds. All of these studies highlight the significant influence of family SES on students’ development.

In addition to the family environment, students were also exposed to school environment. Peske and Haycock (2006) found that youths from poor families were exposed to an environment with less experienced teachers and less academic support. Similarly, a study by Aikens and Barbarin (2008) found that the school systems in low-SES communities were often under-resourced, and in turn this negatively affected students’ academic progress. When considering school neighborhood environment, the students in urban schools performed significantly lower
on national and international science achievement tests than those in low-poverty schools and high-poverty nonurban schools (NCES, 2006). Students with low-SES family background continued to learn in under-resourced schools with less experienced teachers. Therefore, students’ SES affects their learning experiences and learning outcomes throughout most of their learning period.

There have been many studies regarding the effect of SES on students’ achievement. White (1982) conducted a meta-analysis of studies published prior to 1980, examining the relationship between SES and academic achievement. The meta-analysis consisted of 101 studies yielding 636 correlation coefficients estimating the strength of these reported coefficients. White (1982) found that when SES was measured at the student level with these most common indicators--income, occupation, and education level--a weak relationship between SES and academic achievement was found and accounted for less than 5% of the variance at the student level (r= 0.22).

More recently, Sirin (2005) conducted a meta-analysis of SES-achievement research focusing on data from 1990 to 2000. The meta-analysis consisted of 75 samples, and 207 correlation coefficients were coded. The results showed correlations ranging from 0.005 to 0.77 with a mean of 0.29 (SD = 0.19) and a median of 0.24 among these samples. Sirin also found that the strength of these relationships varied significantly and were moderated by factors such as types of SES measure, student characteristics, grade, and minority status.

Reform-Based Science Teaching

Although students from low-SES family environments have disadvantages in science learning, our core belief is that schools should provide equitable learning environments for all
students regardless of their SES background and that is the goal of science education reform (NRC, 1996). The central features of reform-based science teaching and research regarding the efficacy of reform-based science teaching for low-SES students' learning are summarized below.

Based on current international research for effective science teaching and science education reform documents (AAAS, 1989; NRC, 1996), the PISA 2006 Questionnaire Expert Group summarized four science teaching and learning dimensions that provides opportunities for students to engage in science and develop students’ scientific literacy: hands-on activities, student investigations, interactive teaching, and real-life applications.

**Hands-on activities.** Hands-on activities became the major focus in science teaching since the curriculum reform movement resulted from the successful launch of the Soviet’s Sputnik I in 1957. Many elementary science programs being supported by the NSF, such as Science-A Process Approach (SAPA), Science Curriculum Improvement Study (SCIS), and Elementary Science Study (ESS) were designed to emphasize the laboratory method of science teaching and to focus on developing elementary students’ basic skills in the processes of science. These processes included observing, classifying, measuring, predicting, etc. The developers of these programs believed that by teaching students scientific process skills, they would not only learn science content knowledge, but would also improve learning skills that students could use in new subjects and situations (Gagne, 1967). Hands-on activities were supported by constructivism and developers believed that by doing hands-on activities, students would get direct sensory experiences. These experiences provided opportunities for students to take external data into their cognitive structures through sensory data or to modify their cognitive structures to match the external data as a result of disequilibrating experiences (Saunders, 1992). In this study, hands-on activities refer to teaching and learning activities that provide students
with the opportunity to carry out scientific tasks such as planning and conducting an experiment in a laboratory, collecting scientific evidence, and drawing conclusions from their results (Harlen, 1999).

Hands-on activities were found to be helpful for students' science learning. Stohr-Hunt (1996) investigated the relationship between the amount of time students engaged in hands-on activities and science achievement. Using the data from the longitudinal data of the National Educational Longitudinal Study of 1988 (NELS: 88), he reported that eighth-grade students who engaged in hands-on activities every day or once a week had significantly higher scores on a standardized science achievement test than students less frequently engaged in hands-on activities.

**Student investigations.** Student investigations went beyond the idea of hands-on activities and aimed to involve students in the broad process of scientific research, in which students identified their own research question, tested their own ideas, used science process skills, and designed scientific research by themselves (Hofstein & Lunetta, 2004). Investigation activities provide students with more freedom in the process of scientific inquiry. The impact of student investigations on individuals’ science understanding has been investigated. In the science education literature, student investigation sometimes referred to open inquiry-based science teaching (NRC, 2000). A review by Minner, Levy, and Century (2010) showed positive impacts of inquiry-based teaching on students’ science understanding. Among the 138 studies that have investigated inquiry-based teaching, 71% of the studies reported findings with positive effects, 33% of studies showed mixed effects, and 14% of studies showed no effect. Harlen (1999) suggested that the quality of how science was represented by inquiry or investigations was more important than whether students implemented scientific inquiry or investigation.
Interactive science teaching. Science educators agreed that science teaching should emphasize interactive learning activities (Hofstein & Lunetta, 2004). Traditional science teaching primarily focuses on transmitting science knowledge through lectures, textbooks and demonstrating experiments. In contrast, interactive science teaching was oriented towards cooperative learning, student discussions, and collaborative teacher-student interactions and focused on involving students in classroom discourse about scientific topics. Donovan and Bransford (2005) suggested that science teaching should support students to communicate scientific concepts and experience the importance of discourse in scientific inquiry. Schroeder, Scott, Tolson, Huang, and Lee (2007) conducted a meta-analysis of different teaching strategies for science achievement and found that questioning strategies and collaborative learning strategies had the greater positive effect on students’ achievement and were among the highest effect size for different teaching strategies.

Real-life applications. Real-life applications helped students understand the relevance of classroom scientific concepts to the real world and provide them with the ability to use their science knowledge when encountering problems in real-life context. Today, young people encounter more science and technology in different environments than previous generations. Real-life applications provide opportunities for authentic contexts where science is applied to help young people in their development of scientific literacy. Ramsden (1997) investigated the effects of context-based curriculum on student learning and he found that the high school students in the context-based curriculum group had a greater understanding of science concepts than the students in the traditional curriculum group.

Effectiveness of reform-based science Teaching. Some researchers studied the relationship between science teaching and learning and student achievement using the PISA
2006 dataset. Valente, Fonseca, and Conboy (2011) investigated this relationship with the data from Portugal, Spain, France, United Kingdom, Turkey, U.S., Greece and Finland. They found that students reporting higher frequency of student investigations and interactive teaching tended to have lower achievement in science. On the other hand, students reporting higher frequencies of learning science through focusing on real-life application tended to have higher science achievement scores for all of those countries. Students’ learning through hands-on activities had mixed results in regards to science achievement for the different countries. Areepattamannil, Freeman, and Klinger (2011) conducted a similar study with the data from Canada. They found that while science teaching using hands-on activities had a substantial positive effect on science achievement, science teaching using student investigations had a substantial negative effect on science achievement. The research for the Finnish students had a similar result, namely, that student investigation had a negative correlation with students’ performance on the PISA science assessment (Lavonen & Laaksonen, 2009). Given the literature above suggesting the benefits of reform-based teaching for science leaning, these findings of negative correlation between reform-based teaching and achievement seemed counter-intuitive. Some researchers gave two explanations for this counter-intuitive result: 1) reform-based teaching “does not seem to be an effective instructional technique if the goal is limited to increasing proficiency achievement as measured by PISA” (Areepattamannil, Freeman, & Klinger, 2011, p. 249), and 2) “pupils perceive as being positive the fact that new concepts are introduced by a teacher, an expert, who first presents new information and then demonstrates how this information is used for solving problems or performing tasks” (Lavonen & Laaksonen, 2009, p.937).


**SES and Reform-Based Science Teaching and Learning**

These reform-based teaching strategies required students to collaborate with their classmates, actively manipulate experimental materials, think deeply about complex concepts, relate new science content to their daily lives outside school, and self-regulate their behavior. These requirements brought challenges to students and more challenges to students from low-SES families with limited education resources and lower school performance. Geier, Blumenfeld, Marx, Krajcik, Fishman, Soloway and Clay-Chambers (2008) summarized these challenges for low-SES students, including “lack of resources, high levels of poverty, low student achievement, below grade level English proficiency, high student mobility, attendance problems and difficulty in recruiting and retaining highly qualified teachers” (p 923).

While many teachers may hesitate to use reform-based teaching for low-SES students in urban school district, researchers at the University of Michigan collaborated with the Detroit Public Schools to implement a science education reform project (Krajcik, Marx, Blumenfeld, Soloway, & Fishman, 2000). Spanning three years (1999-2002), the project was completed for over 2500 students at 15 different Detroit public middle schools located in lower SES neighborhoods. Over 91% of students were African-American, 70% of students receive free or reduced-price lunches, and 85% of students were reported below grade level at the statewide standardized eighth grade science assessment. The researchers designed project-based curriculum materials that contextualized learning science through real-world problems, engaged students in scientific inquiry, and used learning technologies. The results showed that students not only had significant and consistently high learning gains in science achievement tests designed with close proximity to the curriculum materials (Rivet & Krajcik, 2004), but also showed significantly higher passing rates on the statewide high-stakes test (Geier et al., 2008).
They suggested that when curriculum was carefully developed, assessments were aligned with curriculum materials, and when professional development was designed to change teachers' traditional practice, low-SES students who historically were low achievers in science could successfully learn science content and pass standardized science achievement tests (Geier et al., 2008).

There was another large-scale science education intervention program focused on culturally and linguistically diverse elementary students (Lee, Deaktor, Enders, & Lambert, 2008). From 2001 to 2004, this program was implemented for third, fourth, and fifth graders at six elementary schools in a large urban school district in which 70% of the students participated in free or reduced price lunch programs. The curriculum development was guided by the national policy documents on science education standards (AAAS, 1989, 1993; NRC, 1996). The results showed that there were statistically significant increases on the mean scores of tests between pre- and posttests for all three grade levels. Additional research that focused on students’ inquiry skills (Lee, Buxton, Lewis, & LeRoy, 2006) indicated that by designing learning environments that fostered science inquiry for diverse background students, students from less privileged backgrounds showed greater gains in inquiry abilities than those from more privileged background.

Other studies also found that different components of reform-based teaching successfully improved low-SES students learning. For example, Warren et al. (2001) found that when students work on scientific inquiry, teachers can identify intersections between students’ everyday knowledge and scientific practice and use these intersections to inform their science teaching practices. The results showed that low-income immigrant students with limited science experience were successful in regards to scientific inquiry, reasoning, and argumentation. Other
studies showed the benefits of real-life application for science learning of low-SES students. Rodriguez (1997) thought that because school science was usually disconnected from the multicultural contexts of students’ lives, especially for urban school students, students opted not to engage in science. To address this problem, Mallya, Mensah, Contento, Koch, and Barton (2012) designed a food science curriculum for urban schools where youth obesity was a problem. They found that by creating meaningful and relevant learning opportunities for students through connecting school science with issues of personal and social significance in students’ lives outside of school, students not only could make healthier food choices, but they also valued learning science as a way to transform their living.

On the other hand, there were negative results in regards to reform-based teaching. Wilson, Taylor, Kowalski, and Carlson (2010) conducted a randomized experimental study to compare the effects of implementing inquiry-based teaching organized around the Biological Sciences Curriculum Study (BSCS) 5E Instructional Model and traditional science teaching strategies based on a national teacher survey data for students from different race and SES backgrounds. The result showed that the difference between achievement gain of pre- and post-test between free or reduced lunch free (FRL) students and non-FRL students was larger for the inquiry group than for the traditional teaching group. This may indicate that inquiry teaching was more effective for students from high-SES backgrounds.

**SES, Self-Belief, and Achievement**

Based on the social cognitive theory (Bandura, 1986), personal determinants are cognitive or psychological behaviors that influence an individual’s behaviors. In the Model of Social-Self-Outcome Interaction in Achievement Settings, the personal variables include goals, self-efficacy, outcome expectations, attributions, progress self-evaluation, and self-regulatory
progress (Schunk, 1999). These personal variables explained individuals’ learning behaviors and learning outcomes. Specifically, self-efficacy is a frequently studied concept and the most important predictor for achievement among these personal variables (Pajares & Shunk, 2002) and, thus, was selected as a personal variable for this study.

Bandura (1997) defined self-efficacy as “belief in one’s capabilities to organize and execute the courses of action required to produce given attainments” (Bandura, 1997, p.3). The PISA 2006 (OECD, 2007) used the term "science self-beliefs" to include self-efficacy and self-concept, a similar concept, as an important factor correlated with students’ science achievement. Self-efficacy refers to students’ self-perceived confidence to succeed in specific scientific tasks, science courses, or science-related activities (Britner & Pajares, 2006). On the other hand, self-concept refers to students’ self-perceived general academic ability (Bong & Skaalvik, 2003). Pajares and Schunk (2002) suggested that self-concept is formed by asking questions of being and feeling (e.g., “Am I any good at science?”), whereas self-efficacy is formed by asking “can” questions (e.g., “Can I do this science question?”). Moreover, whereas self-efficacy is more future-oriented and changeable, self-concept is more past-oriented and stable (Bong & Skaalvik, 2003).

Family environment provides initial experiences that influence children’s self-efficacy. Parents in high-SES families tend to better motivate their children to reach certain goals, be better role models for children, and know how to teach tactics that they can use to overcome challenges, and all of this enhances self-efficacy. Blenkinsop et al. (2006) argued that higher SES status parents expressed aspirations that were passed on to their children regarding their academic capability. Alexander and Entwisle (1988) showed that family income are positively associated with parents’ expectations for their child’s educational success. Régner, Huguet, and
Monteil (2002) showed that individuals from a low-SES background tended to develop lower learning self-efficacy.

Bandura (1997) suggested that schools should provide opportunities to engage students in tasks and activities, interpret the results of their activities, and use these interpretations to develop self-efficacy about their ability to engage in subsequent tasks or activities. Experience in mastering tasks is the most important source of establishing students’ self-efficacy. Britner and Pajares (2006) found that mastery experiences is a significant predictor for students’ science self-efficacy.

Regarding the relationship between students’ self-efficacy and achievement, Lent, Brown, and Larkin (1984) found that self-efficacy influences motivation, effort, and persistence in solving problems, and has positive association with achievement. Kupermintz (2002) showed that in high school students, science self-efficacy correlated with science achievement and was a better predictor for achievement than the other factors, such as gender, ethnicity, and parental background. From the results of PISA 2006 (OECD, 2007), students’ self-efficacy and self-concept had the most positive relationship with science achievement among all science–related affective and motivational characteristics variables in both U.S. and Taiwan. Students increased approximately 35 points per unit increase in self-efficacy index in both countries, and increased around 30 points for the U.S. students and 15 points for Taiwanese students per unit increase in self-concept index (OECD, 2007).

**Implementation of Reform-Based Science Teaching and Learning**

**United States**

How is reform-based science teaching implemented as current practice in the U.S.?

Horizon Research, Inc. implemented two large scale studies in 2000 (Hudson, McMahon &
Overstreet, 2002) and 2012 (Weis, 2013) to understand the status of middle school science teaching. In the 2000 study (Hudson, McMahon & Overstreet, 2002), the researchers surveyed 529 middle school science teachers and found that two-thirds of middle school science teachers indicated they were at least somewhat familiar with the NSES (NRC, 1996). Moreover, among those teachers familiar with the national standards, three-fourths of them indicated that they had implemented the national standards in their classroom at least to a moderate extent. For various instructional activities, 80% of teachers included classroom discussion, and 67% of teachers included a hands-on/laboratory science activity or investigation at least weekly. However, students were much more likely to follow specific instructions in completing an activity or investigation (76%) than to design or implement their own investigations (17%). In 2012 study (Weis, 2013), the researcher surveyed 958 middle school science teachers and found that whole class discussion occurred in 90% of classes and that hands-on/laboratory activity occurred in 62% of science lessons. The implementation of these reform-based science learning activities were similar between the 2000 and 2012 studies.

Contrary to the results from teachers’ self-report data, the observation study suggested that even with the emphasis on the importance of reform-based science teaching in the standards (NRC, 1996), reform-based teaching was still not common teaching practice in science classrooms. Simmons et al. (1999) found that only 10% of 116 beginning secondary teachers implemented student-centered instructions in their first year of teaching. Roehrig and Luft (2004) found that only 4 of 14 beginning secondary science teachers implemented more reform-based science lessons. The majority of them taught science through more traditional ways such as asking students to find the right answers by checking textbooks. This low incidence of reform-
based teaching in beginning science teachers’ classrooms raised an important issue regarding the evaluation of teachers' implementation for reform-based teaching.

Taiwan

In Taiwan, teachers always rely on science textbooks. Thus, evaluating the implementation of reform-based science curriculum in science classes starts with evaluating the textbooks. Tuan’s study (Abd-El-Khalick et al., 2004) described how the extent to which inquiry was addressed in the textbooks varied widely for the different grade levels. In elementary school levels, science textbooks were oriented mainly around hands-on and focused on students’ process skills. In middle school, in each chapter, textbooks provided a hands-on activity that students could conduct and discussed natural phenomena in the first session. Then, in the second session, the textbooks presented science content knowledge that related to the hands-on activity. In high school, science textbooks and laboratory manuals were separate. The science textbooks only focused on content knowledge and the manuals provided verification-style experiments that were designed to reinforce what students learned in the science textbooks. The textbooks in Taiwan currently focused less on reform-based curriculum at the secondary level than they did at the elementary level.

Wang and Lin (2009) compared differences between elementary classroom and middle school classroom and found that while elementary school had more classroom discussion, student presentation, and even role-playing activities, middle school classrooms exclusively focused on traditional teaching. The science teachers dominated a science class in which they delivered science knowledge to students through explanations, demonstrations, reviewing, test administration, and post-test discussions. Occasionally, following classroom instruction, students had the opportunity to participate in hands-on activities, in which they observed natural
phenomena and recorded what they observed based on instructions in the laboratory manuals. Students have little opportunity to express themselves in the classroom.

Wang and Lin (2009) conducted a national survey and evaluated various dimensions of the learning environment perceived by elementary and middle school students in Taiwan. They listed scientific inquiry as one of the dimensions of learning environment. They evaluated the extent to which students were involved in science classrooms in terms of developing and evaluating experiments, evidence, hypotheses, explanations, models, and arguments. Compared to the other dimensions, such as conceptual understanding of science, scientific inquiry was used less frequently in both the elementary and middle school science classrooms. Based on these research studies, Taiwanese science teachers did not emphasize students’ inquiry skills, and the inquiry activity was not the major learning activity for students in science classroom.

**Measures of Socio-Economic Status (SES)**

Although SES has been at the core of many different fields of research, there seems to be no widespread consensus in its definition and measure. For example, in the U.S., most of education research used student’s eligibility for free or reduced-price school lunch as an indicator of SES (Wilson et al., 2010). In general, for the research related to education policy and social science, educational attainment, occupational status, and income or wealth were the most common indicators to measure SES (Buchmann, 2002).

For measuring educational attainment in international comparative research, a cross-national comparative scale was needed. The International Standard Classification of Education (ISCED) and Comparative Analysis of Social Mobility in Industrial Nations (CASMIN) were developed to be cross-national comparative scales. The CASMIN categories were developed in Germany for the purpose of facilitating comparative research on social stratification and
mobility. The CASMIN scale has more education attainment choices by distinguishing vocational credentials from general or academic credentials in secondary and tertiary education (Buchmann, 2002). The ISCED was developed and regularly used by United Nations Educational, Scientific and Cultural Organization (UNESCO), and other international organizations for reporting national education statistics. The PISA used the number of years of education according to ISCED classification to indicate students’ father and mother’s educational attainment (OECD, 2009).

Occupational status was measured via developed scales that have the prestige score associated with occupations across a wide range of occupations. The earliest version of occupation measure was the U.S. Socioeconomic Index (SEI) scale formulated by Duncan (1961), and subsequent modified versions were used by other researchers for other countries. In order to develop internationally comparative scales of occupational prestige for international comparative research, based on Duncan’s scale, two scales were developed: the Standard International Occupational Prestige (SIOP) scale (Treiman, 1977) and the International Socioeconomic Index (ISEI) of occupational status (Ganzeboom, DeGraaf, & Treiman, 1992). The PISA used ISEI to measure students’ father and mothers’ occupation status (OED, 2009).

Home possessions as an indicator for family wealth has received more attention in international studies in the field of education because home possessions are believed to reflect a more stable source of wealth (Buchmann, 2002). In the PISA, home possessions included family wealth possessions (i.e. cars, computers, and televisions), cultural possessions (i.e. works of art and classical literatures), and home educational resources (i.e. study desks, computers students can use for school work, educational software) (OECD, 2009).
International Comparison

Founded in 1961, the Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organization with the goal to promote policies that will improve the economic and social well-being of people around the world. While the OECD is basically concerned with economic policy, education has taken on increasing importance because the new emerging topic of knowledge economy as the issues of human capital has become an important factor for national economic competitiveness (OECD, 2007).

The OECD evaluated the education system worldwide through conducting the Programme for International Student Assessment (PISA). Its purpose was to examine the knowledge and skills for 15-year olds because it was the last year of compulsory education in the majority counties around the world. The assessment tasks were the literacies associated with reading, mathematics, and science. The PISA used the term "literacy" to encompass a broad range of competencies relevant to coping with real life. Rather than examining students’ mastery on the school curricula, its main purpose of assessment was to evaluate young people’s ability to practically apply their knowledge and skills to cope with problems occurring in everyday life situations (OECD, 2009).

The PISA is a triennial international survey since 2000. While the assessment always tested for literacy, mathematical literacy, and scientific literacy, the assessment had a main focus each time. In 2000, PISA focused on students’ reading skills; in 2003, PISA focused on students’ mathematical ability; and PISA focused on science literacy in 2006. And in 2009, PISA started a new cycle and focused on students’ literacy again (OECD, 2009). The domains assessed for the scientific literacy were

an individual’s scientific knowledge and use of that knowledge to identify questions, to acquire new knowledge, to explain scientific phenomena, and to
draw evidence-based conclusions about science-related issues, understanding of the characteristic features of science as a form of human knowledge and enquiry, awareness of how science and technology shape our material, intellectual, and cultural environments, and willingness to engage in science-related issues, and with the ideas of science, as a reflective citizen (OECD, 2006, p. 12).

Based on the assessment domain, the goal of the PISA 2006 science assessment is “focus(ing) on competencies that clarify what 15-year-old students know, value, and are able to do within reasonable and appropriate personal, social, and global contexts” (OECD 2006, p. 21).

The PISA 2006 assessed students’ science learning outcomes for four aspects: “1) Context: recognizing life situations involving science and technology; 2) Knowledge: understanding the natural world on the basis of scientific knowledge that includes both knowledge of the natural world and knowledge about science itself; 3) Competencies: demonstrating competencies that include identifying scientific issues, explaining phenomena scientifically, and drawing conclusions based on evidence; and 4) Attitudes: indicating an interest in science, support for scientific enquiry, and motivation to act responsibly towards, for example, natural resources and environments” (OECD 2006, p. 25).

Along with the assessment for the participating students, the survey information from the student questionnaire and the school principal questionnaire was also collected. The student questionnaire gathered general background information (e.g., age, gender, educational resources at home, and school-related activities) as well as asked students a number of questions relating to the main focus of the assessment. The school principal questionnaire surveyed information about instructional practices, school organization, school decision-making, school resources, school climate, and teacher and school autonomy. The PISA adopted established and professionally recognized principles of sampling to ensure that the sampled students represented the full target population of 15-year-old students (OECD, 2009).
In the PISA 2000, there were 43 countries participating in the assessment. For countries’ mean score comparison, Japan and Korea showed the highest performance on the scientific literacy scale. The scores of Australia, Austria, Canada, the Czech Republic, Finland, Ireland, New Zealand, Sweden and the United Kingdom were significantly higher than the OECD average. The U.S. science scores were not significantly different from the OECD average. While the mean scores for white students were significantly higher than the OECD average, the mean scores for Hispanic and black students were significantly lower than the OECD average. There is no gender difference on science scores in the U.S. Taiwan did not participate in the PISA 2000. (OECD, 2001).

In the PISA 2003, there were 41 countries participating in the assessment. Finland, Japan, Korea, and Hong Kong had the highest scores in science assessment. Australia, Belgium, Canada, the Czech Republic, France, Ireland, the Netherlands, New Zealand, Sweden, Switzerland, Liechtenstein, and Macao had science scores above the average score. The U.S. science scores were significantly lower than the OECD average. Similar to the results in the PISA 2000, the mean scores in 2003 for white students were significantly higher than the OECD average and the mean scores for Hispanic and black students were significantly lower than the OECD average. There was no gender difference on science score in the U.S., and Taiwan did not participate in the PISA 2003 (OECD, 2004; Lemke et al., 2004).

In the PISA 2006, there were 57 countries participating in the assessment. Finland’s students performed clearly ahead of students in all other countries. Canada, Japan, New Zealand, Australia, Hong Kong, Taiwan, and Estonia students’ scores were well above the OECD average. The U.S. science scores showed no significant difference from the OECD average. As the previous results, the mean scores for white students were significantly higher than the OECD
average and the mean scores for Hispanic and black students were significantly lower than the OECD average. The gender differences in science performance tended to be small in comparison to reading and mathematics performance. The majority of the countries, including the U.S. and Taiwan, showed no statistically significant differences between males and females (OECD, 2007; Baldi et al., 2007).

**Summary**

This chapter introduces social cognitive theory as the theoretical framework. Under this theoretical framework, the relationship between SES and three factors--reform-based science teaching, self-belief, and science achievement--are described. As documented in the literature, individuals’ SES and self-belief play a significant role in their learning process and learning outcomes. This chapter also presents the proposed conceptual model and how this study views the relationship among reform-based learning activities, individuals’ self-beliefs, SES, and learning outcomes. Finally, this chapter describes currently implementation of reform-based teaching in science classrooms in the U.S. and Taiwan, and presents past studies of using the PISA for international comparisons.
CHAPTER THREE

METHODOLOGY

Chapter Three describes the research design of the current study. The first section describes the research questions and the nature of the data from the Programme for International Student Assessment (PISA) 2006. The second section outlines the measurement, variables, and the data analytic procedures used to answer each research question. Finally, the third section describes the limitation of this study and offers a brief summary of this chapter.

Research Questions and Hypotheses

The main purpose of this study was to investigate how reform-based science learning has been implemented and whether reform-based science learning promoted education equity by being available and minimizing the achievement gap for students from different socioeconomic status (SES) backgrounds in the U.S. and Taiwan. This study was tri-folded, and each section was designed to answer a specific research question. The research questions and analytic techniques used are described below.

Implementation of reform-based science learning

RQ 1: How broadly has reform-based science learning been implemented in classrooms in the U.S. and Taiwan?

This research question was empirically examined using latent profile analysis (LPA) to classify students into different science learning subgroups based on the usage and foci of learning activities in their science classrooms. Each science learning subgroup was statistically representative of the extent to which students learned science through reform-based learning in classrooms. The results of LPA model were used to understand the nature and prevalence of science learning activities in the U.S. and Taiwan.
Availability of Reform-Based Science Learning

RQ2: Was reform-based science learning available for all students regardless of their SES background in the U.S. and in Taiwan? Two different analyses, Latent Class Regression (LCR) and Structural Equation Modeling (SEM), were used to answer this research question. The following hypotheses was tested separately for the U.S. and Taiwan.

H₀1: Students’ SES is not significantly associated with an increase (or decrease) in the odds of membership in a specific science learning subgroup relative to a most usage reform-based science learning subgroup.

Analysis 1. Building on the results of the LPA from the first research question, students’ SES was added as a covariate (LCR) to examine associations between students’ SES and their science learning subgroup membership. These associations were modeled using a multinomial logistic regression to show how students’ SES predicted their subgroup membership. This study hypothesized that students with higher SES would have more opportunity to learn science through reform-based learning activities.

H₀2: There is no statistically significant association between students’ SES and their frequency to learn science through each of four dimensions of reform-based science learning (i.e., hands-on activities, student investigation, interactive teaching, and real-life application).

Analysis 2. After examining reform-based science learning as a whole construct, this study was individualized with more detail to investigate each of four science learning dimensions. This research question was empirically examined using structural equation modeling (SEM) to investigate the associations between students’ SES and each of the four dimensions of science learning. This study hypothesized that students with higher SES would have more opportunity to learn science through each of four dimensions of science learning.
SES Achievement Gap and Reform-Based Science Learning

RQ3: Did reform-based science learning minimize achievement gap for students from different SES backgrounds in the U.S. and Taiwan? Two different analyses, Regression Mixture Modeling and SEM, were used to answer this research question. The following hypotheses was tested separately both for the U.S. and Taiwan.

H03: There is no statistically significant difference on the influences of students’ SES on their science achievement across different science learning subgroups.

Analysis 1. Regression Mixture models were used to examine the influences of students’ SES on their science achievement within each science learning subgroup. This study hypothesized that students in the subgroup with more frequent usage and more variety choice of reform-based learning activities would have smaller SES-associated achievement gap.

H04: There are no statistically significant associations between the interaction terms between students’ SES and each of four dimensions of science learning (i.e., hands-on activities, student investigation, interactive teaching, and real-life application) and student achievement.

Analysis 2. Building on the results of the SEM, the interaction terms between students’ SES and each of four dimensions of science learning were added to examine associations between these four interaction terms and students’ science achievement. This study hypothesized that students who more frequently learned science through each of the four dimensions of science learning would have smaller SES-associated achievement gaps.

Data Source

The data source for this study was from the PISA 2006. Along with the assessment for the participating students, a number of questions relating to science learning were included in the student questionnaire. These surveys allowed researchers using PISA to interpret and analyze the
processes of educational policy and governance at a national and an international level (OECD, 2009). Students’ background information, confidence in science, and their classroom learning activities in science, as reported in the student questionnaire, as well as students’ test scores in the science assessment, were used as data for this study.

The PISA 2006 U.S. dataset contained 5611 students from 166 schools and the Taiwan dataset contained 8815 students from 236 schools. For the current study, general education was the focus, so students in vocational schools, such as vocational senior high schools and 5-year colleges in Taiwan were excluded from this study. Therefore, the sample size for Taiwan was 5995 students from 179 schools.

Table 1
Student Composition by Gender and Race for the PISA 2006 U.S. and Taiwan Data.

<table>
<thead>
<tr>
<th></th>
<th>U.S. (%)</th>
<th>Taiwan (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>49.3</td>
<td>46.6</td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>57.9</td>
<td>NA</td>
</tr>
<tr>
<td>Black</td>
<td>13.1</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>17.9</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>Multiracial</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1.7</td>
<td></td>
</tr>
</tbody>
</table>

Note: NA=not available

Measurement and Variables

Science Learning and Teaching

Based on current international research for science education and science education reform documents (AAAS, 1989; NRC, 1996), the PISA Questionnaire Expert Group examined four reform-based science learning dimensions: hands-on activities, student investigations, interactive teaching, and real-life applications. These dimensions were the basic aspects of
science learning activities effective for students’ development of scientific literacy. Based on these four dimensions, they developed 15 items for the PISA science teaching and learning survey. Using the questionnaire, students reported their views on the frequency of the 15 learning activities in their science classroom for the current year. Students were asked to rate each item on a 4-point Likert scale (1=in all lesson, 2=in most lessons, 3=in some lessons, and 4=never or hardly ever) (OECD, 2009). The ratings were reverse-coded to show that higher numbers indicated higher frequency and lower numbers indicated lower frequency.

The 15 learning activity items with their dimensions are listed in the following:

**Hands-on activities:**

a) Students spend time in the laboratory doing practical experiments (do experiments),

b) Students are required to design how a school science question could be investigated in the laboratory (design for lab),

c) Students are asked to draw conclusions from an experiment they have conducted (draw conclusions), and

d) Students do experiments by following the instructions of the teacher (follow instructions).

**Student investigations:**

e) Students are allowed to design their own experiments (design own experiments),

f) Students are given the chance to choose their own investigations (choose own investigations), and

g) Students are asked to do an investigation to test out their own ideas (test own ideas).

**Interactive teaching:**

h) Students are given opportunities to explain their ideas (explain ideas),
i) The lessons involve students’ opinions about the topics (student opinion),

j) There is a class debate or discussion (class debate), and

k) The students have discussions about the topics (discussion topics).

Real-life applications

l) The teacher explains how a school science idea can be applied to a number of different phenomena (apply different phenomena),

m) The teacher uses science to help students understand the world outside school (understand world outside),

m) The teacher clearly explains the relevance of broad science concepts to our lives (explain relevance), and

o) The teacher uses examples of technological application to show how school science is relevant to society (society relevance).

The PISA 2006 Technical Report (2009) supplied the validity and reliability information for the four-dimension model of the science teaching and learning items. The model fit for the four-dimensional model was acceptable for the U.S. sample with the root-mean square error of approximation (RMSEA) equaled to 0.094, the root mean square residual (RMR) equaled to 0.049, and the comparative fit index (CFI) equaled to 0.89. The scale reliabilities (Cronbach’s alpha) of these four dimensions for the U.S. sample was 0.75 for “Hands-on activities”, 0.79 for “Student investigations”, 0.80 for “Interactive teaching”, and 0.76 for “Real-life applications” (OECD, 2009). There was no such information for Taiwan sample because the PISA 2006 Technical Report did not supply validity and reliability information for the non-OECD countries.
A total of 130 students in the U.S. sample and six students in the Taiwan Sample were excluded from this study because they did not respond to any of these 15 items about classroom learning activities. The missing rate is 2.31% for the U.S. sample and 0.10% for Taiwan sample.

Socioeconomic Status (SES)

The PISA’s definition of SES combined economic capital (e.g., material resources), cultural capital (non-material resources), and social capital (e.g., social connections). Thus, parents’ education, occupational status and income were chosen as three components of SES. In the PISA 2006, the variable related to parents’ occupational status was “highest occupational status of parents”, in which parents’ occupations were mapped to the international socio-economic index of occupational status (ISEI) (Ganzeboom et al., 1992) and the higher ISEI score of either parent was used for this variable. The variable related to parents’ education levels was “highest educational level of parents”, in which the higher years of schooling of either parent were used. The PISA did not survey income and instead used the variable “household possessions” as an indicator of family wealth. Household possessions were believed to capture family wealth better than income because they reflected a more stable source of wealth (Buchmann, 2002). In PISA 2006, students reported the availability of 16 different household items and how many books were in the home, and their responses were scaled using the one-parameter (Rasch) model of item response theory (IRT) scaling methodology. Finally, the SES scores were obtained as component scores for the first principal component with zero being the score of an average OECD student and the standard deviation of one across equally weighted OECD countries (OECD, 2009). On average, parents in the U.S. had a SES score above average, where as parents in Taiwan had a SES score below average.
Table 2
Means and Standard Deviations of SES and their Three Components for the PISA 2006 U.S. and Taiwan Data.

<table>
<thead>
<tr>
<th></th>
<th>U.S.</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>SES</td>
<td>0.14</td>
<td>0.91</td>
</tr>
<tr>
<td>Highest occupational status of parents</td>
<td>52.53</td>
<td>16.77</td>
</tr>
<tr>
<td>Highest educational level of parents</td>
<td>13.64</td>
<td>2.47</td>
</tr>
<tr>
<td>Household possessions</td>
<td>-0.12</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Science Achievement

In this study, the PISA science achievement score was the outcome variable. The science assessment in the 2006 PISA contained 108 items allocated to three scientific competencies (i.e., identifying scientific issues, explaining scientific phenomena, and using scientific evidence). Because students could not finish all items in addition to the literacy and math items within the two-hour time limit, the PISA 2006 used the matrix-sampling technique. With this technique, the 108 science items in addition to literacy and mathematics items were allocated to 13-item clusters (seven science clusters, two reading clusters and four mathematics clusters) with each cluster representing 30 minutes of test time. The 13-item clusters were distributed and presented to students in 13 test booklets, and each booklet contained four 13-item clusters. Each student was assigned one booklet during test administration. Since each student only responded to one booklet, item response theory scaling methods and the imputation process were used to derive five plausible values for each student to represent their competency as if they had completed the entire assessment (OECD, 2009). The PISA 2006 reported five plausible values for both the science competencies in general and three science competencies in specific. The assessment scale is established by setting the mean of the scale at 500 and the standard deviation at 100 for the pooled and equally weighted OECD countries. The mean science assessment score is 489 for
the U.S. and 532 for Taiwan. On average, students in Taiwan had higher science scores than students in the U.S.

**Table 3**  
*Means and Standard Deviations of Science Assessment for the PISA 2006 U.S. and Taiwan Data.*

<table>
<thead>
<tr>
<th></th>
<th>U.S.</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science Score</td>
<td>489</td>
<td>532</td>
</tr>
<tr>
<td><strong>S.D.</strong></td>
<td>104.7</td>
<td>89.4</td>
</tr>
</tbody>
</table>

**Science Self-Belief**

The PISA 2006 identified two types of science self-belief: (1) students’ specific beliefs of their own ability to handle science-related tasks effectively and to overcome challenges (i.e., self-efficacy) and (2) students’ general beliefs in their own science academic abilities (i.e., self-concept) (OECD, 2006). Pajares and Schunk (2002) suggested that self-concept is formed by asking questions of being and feeling (i.e., “Am I any good at science?”), whereas self-efficacy is formed by asking “can” questions (i.e., “Can I do this science question?”).

The PISA 2006 measured both of these constructs in relation to science achievement through the student questionnaire. Eight tasks measuring students’ science self-efficacy are listed in the following:

(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 formation of acid rain.

Students reported their confidence and ability in solving these questions by rating them on a four-point Likert scale (1=I could do this easily; 2= I could do this with a bit of effort; 3= I would struggle to do this on my own; and 4= I couldn’t do this). The scale reliability (Cronbach’s alpha) is 0.87 for the U.S. sample and 0.85 for Taiwan sample. All items were reverse coded for IRT scaling so that higher scores on this index indicated higher levels of self-efficacy (OECD, 2009). The mean score of the index of science self-efficacy was 0.22 for the U.S. sample and 0.30 for the Taiwan sample. Comparing the U.S. and Taiwan to other countries, the results of PISA 2006 indicated that students in both the U.S. and Taiwan had high self-efficacy, with the U.S. ranked second and Taiwan ranked seventh in 57 countries.

Six items measuring students’ science self-confidence are the following:

(a) Learning advanced school science topics would be easy for me,

(b) I can usually give good answers to test questions on school science topics,

(c) I learn school science topics quickly,

(d) School science topics are easy for me,

(e) When I am being taught school science, I can understand the concepts very well, and

(f) I can easily understand new ideas in school science.

Students answered these items on a four-point Likert scale (1=strongly agree, 2= agree, 3=disagree, and 4=strongly disagree). The scale reliability (Cronbach’s alpha) is 0.93 for both the U.S. and Taiwan sample. All items were reverse coded for IRT scaling so that higher scores on this two index indicated higher levels of self-confidence (OECD, 2009). The mean score for the index of science self-concept is 0.20 for the U.S. sample and -0.36 for the Taiwan sample.
While students in the U.S. still rated higher than the OECD average score, Taiwan was ranked 3rd from the bottom (OECD, 2007).

Table 4

<table>
<thead>
<tr>
<th></th>
<th>U.S.</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>0.22</td>
<td>1.13</td>
</tr>
<tr>
<td>Self-concept</td>
<td>0.20</td>
<td>1.03</td>
</tr>
</tbody>
</table>

**Data Analysis**

**Implementation of Reform-Based Science Learning**

**Latent Profile Analysis.** A latent profile analysis (LPA) was used to classify students into different science learning subgroups based on frequency and types of the reform-based learning activities in the science classroom. That is, different latent subgroups would represent different usage and foci of the learning activities that students experienced in the classroom.

Figure 3 refers to the latent profile model using 15 reform-based learning activities (e.g. doing experiments and draw conclusions) as indicators of the latent profile for science learning. This analysis was conducted for both U.S. and Taiwan samples, separately.

---

Figure 3. Latent profile model diagram of science learning
Latent class analysis (LCA) and latent profile analysis (LPA) are two widely used latent variable models. Both analyses are powerful methods for finding unobservable homogeneous subgroups within a population and are used to classify groups of individuals who are similar to each other and different from those in other groups. The difference between them is that indicators for LCA are categorical variables and indicators for LPA are continuous variables. In this study, LPA was used to answer the first research question, which was intended to measure how reform-based science learning has been implemented in the classrooms. This study used 15 reform-based learning activities as indicators of the latent variable, and students’ responses to these items on a four-point Likert scale were treated as continuous variables.

A useful framework to understand LPA is to distinguish person-centered and variable-centered approaches (Muthén & Muthén, 2000). As Muthén and Muthén described, variable-centered approaches, including regression, factor analysis, and structural equation modeling, focused on describing the relationships among variables, such as how dependent and independent variables were related. The goal for the variable-centered approach was to identify significant predictors among independent outcome variables. Person-centered approaches, on the other hand, included methods such as cluster analysis, LCA, and LPA focusing on the relationships among individuals. The goal for the person-centered approach was to classify individuals into distinct subgroups based on individual response patterns so that individuals within a subgroup were more similar than individuals between subgroups.

Because of LPA’s exploratory nature, the number of the latent subgroups was not determined as a priori. Thus, the first step in a LPA was to run models with different numbers of subgroups and to compare them to determine the number of subgroups with well-defined, differentiated profiles. Generally, researchers have relied on information criterion (IC) indices,
specifically Akaike’s information criterion (AIC), the Bayesian information criterion (BIC), and the sample-size-adjusted BIC (a-BIC). IC indices are descriptive indices for model comparisons. They take into account both the goodness of fit of a model and the parsimony of a model. The basic idea is that the model which fits the data well and uses fewer parameters is the better model. The AIC is the maximized log likelihood function with the penalty of the two times the numbers of parameters estimated, and the BIC is the maximized log likelihood function with the penalty of the log of N times the number of parameters estimated. Comparing different models, the model with the smallest AIC, BIC, or a-BIC value is the best model.

Another approach to compare models was to test the statistical significance to determine whether a more complex model fit the data significantly better than a more parsimonious model, when the models being compared were nested. The Vuong–Lo–Mendell–Rubin (VLMR) test (Lo, Mendell, & Rubin, 2001) was developed based on a principle similar to the likelihood ratio difference test. A significant p-value indicates that a model with K classified classes fits the data better than a more parsimonious model with K-1 class.

Finally, another way to evaluate models was to see whether an individual has been correctly assigned to a subgroup. Entropy was a summary measure used for the quality of the classification for LPA. Entropy shows an estimated probability between 0 and 1 with values approaching to 1 indicating good classification accuracy and values approaching 0 indicating lack of accuracy (Geiser, 2012).

This study considered together AIC, BIC, a-BIC, VLMR Test, and Entropy as statistical indicators to make decisions for the number of subgroups for LPA model. However, Lanza, Bray, and Collins (2013) suggested that when deciding the number of subgroups, researchers
should not only consider these statistics indices, but also make sure to select the model that makes most sense in relation to theory, previous research, and interpretation of each subgroup.

In this study, LPA was used to classify students into different science learning subgroups based on the usage and foci of the reform-based learning activities in science classrooms. Students within the same science learning subgroup exhibited similar patterns of responses for their science learning activities, while those between science learning subgroups exhibited significant differences. That is, each science learning subgroup represents a unique learning environment based on the combination of usage and foci of the learning activities that students experienced in the classroom. Therefore, each learning subgroup should have a distinct feature and is different from others. LPA produced average frequency for each science learning activity in each science learning subgroups and each subgroup was labeled based on the nature and prevalence of the activities.

LPA was chosen for this study for several reasons. First, research in this area traditionally had relied on multiple regression analyses to determine which learning activities were the predictors for students’ learning outcomes, such as achievement or attitudes. However, in real classroom settings, learning activities did not occur solely in class. Wise (1996) suggested that innovative science instruction is a mixture of different teaching activities and that using one strategy is not as powerful as utilizing a combined strategies approach. The author was more interested in whether specific classroom learning and learning patterns can be identified as being associated with students’ learning outcomes. LPA categorized students based on their exposure to hands-on activities, student investigations, interactive teaching, and real-life applications concurrently. This approach allowed researchers to shift from the conventional approach of
studying these several learning activities as individual phenomena to a new approach that acknowledged the interaction among them.

Second, researchers, traditionally, classified individuals in a sample based on predetermined values as cut-scores based on the sample mean or percentile (Lang, 1982). However, LPA, which assumed an underlying categorical latent variable determining an individual’s class membership, was model-based, and this suggested that the model could be replicated with another sample (Muthén & Muthén, 2000). LPA also provided statistical indices, which allowed researchers to assess model fit to decide the number of subgroups. This allowed researchers to assess and confirm the appropriateness of the classification.

**Availability of Reform-Based Science Learning**

Latent Class Regression (LCR) and Structural Equation Modeling (SEM) were used to test whether there was a relationship between students’ SES and the frequency students learn science through reform-based science learning. The LCR and SEM intended to address the availability issues using different approaches. The analyses were conducted for both the U.S. and Taiwan samples.

**Latent Class Regression (LCR).** Latent class regression (Masyn, 2013) is LCA or LPA with hypothesized covariates or outcome variables of latent class (or profile) membership. LPA with covariates uses multinomial logistic regression to parameterize the relationship between the probability of latent subgroup membership and covariates. On the other hand, LPA with outcomes adds outcome variables as additional indicators and allows them to vary across different subgroups. Including covariates or outcome variables can validate the latent subgroups that emerged and to gain a richer characterization and interpretation of the latent subgroups. If covariates are included in a LCR model and the results indicate meaningful relationships
between latent subgroups and a set of selected covariates, the results can support the validity of the subgroups. Including covariates or outcome variables can also be used to test hypotheses related to subgroup membership and adding variables based on theories. Figure 4 displays the LCR model diagram used to answer the second research question in this study. The model included the covariate of SES to test the association between students’ SES and students’ science learning subgroup membership.

Figure 4. Latent class regression model diagram with covariate of SES

**Structural Equation Modeling (SEM).** After examining 15 reform-based science learning activities as a whole to represent the latent variable for classroom science learning, this study was individualized with more detail to investigate each science learning dimensions (i.e., hands-on activities, student investigation, interaction, and application). SEM, a combination of factor analysis and path analysis, is appropriate for testing proposed theoretical models with latent variables. More specifically, path analysis models were analyzed because this study directly used indices of each latent construct that PISA 2006 had supplied. In an SEM model, causal relationships among theoretical variables were represented by a series of structural equations (regression equations), and the relationships were estimated simultaneously in a
model. Then, the goodness of fit between the theory-based model and the data was tested statistically. If the goodness of fit was adequate, the postulated model would be plausible and consistent with the data; if the goodness of fit was poor, the model would not be plausible and would need to be rejected or re-specified (Byrne, 1998).

The SEM was guided using Schunk’s (1999) model of Social-Self-Outcome Interaction in achievement settings and can investigate the relationship among the following environment variables: SES and reform-based science learning; personal variables: self-efficacy and self-concept; and behavior outcome variables: science achievement. For example, the model tested the direct effect of students’ SES and the science learning dimension of hands-on activities on students’ science achievement, as well as the indirect effect of students’ SES and the science learning dimension of hands-on activities on students’ science achievement through students’ science self-efficacy and science self-concept. This model was used to answer the research question for understanding the associations between students’ SES and each of the four dimensions of science learning. Figure 5 displays the proposed SEM modeling and this model was tested for both the U.S. and Taiwan samples.

![Figure 5. Proposed SEM models of SES, science achievement, self-efficacy, self-concept, and science learning (i.e. hands-on activities, student investigation, interaction, and application).](image)

56
SES Achievement Gap and Reform-Based Science Learning

Regression Mixture Modeling and SEM were used to test whether there was a relationship between the frequency of students learning science through reform-based science learning and SES-associated achievement gaps. These analyses were conducted for both U.S. and Taiwan samples, separately.

**Regression Mixture Model.** Regression mixture models (Ding, 2006; Muthén & Asparouhov, 2009) can be viewed as a combination of the conventional regression model and the classic LCA (or LPA). Conventional regression analysis assumes that a sample comes from a single population and that estimated parameters are the same for the entire population. Regression mixture analysis relaxes the assumption of single population to allow slope and intercept to vary across several unobserved subpopulations, which correspond to different subgroups emerging from LPA. Figure 6 displays the Regression mixture model for this study. The solid arrow represents conventional regression of students’ science achievement on their SES. The dashed arrows from science learning latent subgroup to the regression of achievement on SES indicates that the slope varies across the latent subgroups. Students’ gender and their self-beliefs (self-efficacy and self-concept) were controlled for achievement when examining the relationship between SES and achievement.
Structural Equation Modeling (SEM). As in the previous section, after examining 15 reform-based science learning activities as a whole to represent the latent variable for classroom science learning, this study was individualized with more detail to investigate each science learning dimensions’ (i.e. hands-on activities, student investigation, interaction, and application). Building on the results of the SEM model from the previous section, the interaction terms between students’ SES and each of four dimensions of science learning was added to examine the interaction effects between four dimensions of reform-based science learning and students’ SES on students’ science achievement. Figure 7 shows that the proposed SEM model and the model was tested both for the U.S. and Taiwan samples, separately.
Figure 7. Proposed SEM interaction models of SES, science achievement, science learning (i.e. hands-on activities, student investigation, interaction, and application), and the interaction between SES and science learning.

In this study, the Mplus 7.2 (Muthén & Muthén, 1998-2007) statistical software package was employed to run the LPA, LCR and Regression Mixture Model. Mplus utilized the EM algorithm and full information maximum likelihood estimates to compute the LPA model parameters, standard errors and fit statistics. To ensure that the estimates represented global likelihood maxima solutions and not local optima, 500 random start values were used.

Mplus 7.2 also was used for its ability to deal with SEM analysis with complex survey data (Muthen & Muthen, 1998-2007). Student weighting variables in PISA 2006 are used to adjust for unequal probability of selection. To examine student achievement measured by five plausible values, the function of multiple imputation in Mplus 7.2 was used in the models.

Schafer (1997) explained how the model was analyzed: “Parameter estimates are averaged over
the set of analyses, and standard errors are computed using the average of the standard errors over the set of analyses and the between analysis parameter estimate variation” (p. 72). Given that some variables may not be normally distributed, the study used MLR estimator in Mplus 7.2 to deal with these non-normal distribution variables. MLR is defined as “maximum likelihood parameter estimates with standard errors and a chi-square test statistic (when applicable) that are robust to non-normality and non-independence of observations” (Muthen & Muthen, 1998-2007, p. 484).

Because PISA 2006 used complex sampling designs with students being nested in school, multi-level SEM was used for its capability to deal with the problems of underestimation of sampling variance by correcting estimation of standard errors. Stapleton (2006) suggested the first step to undertake multilevel SEM was to consider the intraclass correlation coefficient (ICC) of the variables. ICC is the proportion of the variance in each of the variables accounted by a function of differences in cluster means. A higher ICC value indicates greater variance, and thus multilevel analysis is needed. Kline (2010) suggested that there was a need for multilevel SEM instead of single-level SEM if the ICC was greater than 0.10. On the other hand, if the value of the ICCs was close to zero, multilevel SEM would be difficult to conduct due to convergence problems. Thus, if the SEM model had convergence problem for this study, any variables with ICC less than 0.10 would be set to the student-level variables.

The hypothesized model for this study only focused on testing relationships among variables at the student-level. Thus, in order to get a better model fit, the variables at the school-level were allowed to freely co-vary with each other as a saturated model.

The following model fit indices were used for evaluating the model in this study: (a) the chi-squared statistics and p-value. When the p value is greater than .05, it indicates that there are
no statistically significant discrepancies between the sample variance-covariance matrix and the reproduced implied covariance matrices. As the Chi-squared statistic is very sensitive to sample size and departures from multivariate normality, it may easily reject a well-fitting model; (b) the Comparative Fit Index (CFI). A CFI value close to .95 reflects a good fit and 1.0 indicates a perfect fit; (c) Root-Mean-Square Error of Approximation (RMSEA). A RMSEA value less than .05 indicates a good model fit; and (d) Standardized Root Mean Squared Residual (SRMR). Hu and Bentler (1999) suggested that a good model would have values above .95 in CFI and below .08 in SRMR, or values below .06 in RMSEA.

**Limitations**

Although the secondary dataset in this study provided a wealth of educational data and has several advantages, it has some limitations. From the issues of questionnaire design, the PISA 2006 science teaching and learning questionnaire asked students which science teaching and learning activities occurred in their classroom. The questionnaire did not provide indicators that completely captured the nature of reform-based science teaching and only focused on four reform strategies: hands-on activities, student investigations, interactive teaching, and real-life applications. Other important aspects of reform-based science learning and teaching practices, such as modeling, were not captured using the PISA. The data from the student questionnaire also were limited by the questions asked, the directions for these questions, and the response choices provided. A major concern was that the student questionnaire was the only data source for examining students’ classroom practices. The study relied exclusively on student self-reported data. Therefore, the data used in this study was from student perspectives. There was measurement error in such data and that would cause reliability concerns (Le, et al., 2006). Students were asked to recall past science classes to answer the questions used in the dataset.
Discrepancies in students’ memory of a class could have influenced the study. Any misinterpretations by students would have also influenced the results of this study. Moreover, this study compared two countries, and any cultural differences in students’ responses to the questions and students’ rating inclinations might mean that the results cannot be compared directly. The PISA 2006 addressed this reliability issue when developing the questionnaire items. The classroom activities that the questionnaire focused on were international in scope and carried out in both countries. The questions only referred to observable class activities, which made it easier for students to evaluate lesson characteristics objectively. The question asked about the frequency with which clearly definable teaching and learning activities occurred in science lessons. With this methodology, PISA was able to provide reliable data and allowed a description of teaching and learning activities in the context of an international comparison (Kobarg, Prenzel, Seidel, Walker, McCrae, Cresswell, & Wittwer, 2011).

In terms of sampling design, the PISA’s sample target was 15-year-olds because this age was the last year of the compulsive education for majority of the countries in the world. Thus, an age-based sample was at the core of the sampling design for the PISA (OECD, 2007). However, the fact that the PISA sample was based on age rather than grade precluded PISA from sampling at the classroom level. Through this sampling design, the responses from students could not be aggregated to the classroom level, and it was therefore difficult to provide a complete description of science learning and teaching in class (Wittwer, 2008).

In terms of assessment design, there was a concern that what motivated students to do their best on the PISA science assessment. This assessment did not match their school science curricula and was not evaluated for course grades, and often teachers may put little emphasis on the PISA assessment. Another concern was whether the science achievement as measured by a
paper and pencil test was an appropriate measure of performance for students who engaged in reform-based learning activities. This concern arose because reform-based activities emphasized the learning process, and the paper and pencil test could not capture students’ performance during the learning process. Finally, many studies pointed out that students’ science classroom learning and teaching activities were related to students’ science assessment scores and student self-reported science learning self-efficacy. However, the PISA asked about science learning activities in their classroom for the current school year, while student science performance and self-efficacy was a result of many years of collective learning and life experiences. Thus, the results of the associations between teaching and learning practices and students’ learning outcome needed to be interpreted with caution (Kobarg et al., 2011).

**Summary**

This chapter summarizes three main issues regarding implementation of reform-based science learning, availability of reform-based learning, and the SES achievement gap in the U.S. and Taiwan. Various statistical approaches are presented in this chapter and are tied to two perspectives. The first perspective uses the reform-based science learning activities as a whole (15 items) and LPA with covariate, and Regression Mixture modeling is described. The second perspective uses the reform-based science learning activities as four dimensions: hands-on activities, student investigation, interactive teaching, and real-life applications, and SEM is described. This chapter organizes the research questions, analytic procedures, and hypotheses into the three main sections. The limitations are presented at the end of this chapter.
CHAPTER FOUR

RESULTS

Chapter Four describes the research results of the study. The descriptive statistics of the reform-based learning variables are presented. The rest of the results are organized according to the three main foci of the study (i.e., implementation, availability, and achievement gap). First, to examine how reform-based science learning has been implemented in the science classrooms, the results of LPA are reported. Second, to test whether there was a relationship between students’ SES and the frequency of students learning science through reform-based learning activities, the results of latent class regression (LCR) and structural equation modeling (SEM) are reported. Finally, to investigate whether there was a relationship between the frequency of students learning science through reform-based learning activities and SES-associated achievement gaps, the regression mixture model and SEM are reported.

Descriptive Statistics

Research question 1 was empirically examined using LPA to classify students into different science learning subgroups by the usage and foci of learning activities in their science classrooms separately for U.S. and Taiwan. The PISA 2006 science teaching and learning survey asked students how often they learned science through the 15 reform-based learning activities. In the PISA 2006, these 15 reform-based learning activities have been factor-analyzed and belong to four dimensions: hands-on activities, student investigations, interactive teaching, and real-life applications. Table 5 summarizes students’ responses and presents means and standard deviations of these reform-based teaching and learning activities.
Table 5.
Means and Standard Deviations of Fifteen Teaching and Learning Activities for the PISA 2006 U.S. and Taiwan Data.

<table>
<thead>
<tr>
<th>Hands-on Activities</th>
<th>U.S.</th>
<th>Taiwan</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Do Experiments</td>
<td>5458</td>
<td>2.46</td>
<td>.801</td>
<td>5981</td>
<td>1.94</td>
<td>.625</td>
</tr>
<tr>
<td>Design for Lab</td>
<td>5441</td>
<td>2.43</td>
<td>.870</td>
<td>5981</td>
<td>1.83</td>
<td>.728</td>
</tr>
<tr>
<td>Draw Conclusions</td>
<td>5444</td>
<td>2.91</td>
<td>.858</td>
<td>5967</td>
<td>2.26</td>
<td>.820</td>
</tr>
<tr>
<td>Follow Instruction</td>
<td>5423</td>
<td>2.89</td>
<td>.859</td>
<td>5960</td>
<td>2.62</td>
<td>.913</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Student Investigations</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Design Own Experiments</td>
<td>5448</td>
<td>2.05</td>
<td>.961</td>
<td>5974</td>
<td>1.61</td>
<td>.785</td>
</tr>
<tr>
<td>Choose Own Investigations</td>
<td>5422</td>
<td>2.03</td>
<td>.915</td>
<td>5960</td>
<td>1.76</td>
<td>.836</td>
</tr>
<tr>
<td>Test Own Ideas</td>
<td>5423</td>
<td>2.24</td>
<td>.928</td>
<td>5964</td>
<td>1.95</td>
<td>.859</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interaction</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Explain Ideas</td>
<td>5468</td>
<td>2.98</td>
<td>.840</td>
<td>5982</td>
<td>2.78</td>
<td>.918</td>
</tr>
<tr>
<td>Student Opinion</td>
<td>5427</td>
<td>2.63</td>
<td>.934</td>
<td>5971</td>
<td>2.32</td>
<td>.863</td>
</tr>
<tr>
<td>Class Debate</td>
<td>5448</td>
<td>2.44</td>
<td>.941</td>
<td>5967</td>
<td>2.17</td>
<td>.889</td>
</tr>
<tr>
<td>Discussion Topics</td>
<td>5420</td>
<td>2.70</td>
<td>.920</td>
<td>5965</td>
<td>2.43</td>
<td>.855</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Applications</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
<td>N</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Apply Different Phenomena</td>
<td>5455</td>
<td>2.89</td>
<td>.885</td>
<td>5970</td>
<td>2.80</td>
<td>.801</td>
</tr>
<tr>
<td>Understand World Outside</td>
<td>5424</td>
<td>2.68</td>
<td>.883</td>
<td>5965</td>
<td>2.25</td>
<td>.867</td>
</tr>
<tr>
<td>Explain Relevance</td>
<td>5427</td>
<td>2.68</td>
<td>.912</td>
<td>5956</td>
<td>2.74</td>
<td>.827</td>
</tr>
<tr>
<td>Society Relevance</td>
<td>5434</td>
<td>2.51</td>
<td>.900</td>
<td>5971</td>
<td>2.47</td>
<td>.863</td>
</tr>
</tbody>
</table>

When comparing the mean scores of Taiwan students, the U.S. students had higher frequency of learning science through all 15 activities in classroom except the activity of “explain relevance”. In general, the biggest differences for the frequency between the U.S. and Taiwan were the “hands-on activities”. For example, the difference for the frequency of the “do experiments”, “design for lab”, and “draw conclusions” between the U.S. and Taiwan equaled to 0.52, 0.60, and 0.65 units, respectively. On the other hand, the smallest differences were the learning activities of the dimension of “applications”. For example, the difference of the frequency of the “apply different phenomena”, “explain relevance”, and “society relevance” equaled to 0.09, 0.06, and 0.04 units, respectively. The other learning activities of the dimensions
of “student investigation” and “interaction” were in the middle in regards to their differences. The range of the differences were from 0.20 (the difference for the “explain ideas”) to 0.44 (the difference for the “Design Own Experiments”).

When comparing for similarities between U.S. and Taiwan, the students in both countries had higher frequencies in learning science through the activities of the dimension of “interaction” (e.g., the mean value of “explain ideas” was 2.98 in the U.S. and 2.78 in Taiwan) and “applications” (e.g., the mean value of “apply to different phenomena” was 2.89 in the U.S. and 2.80 in Taiwan). Students in both countries had the lowest frequencies in learning science through the activities of the dimension of “student investigations” (e.g., the mean value of “design own experiments” is 2.05 in the U.S. and 1.61 in Taiwan).

Implementation of Reform-Based Science Learning

Latent Profile Analysis

United States. The LPA was applied to the 15 learning activities. Based upon literature review, there is no strong theory that allows us to make an assumption about the number of science learning subgroups to extract; so the number of subgroups had to be determined through model comparisons, and the model with the best model fit was selected. This study began by conducting LPA with tests of a one-subgroup model, which simply yielded the observed means in the data, and was followed by tests of models, each with one more subgroup than the previous model, until the model could not converge correctly. Table 6 presents five indices of fit, AICs, BICs, a-BICs, entropy and VLMR, between one- and six-subgroup models, which were commonly used in LPA research. First, the three information criteria (AIC, BIC, and a–BIC) was evaluated. Typically, when evaluating the information criteria, the best model should be the majority of the fit indices with the lowest values. However, the results from the PISA 2006 USA
data showed that the values continued to decrease across all of models considered, therefore suggesting that we should consider at least six subgroups. Thus, these three information indices did not provide sufficient information for determining the number of subgroups. Such a situation is a methodological issue that is still unresolved in the literature (Muthén, 2003). Second, the VLMR test was used to determine if a more complex model was able to fit the data significantly better than a more parsimonious model. The results of the VLMR test showed that there were significant differences when comparing the two- to one-subgroup model (p < .001), the three- and two-subgroup model (p < .001), and the four- to three-subgroup model (p < .01), whereas there was no significant difference when comparing the five- to four-subgroup model (p = .760), suggesting that the five-subgroup solution was not statistically significantly better than the four-subgroup solution. Finally, entropy was used as another indicator to summarize the quality of the classification in an LPA model and should be considered. Our results showed that the four-subgroup model was among the best (entropy=0.852), implying that 85.2% of the students are accurately categorized in a subgroup. Based on these model comparisons, the four-subgroup model seemed to have the best model fit and was selected for modeling the U.S. sample.

Table 6.  
*Goodness-of-Fit Criteria for Various Latent Profile Models for the PISA 2006 U.S. data*

<table>
<thead>
<tr>
<th>Number of Subgroups</th>
<th>Number of Free Parameters</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>a-BIC</th>
<th>VLMR</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>-106469.1</td>
<td>212998.2</td>
<td>213196.5</td>
<td>213101.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>-97446.7</td>
<td>194985.5</td>
<td>195289.6</td>
<td>195143.4</td>
<td>0.000</td>
<td>0.856</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>-94572.6</td>
<td>189269.3</td>
<td>189679.1</td>
<td>189482.0</td>
<td>0.000</td>
<td>0.849</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>-93365.7</td>
<td>186887.4</td>
<td>187402.9</td>
<td>187155.0</td>
<td>0.008</td>
<td>0.852</td>
</tr>
<tr>
<td>5</td>
<td>94</td>
<td>-92542.5</td>
<td>185273.0</td>
<td>185894.3</td>
<td>185595.6</td>
<td>0.760</td>
<td>0.824</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>-91986.0</td>
<td>184192.1</td>
<td>184919.0</td>
<td>184569.5</td>
<td>0.128</td>
<td>0.810</td>
</tr>
</tbody>
</table>

*Note. N = 5481. AIC = Akike Information Criteria; BIC = Bayesian Information Criteria; a-BIC = Sample-Size Adjusted BIC; VLMR = p values for the Vuong-Lo-Mendell–Rubin likelihood for K versus K-1 classes.*
Figure 8 presents the results of latent profile for four identified science learning subgroups in the U.S. sample. The profiles of mean scores for the four subgroups did not differ qualitatively (profile shape) but did quantitatively (profile level). All four profiles had higher values at the activities of “draw conclusions”, “students ideas”, and “different phenomena”; and lower values at the activities of “choose own investigation”, “class debates”, and “society relevance”. For quantitative levels, the profile with diamond symbols had uniformly highest scores around 3.5 points across all learning activities. In this learning subgroup, students learned science through all 15 reform-based learning activities in most or all of their lessons, so this subgroup was labeled as the “highest usage reform-based learning” subgroup. Second, the profile with rectangle symbols had uniformly the second highest scores with score ranging from 2.2 to 3.2 for all 15 learning activities. That is, in this learning subgroup, students learned science through these 15 activities in most lessons, and this subgroup was labeled as the “high usage reform-based learning” subgroup. Third, the profile with triangle symbols had uniformly the third to the highest scores, with around 2 to 2.5 across all 15 learning activities. In this learning subgroup, students learned science through these 15 activities only in some lessons, and this group was labeled as the “medium usage reform-based learning” subgroup. The last profile with cross symbols had uniformly the lowest scores, with around 1.5 across all 15 learning activities. In this learning subgroup, students seldom learned science through these 15 activities, and this subgroup was labeled as the “low usage reform-based learning” subgroup.
Figure 8. *Latent profile of four identified science learning groups for the U.S.*

Note. ♦: Highest usage reform-based learning subgroup (10.52%)  
■: High usage reform-based learning subgroup (40.66 %)  
▲: Medium usage reform-based learning subgroup (39.92%)  
X: Low usage reform-based learning subgroup (8.89%)

The results of LPA also showed the relative distribution of the sample across the four learning subgroups as 10.52% for highest usage reform-based learning subgroup, 40.66 % for the high, 39.92% for the medium, and 8.89% for the low. This indicated that after the science education community promoted reform-based learning activities, more than 50% of students in the U.S. learned science through these activities on the regular basis (in most lessons), and their learning not only focused on one activity but on all 15 reform-based science learning activities.
After the subgroups were identified, it was important to verify that the subgroups were differentiated in variables that were not used in the classification process. If such differences exist, they may provide additional information on the nature of the subgroups. For the purposes of this study, the variables used for validation chosen were related to students’ science learning outcomes. Examining the mean scores for students’ achievement, self-efficacy, and self-concept within each subgroups allowed the author to investigate the associations between reform-based science learning and students’ learning outcomes. Table 7 showed the mean scores and their significant contrast results by calculating their 95% confidence interval. The results showed that students in the medium subgroup had the significantly highest achievement score and that students in the highest subgroup had the significantly lowest achievement score. However, students in higher usage reform-based science learning subgroups had significantly higher science self-efficacy and science self-concept than students in lower usage reform-based science learning subgroups.

**Table 7.**

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Highest subgroup (n=547)</th>
<th>High subgroup (n=2229)</th>
<th>Medium subgroup (n=2196)</th>
<th>Low subgroup (n=482)</th>
<th>Total Sample (n=5454)</th>
<th>Significant Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>446.22</td>
<td>484.35</td>
<td>509.82</td>
<td>489.49</td>
<td>489.83</td>
<td>M&gt;L,H&gt;Highest</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>0.70</td>
<td>0.31</td>
<td>0.10</td>
<td>-0.18</td>
<td>0.21</td>
<td>Highest&gt;H&gt;M&gt;L</td>
</tr>
<tr>
<td>Self-concept</td>
<td>0.79</td>
<td>0.32</td>
<td>0.05</td>
<td>-0.25</td>
<td>0.20</td>
<td>Highest&gt;H&gt;M&gt;L</td>
</tr>
</tbody>
</table>

**Taiwan.** Similarly, LPA was conducted on the 15 reform-based learning activities. Table 8 presents the model fit indices of the one- to six-subgroup model for Taiwan. As the results of the U.S., the results of Taiwan also showed that three information criteria (AIC, BIC, and a–BIC) continued to decrease as the number of subgroups increased in the model. Thus, the fit indices...
did not provide sufficient information for determining the number of classes. Then, the VLMR test was conducted and the results showed that there were significant differences when comparing the two- to one-subgroup model (p < .001), the three- and two-subgroup model (p < .001), the four- to three-subgroup model (p < .001), and the five- to four-subgroup model (p < .05), whereas there was no significant difference when comparing the six- to five-subgroup model (p = .128). The result indicated that the six-subgroup solution was not statistically significantly better than the five-subgroup solution. This suggested that the five-subgroup solution was a better choice. However, the entropy value showed that the four-subgroup models was a better choice because the entropy for the five-subgroup model (0.835) is higher than the entropy for the four-subgroup model (0.810). Therefore, the latent profiles of both four- and five-subgroup models were considered and compared thoroughly. When examining the latent profile for the five-subgroup models (Figure 9), two of the subgroups, the circle symbols and the triangle symbols, seemed twisted together, and that caused misclassification for these two subgroups. Table 9 showed the average latent subgroup probabilities for most likely latent subgroup membership (row) by latent subgroup (column) for the five-subgroup model. The results showed that these two subgroups did not have satisfactory average probabilities (0.82 for the subgroup with circle symbol and 0.86 for the subgroup with triangle symbol; and over 0.90 was considered satisfactory). Therefore, misclassification was the reason for its low entropy value. On the other hand, Table 10 showed that all the average probabilities were satisfactory for the four-subgroup model. Moreover, the four-subgroup model was selected for the U.S. data, and in order to compare the model between Taiwan and U.S., the four-subgroup model was chosen for modeling the Taiwan data.
Table 8.
**Goodness-of-Fit Criteria for Various Latent Class Models for the PISA 2006 Taiwan data**

<table>
<thead>
<tr>
<th>Number of Classes</th>
<th>Number of Free Parameters</th>
<th>Log-Likelihood</th>
<th>AIC</th>
<th>BIC</th>
<th>a-BIC</th>
<th>VLMR</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30</td>
<td>-109955.6</td>
<td>219971.3</td>
<td>220172.3</td>
<td>220076.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>-100861.0</td>
<td>201814.0</td>
<td>202122.1</td>
<td>201975.9</td>
<td>0.000</td>
<td>0.866</td>
</tr>
<tr>
<td>3</td>
<td>62</td>
<td>-98258.1</td>
<td>196640.2</td>
<td>197055.5</td>
<td>196858.5</td>
<td>0.000</td>
<td>0.833</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>-97016.1</td>
<td>194188.3</td>
<td>194710.7</td>
<td>194462.9</td>
<td>0.000</td>
<td>0.835</td>
</tr>
<tr>
<td>5</td>
<td>94</td>
<td>-96133.2</td>
<td>192454.4</td>
<td>193084.0</td>
<td>192785.3</td>
<td>0.019</td>
<td>0.810</td>
</tr>
<tr>
<td>6</td>
<td>110</td>
<td>-91986.0</td>
<td>184192.1</td>
<td>184919.0</td>
<td>184569.5</td>
<td>0.128</td>
<td>0.810</td>
</tr>
</tbody>
</table>

Note. N = 5989. AIC = Akike Information Criteria; BIC = Bayesian Information Criteria; a-BIC = Sample-Size Adjusted BIC; VLMR = p values for the Vuong-Lo–Mendell–Rubin likelihood for K versus K-1 classes.

Figure 9. *Latent profile of five-subgroup model for Taiwan*
Table 9.
Average latent subgroup probabilities for most likely latent subgroup membership (row) by latent subgroup (column) for the 5-subgroup model for Taiwan

<table>
<thead>
<tr>
<th>Latent Subgroup</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Circle)</td>
<td><strong>0.82</strong></td>
<td>0.03</td>
<td>0.11</td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>2 (Cross)</td>
<td>0.03</td>
<td><strong>0.91</strong></td>
<td>0.06</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>3 (Triangle)</td>
<td>0.06</td>
<td>0.04</td>
<td><strong>0.86</strong></td>
<td>0.04</td>
<td>0.00</td>
</tr>
<tr>
<td>4 (Square)</td>
<td>0.03</td>
<td>0.00</td>
<td>0.05</td>
<td><strong>0.90</strong></td>
<td>0.02</td>
</tr>
<tr>
<td>5 (Diamond)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td><strong>0.94</strong></td>
</tr>
</tbody>
</table>

Note. On-diagonal of the matrix for average posterior probabilities indicate precision of the latent subgroup membership estimation.

Table 10.
Average latent subgroup probabilities for most likely latent subgroup membership (row) by latent subgroup (column) for the 4-subgroup model for Taiwan

<table>
<thead>
<tr>
<th>Latent Subgroup</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Cross)</td>
<td><strong>0.90</strong></td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2 (Triangle)</td>
<td>0.04</td>
<td><strong>0.91</strong></td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>3 (Square)</td>
<td>0.00</td>
<td>0.08</td>
<td><strong>0.90</strong></td>
<td>0.02</td>
</tr>
<tr>
<td>4 (Diamond)</td>
<td>0.00</td>
<td>0.00</td>
<td>0.06</td>
<td><strong>0.94</strong></td>
</tr>
</tbody>
</table>

Note. On-diagonal of the matrix for average posterior probabilities indicate precision of the latent subgroup membership estimation.

Figure 10 presents the results of latent profiles for four identified science learning subgroups for the Taiwan data. Similar to the U.S., the profiles of mean scores for the four subgroups did not differ qualitatively (profile shape) but differed quantitatively (profile level). They all had higher values at the activities of “follow instruction”, “students ideas”, and “different phenomena”; and lower values at the activities of “choose own investigation”, “class debates”, and “world outside”. For the quantitative level, the profile with diamond symbols had
the uniformly highest scores across all 15 learning activities ranging from 3 to 3.5 units. In this learning subgroup, students learned science through these 15 reform-based learning activities in most of their lessons, and this subgroup was labeled as the “highest usage reform-based learning” group. Second, the profile with rectangle symbols had uniformly the second scores with score ranging from 1.8 to 3.2 units for all 15 learning activities. In this learning subgroup, students learned science through these 15 activities in most lessons, and this subgroup was labeled as the “high usage reform-based learning” subgroup. Third, the profile with triangle symbols had uniformly the third scores with around 1 to 2.7 units across all 15 learning activities. In this learning subgroup, students learned science through these 15 activities only in some lessons, and this subgroup was labeled as the “medium usage reform-based learning” subgroup. The last profile with cross symbols had uniformly the lowest scores ranging from 1.0 to 2.1 units across all 15 learning activities. In this learning subgroup, students seldom learned science through these 15 reform-based activities in lessons, and this group was labeled as a low usage reform-based learning subgroup. Although the labels for the learning subgroups are the same for the U.S. and Taiwan, the quantitative profile levels in the Taiwan sample were lower than the ones in U.S. sample. This was consistent for all four subgroups, and the differences were about 0.5 units. For example, while the mean scores of the 15 learning activities for the highest usage reform-based learning group in the U.S. was around 3.5, the ones in Taiwan was just around 3.2.

The results of LPA also showed that the relative distribution of the sample across the four learning subgroups was 7.2% for highest usage reform-based learning subgroup, 28.2% for the high, 47.2% for the medium, and 17.5% for the low subgroup.
Figure 10. *Latent profile of four identified science learning subgroups for Taiwan.*

Note. ♦ : Highest usage reform-based learning subgroup (7.15%)
■ : High usage reform-based learning subgroup (28.16%)
▲ : Medium usage reform-based learning subgroup (47.22%)
X : Low usage reform-based learning subgroup (17.45%)

Table 11 showed the mean scores and their significant contrast results by calculating 95% confidence interval for science achievement, self-efficacy, and science self-concept. The results showed that students in the highest subgroup had the significantly lower achievement score than the other three subgroups with lower frequencies in learning science through reform-based learning. However, generally, students had significantly higher science self-efficacy and science
self-concept when they had higher frequency to learn science through these reform-based learning activities.

Table 11. *Mean scores of outcome variables for the four science learning subgroups for Taiwan*

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Highest subgroup (n=417)</th>
<th>High subgroup (n=1595)</th>
<th>Medium subgroup (n=2877)</th>
<th>Low subgroup (n=1089)</th>
<th>Total Sample (n=5978)</th>
<th>Significant Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>488.16</td>
<td>566.92</td>
<td>562.34</td>
<td>567.07</td>
<td>559.16</td>
<td>H, M, L&gt;Highest</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>0.28</td>
<td>0.48</td>
<td>0.27</td>
<td>0.14</td>
<td>0.30</td>
<td>H&gt;Highest, M&gt;L</td>
</tr>
<tr>
<td>Self-concept</td>
<td>-0.05</td>
<td>-0.22</td>
<td>-0.39</td>
<td>-0.60</td>
<td>-0.36</td>
<td>Highest, H&gt;M&gt;L</td>
</tr>
</tbody>
</table>

**Availability of Reform-Based Science Learning**

**Latent Class Regression (LCR)**

**United States.** The SES was introduced as a covariate in the 4-subgroup model to predict membership in a certain latent subgroup. The latent subgroup membership was regressed on the SES in a multinomial logistic regression. Table 12 includes the multinomial logistic regression coefficients, standard errors, p-value and odds ratio values, where the highest usage reform-based learning subgroup is the reference group. Results indicated that students’ SES was not a predictor of students’ likelihood of being in the less usage reform-based learning subgroups compared to the highest subgroup (i.e., the regression coefficients were not significant). This indicated that students with different SES were equally likely to be in the four science learning subgroups. The results supported the hypothesis that there was no association between students’ SES and their science learning subgroup membership. Although the result was counter-intuitive, it was a good result because it indicated that reform-based science learning was available for all students regardless of their SES background in the U.S. Table 13 showed the mean score of SES in each subgroup.
Table 12. *Logistic regression coefficients and odds ratio for 4-subgroup model with SES and gender as a covariate using the highest usage reform-based learning subgroup as the comparison subgroup for U.S.*

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>B</th>
<th>S.E.</th>
<th>p-value</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest (Reference)</td>
<td>0.000</td>
<td>-</td>
<td>-</td>
<td>1.000</td>
</tr>
<tr>
<td>High</td>
<td>0.072</td>
<td>0.069</td>
<td>0.296</td>
<td>1.075</td>
</tr>
<tr>
<td>Medium</td>
<td>0.083</td>
<td>0.065</td>
<td>0.199</td>
<td>1.086</td>
</tr>
<tr>
<td>Low</td>
<td>-0.007</td>
<td>0.129</td>
<td>0.956</td>
<td>0.993</td>
</tr>
</tbody>
</table>

Table 13. *Mean scores of covariate variables for the four science learning subgroups for U.S.*

<table>
<thead>
<tr>
<th></th>
<th>Highest subgroup (n=547)</th>
<th>High subgroup (n=2229)</th>
<th>Medium subgroup (n=2196)</th>
<th>Low subgroup (n=482)</th>
<th>Total Sample (n=5454)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>0.089</td>
<td>0.149</td>
<td>0.158</td>
<td>0.083</td>
<td>0.133</td>
</tr>
</tbody>
</table>

**Taiwan.** Similarly, the SES was introduced as a covariate in the 4-subgroup model to predict membership in a certain latent subgroup. Table 14 provides the multinomial logistic regression coefficients, standard errors, p-value and odds ratio values, where the highest usage reform-based learning subgroup is the reference group for Taiwan. The results indicated that there was a significant SES effect for the three less usage of reform-based learning subgroups compared to the highest usage reform-based learning subgroup. The SES logistic regression coefficient (0.322, p = .006, odds ratio = 1.37 for the high subgroup; 0.165, p = 0.081, odds ratio = 1.17 for the medium subgroup; 0.245, p = 0.021, odds ratio = 1.27 for the low subgroup) indicated that students with higher SES had a greater chance of being in the less usage of reform-based learning subgroups instead of in the highest usage reform-based learning subgroup. Table 15 shows students’ mean SES value in each of the subgroup. This study found that students in
The highest usage subgroup had lower SES (-0.311) than the other three subgroups. The result indicated that reform-based science learning was not equally available for all students regardless of their SES background in Taiwan. The result was counter-intuitive because it indicated that students with lower SES had more opportunity to learn science through reform-based learning activities.

**Table 14.**
*Logistic regression coefficients and odds ratio for 4-subgroup model with SES and gender as a covariate using the highest usage reform-based learning subgroup as the comparison subgroup for Taiwan*

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>B</th>
<th>S.E.</th>
<th>p-value</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest (Reference)</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
<td>1.00</td>
</tr>
<tr>
<td>High</td>
<td>0.322</td>
<td>0.117</td>
<td>0.006</td>
<td>1.37</td>
</tr>
<tr>
<td>Medium</td>
<td>0.165</td>
<td>0.094</td>
<td>0.081</td>
<td>1.17</td>
</tr>
<tr>
<td>Low</td>
<td>0.245</td>
<td>0.107</td>
<td>0.021</td>
<td>1.27</td>
</tr>
</tbody>
</table>

**Table 15.**
*Mean scores of covariate variables for the four science learning subgroups for Taiwan*

<table>
<thead>
<tr>
<th></th>
<th>Highest subgroup (n=417)</th>
<th>High subgroup (n=1595)</th>
<th>Medium subgroup (n=2877)</th>
<th>Low subgroup (n=1089)</th>
<th>Total Sample (n=5978)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>-0.311</td>
<td>-0.108</td>
<td>-0.207</td>
<td>-0.156</td>
<td>-0.180</td>
</tr>
</tbody>
</table>

**Structural Equation Modeling**

**United States.** After examining 15 reform-based science learning activities as a whole to represent the latent variable for classroom science teaching, this study was individualized with more detail to examine each dimension of reform-based science learning. The SEM helped to understand how students’ SES influenced students’ science learning outcomes (i.e., science achievement, science self-efficacy, and science self-concept) through four dimensions of reform-
based science learning (i.e., hands-on activities, student investigation, interaction, and application). Particularly, the SEM was to investigate whether each of the four dimensions of reform-based science learning was available for all students regardless of their SES background. Figure 11 presents the SEM model that was tested for both in the U.S. and in Taiwan.

![SEM diagram]

Figure 11. SEM models of SES, science achievement, self-efficacy, self-concept, and science learning (i.e. hands-on activities, student investigation, interaction, and application).

First, the fit indices of the SEM models for the U.S. data were examined. Given that the values of RMSEA were 0.041, SRMR were 0.037, and CFI were 0.987, the fit indices in the models indicated a satisfactory fit, according to Hu and Bentler’s (1999) suggestion. The estimates of each path for the SEM model are shown in Table 16 in both unstandardized and standardized solutions. It was recommended that the effect sizes of the path coefficients be interpreted as follows: a small effect may indicate standardized path coefficients with absolute
values less than 0.10; a medium effect would be around 0.30; and a large effect would be greater than 0.50 (Kline, 2010).

When looking at the relationship between students’ SES and their learning through reform-based learning activities, the results showed that student with higher SES had significantly more opportunity to learn science through three of four dimensions: hands-on (standardized coefficient=0.067, p=0.024), interaction (standardized coefficient=0.093, p=0.001), and applications (standardized coefficient=0.086, p<0.001). All of their standardized coefficients were below 0.1, indicating students’ SES had a small effect on their frequency of learning through these reform-based learning activities. The results indicated that reform-based learning activities were not equally available for all students. Based on their SES backgrounds, the higher SES background that students have, the more opportunity they have to learn science through hands-on activities, interactive teaching, and real-life applications.

Controlling for students’ SES, two of the four dimensions had a positive relationship with their science self-efficacy and self-concept (hands-on: standardized coefficient=0.130, p=0.002 for self-concept and 0.101, p=0.011 for self-efficacy; applications: 0.112, p=0.037 for self-concept and 0.121, p<0.001 for self-efficacy). The range of these standardized values was around 0.1, indicating that these reform-based learning activities had a small positive effect on students’ science self-concept and self-efficacy. Lastly, controlling for students’ gender, SES, and self-efficacy and self-concept, the relationship between the reform-based science learning and students’ science achievement was investigated. The results indicated that while the dimension of application had a small positive effect on students’ achievement (standardized coefficient=0.090, p=0.020), the dimension of investigation had a medium negative effect on students’ achievement (standardized coefficient=−0.360, p<0.001).
Table 16. The unstandardized and standardized estimates of structural equation model for U.S.

<table>
<thead>
<tr>
<th>Path</th>
<th>Unstandardized values</th>
<th>Standardized values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>S.E.</td>
</tr>
<tr>
<td>SES to Hands-on</td>
<td>0.079</td>
<td>0.036</td>
</tr>
<tr>
<td>SES to Investigation</td>
<td>-0.045</td>
<td>0.043</td>
</tr>
<tr>
<td>SES to Interaction</td>
<td>0.115</td>
<td>0.034</td>
</tr>
<tr>
<td>SES to Application</td>
<td>0.115</td>
<td>0.033</td>
</tr>
<tr>
<td>SES to Self-efficacy</td>
<td>0.287</td>
<td>0.035</td>
</tr>
<tr>
<td>Hands-on to Self-efficacy</td>
<td>0.127</td>
<td>0.051</td>
</tr>
<tr>
<td>Investigation to Self-efficacy</td>
<td>-0.003</td>
<td>0.031</td>
</tr>
<tr>
<td>SES to Self-concept</td>
<td>0.184</td>
<td>0.027</td>
</tr>
<tr>
<td>Hands-on to Self-concept</td>
<td>0.152</td>
<td>0.051</td>
</tr>
<tr>
<td>Investigation to Self-concept</td>
<td>0.029</td>
<td>0.024</td>
</tr>
<tr>
<td>Interaction to Self-concept</td>
<td>0.066</td>
<td>0.047</td>
</tr>
<tr>
<td>Application to Self-concept</td>
<td>0.113</td>
<td>0.053</td>
</tr>
<tr>
<td>Gender to Achievement</td>
<td>-4.359</td>
<td>3.802</td>
</tr>
<tr>
<td>SES to Achievement</td>
<td>15.361</td>
<td>2.916</td>
</tr>
<tr>
<td>Self-efficacy to Achievement</td>
<td>23.030</td>
<td>2.54</td>
</tr>
<tr>
<td>Self-concept to Achievement</td>
<td>18.123</td>
<td>2.353</td>
</tr>
<tr>
<td>Hands-on to Achievement</td>
<td>3.377</td>
<td>3.008</td>
</tr>
<tr>
<td>Investigation to Achievement</td>
<td>-31.852</td>
<td>2.797</td>
</tr>
<tr>
<td>Interaction to Achievement</td>
<td>-4.134</td>
<td>2.808</td>
</tr>
<tr>
<td>Application to Achievement</td>
<td>7.626</td>
<td>3.339</td>
</tr>
</tbody>
</table>

Correlation

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-concept with Self-efficacy</td>
<td>0.430</td>
<td>0.041</td>
<td>&lt;0.001</td>
<td>0.433</td>
<td>0.033</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on with Investigation</td>
<td>0.482</td>
<td>0.033</td>
<td>&lt;0.001</td>
<td>0.573</td>
<td>0.028</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on with Interaction</td>
<td>0.472</td>
<td>0.022</td>
<td>&lt;0.001</td>
<td>0.592</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on with Application</td>
<td>0.535</td>
<td>0.022</td>
<td>&lt;0.001</td>
<td>0.612</td>
<td>0.014</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Investigation with Interaction</td>
<td>0.497</td>
<td>0.026</td>
<td>&lt;0.001</td>
<td>0.561</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Investigation with Application</td>
<td>0.551</td>
<td>0.031</td>
<td>&lt;0.001</td>
<td>0.567</td>
<td>0.022</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Interaction with Application</td>
<td>0.620</td>
<td>0.026</td>
<td>&lt;0.001</td>
<td>0.673</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Residual

<table>
<thead>
<tr>
<th></th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement</td>
<td>5053.591</td>
<td>262.84</td>
<td>&lt;0.001</td>
<td>0.690</td>
<td>0.019</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>1.065</td>
<td>0.053</td>
<td>&lt;0.001</td>
<td>0.897</td>
<td>0.015</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Self-concept</td>
<td>0.924</td>
<td>0.028</td>
<td>&lt;0.001</td>
<td>0.896</td>
<td>0.016</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on</td>
<td>0.757</td>
<td>0.026</td>
<td>&lt;0.001</td>
<td>0.995</td>
<td>0.004</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Investigation</td>
<td>0.936</td>
<td>0.031</td>
<td>&lt;0.001</td>
<td>0.999</td>
<td>0.002</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.839</td>
<td>0.034</td>
<td>&lt;0.001</td>
<td>0.991</td>
<td>0.005</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Application</td>
<td>1.009</td>
<td>0.025</td>
<td>&lt;0.001</td>
<td>0.993</td>
<td>0.004</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Taiwan. Similarly, the SEM analysis was applied to the Taiwan sample. The SEM was to investigate whether each of the four dimensions of reform-based science learning was available for all students regardless of their SES background. The fit indices of the SEM models for the Taiwan data were investigated. Given the values of RMSEA were 0.074, SRMR were 0.041, and CFI were 0.948, the fit indices in the models indicated an adequate fit, according to Hu and Bentler’s (1999) suggestion.

The estimates of each path for the four SEM models are shown in Table 17 for both unstandardized and standardized solutions. When examining the relationship between students’ SES and their learning through reform-based learning activities, the results showed that students’ SES had no significant relationship with their frequency of learning science through each of four dimensions. The results indicated that reform-based learning activities were equally available for all students regardless their SES backgrounds in Taiwan.

Controlling for students’ SES, the dimension of application had a medium positive effect on self-efficacy (standardized coefficient=0.200, p<0.001); and the dimension of hands-on (standardized coefficient=0.087, p=0.009), investigation (standardized coefficient=0.070, p=0.013) and application (coefficient=0.062, p=0.020) had a small positive effect on their science self-concept. Lastly, controlling for students’ gender, SES, self-efficacy and self-concept, the relationship between the reform-based science learning and students’ science achievement were investigated. While the dimension of application had a medium positive effect on students’ achievement (standardized coefficient=0.206, p<0.001), the dimension of hands-on had a small negative effect on students’ achievement (standardized coefficient=-0.069, p=0.036), and the dimension of investigation had a medium negative effect on students’ achievement (standardized coefficient=-0.285, p<0.001).
Table 17.  
*The unstandardized and standardized estimates of structural equation model for Taiwan*

<table>
<thead>
<tr>
<th>Path</th>
<th>Unstandardized values</th>
<th>Standardized values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>S.E.</td>
</tr>
<tr>
<td>SES to Hands-on</td>
<td>0.004</td>
<td>0.026</td>
</tr>
<tr>
<td>SES to Investigation</td>
<td>0.011</td>
<td>0.035</td>
</tr>
<tr>
<td>SES to Interaction</td>
<td>0.027</td>
<td>0.028</td>
</tr>
<tr>
<td>SES to Application</td>
<td>0.055</td>
<td>0.03</td>
</tr>
<tr>
<td>SES to Self-efficacy</td>
<td>0.254</td>
<td>0.032</td>
</tr>
<tr>
<td>Hands-on to Self-efficacy</td>
<td>0.016</td>
<td>0.054</td>
</tr>
<tr>
<td>Investigation to Self-efficacy</td>
<td>-0.039</td>
<td>0.031</td>
</tr>
<tr>
<td>Interaction to Self-efficacy</td>
<td>0.005</td>
<td>0.032</td>
</tr>
<tr>
<td>Application to Self-efficacy</td>
<td>0.209</td>
<td>0.026</td>
</tr>
<tr>
<td>SES to Self-concept</td>
<td>0.194</td>
<td>0.038</td>
</tr>
<tr>
<td>Hands-on to Self-concept</td>
<td>0.107</td>
<td>0.041</td>
</tr>
<tr>
<td>Investigation to Self-concept</td>
<td>0.071</td>
<td>0.029</td>
</tr>
<tr>
<td>Interaction to Self-concept</td>
<td>0.007</td>
<td>0.033</td>
</tr>
<tr>
<td>Application to Self-concept</td>
<td>0.063</td>
<td>0.027</td>
</tr>
<tr>
<td>Gender to Achievement</td>
<td>-12.812</td>
<td>3.521</td>
</tr>
<tr>
<td>SES to Achievement</td>
<td>13.929</td>
<td>3.117</td>
</tr>
<tr>
<td>Self-efficacy to Achievement</td>
<td>21.348</td>
<td>2.237</td>
</tr>
<tr>
<td>Self-concept to Achievement</td>
<td>2.084</td>
<td>2.378</td>
</tr>
<tr>
<td>Hands-on to Achievement</td>
<td>-6.566</td>
<td>3.117</td>
</tr>
<tr>
<td>Investigation to Achievement</td>
<td>-22.488</td>
<td>2.518</td>
</tr>
<tr>
<td>Interaction to Achievement</td>
<td>-2.560</td>
<td>2.453</td>
</tr>
<tr>
<td>Application to Achievement</td>
<td>16.390</td>
<td>2.388</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-concept with Self-efficacy</td>
<td>0.290</td>
<td>0.028</td>
</tr>
<tr>
<td>Hands-on with Investigation</td>
<td>0.444</td>
<td>0.031</td>
</tr>
<tr>
<td>Hands-on with Interaction</td>
<td>0.398</td>
<td>0.034</td>
</tr>
<tr>
<td>Hands-on with Application</td>
<td>0.414</td>
<td>0.033</td>
</tr>
<tr>
<td>Investigation with Interaction</td>
<td>0.468</td>
<td>0.029</td>
</tr>
<tr>
<td>Investigation with Application</td>
<td>0.459</td>
<td>0.035</td>
</tr>
<tr>
<td>Interaction with Application</td>
<td>0.502</td>
<td>0.037</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>4449.202</td>
<td>178.81</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>0.900</td>
<td>0.049</td>
</tr>
<tr>
<td>Self-concept</td>
<td>0.877</td>
<td>0.029</td>
</tr>
<tr>
<td>Hands-on</td>
<td>0.615</td>
<td>0.033</td>
</tr>
<tr>
<td>Investigation</td>
<td>0.908</td>
<td>0.029</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.830</td>
<td>0.033</td>
</tr>
<tr>
<td>Application</td>
<td>0.888</td>
<td>0.041</td>
</tr>
</tbody>
</table>
SES Achievement Gap and Reform-Based Science Learning

**Regression Mixture Model**

**United States.** This study also investigated whether reform-based science learning promoted education equity by minimizing the achievement gap for students from different SES backgrounds in the U.S. A regression mixture model allowed the intercept and slope of a linear regression to vary across four science learning subgroups. This analysis allowed testing the difference of the effect of SES on students’ science achievement across different science learning subgroups. The estimate for regression mixture analysis is presented in Table 18. In this analysis, students’ gender, self-efficacy, and self-concept were controlled the same for all four subgroups. The results showed that when students’ SES increased by one standard deviation, students’ science achievement scores increased approximately 38 units for the students’ in the highest, high, and medium subgroups, and 25.95 units for students in the low subgroup. Although Wald test of parameter constraints was not significant (p=0.085), the estimated SES effects on the low and medium reform-based usage subgroups were noticeable. The results from the PISA 2006 dataset exhibited that students in the low subgroup had a smaller achievement gap than the other three subgroups with more opportunity to learn science through reform-based activities.

*Table 18.*

*Parameter Estimates and Standard Errors for Four-Subgroup Mixture Regression Model for U.S.*

<table>
<thead>
<tr>
<th></th>
<th>Highest Subgroup</th>
<th>High Subgroup</th>
<th>Medium Subgroup</th>
<th>Low Subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S.E.</td>
<td>B</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>411.57</td>
<td>6.77</td>
<td>464.18</td>
<td>5.13</td>
</tr>
<tr>
<td>SES</td>
<td>37.73</td>
<td>5.35</td>
<td>37.91</td>
<td>2.49</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>3.85</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>25.48</td>
<td>1.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-concept</td>
<td>14.51</td>
<td>1.52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Taiwan. Similarly, a regression mixture model was applied to the Taiwan sample to investigate whether reform-based science learning promoted education equity by minimizing the achievement gap for students from different SES backgrounds. The estimates for regression mixture analysis for Taiwan is presented in Table 19. The results showed that when students’ SES increased by one standard deviation, students’ science achievement scores increased 44.15 units for the students in the highest subgroup and approximately 27 units for students in each of the other three subgroups. Wald test of parameter constraints indicated that there was a statistically significant difference in comparing the highest subgroup to the other three subgroups (p=0.009). This result indicated that students’ SES had significantly more effect on students’ achievement in the highest reform-based usage subgroup than on the other three subgroups.

Table 19.  
*Parameter Estimates and Standard Errors for Four-Subgroup Mixture Regression Model for Taiwan*

<table>
<thead>
<tr>
<th></th>
<th>Highest subgroup</th>
<th>High Subgroup</th>
<th>Medium subgroup</th>
<th>Low subgroup</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S.E.</td>
<td>B</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>497.78</td>
<td>6.39</td>
<td>562.32</td>
<td>3.44</td>
</tr>
<tr>
<td>SES</td>
<td>44.15</td>
<td>5.88</td>
<td>26.85</td>
<td>3.51</td>
</tr>
<tr>
<td>Covariates</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-7.54</td>
<td>2.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>25.41</td>
<td>1.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-concept</td>
<td>5.95</td>
<td>1.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Structural Equation Modeling (SEM)**

United States. The SEM model was applied to investigate whether any of four dimensions of reform-based science learning (i.e., hands-on activities, student investigation, interaction, and application) promoted education equity by minimizing the achievement gap for students from different SES backgrounds. Based on the results of the SEM model in the previous
section, interaction terms of SES and four dimensions of reform-based science learning were added as predictors for science achievement (Figure 12).

Table 20 only shows the results for the newly-added four interaction terms because the other coefficients were almost the same as the previous SEM results. Controlling for gender, self-efficacy, self-concept, SES, and four dimensions of science learning, there were no significant interaction effects between SES and four dimensions of science learning on science achievement. This indicated that the SES effect on students’ science achievement was not
conditioned on the frequencies of using reform-based learning activities. Therefore, all of dimensions of reform-based science learning (i.e., hands-on activities, student investigation, interaction, and application) did not help to minimize the achievement gap for students from different SES backgrounds in the U.S.

Table 20.
The unstandardized and standardized estimates of interaction model for U.S.

<table>
<thead>
<tr>
<th>Path</th>
<th>Unstandardized values</th>
<th></th>
<th>Standardized values</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>S.E.</td>
<td>p-value</td>
<td>Coefficient</td>
</tr>
<tr>
<td>Hands-on X SES to Achievement</td>
<td>0.134</td>
<td>3.032</td>
<td>0.965</td>
<td>0.001</td>
</tr>
<tr>
<td>Investigation X SES to Achievement</td>
<td>1.509</td>
<td>3.552</td>
<td>0.671</td>
<td>0.017</td>
</tr>
<tr>
<td>Interaction X SES to Achievement</td>
<td>-4.568</td>
<td>2.921</td>
<td>0.118</td>
<td>-0.048</td>
</tr>
<tr>
<td>Application X SES to Achievement</td>
<td>2.069</td>
<td>3.015</td>
<td>0.493</td>
<td>0.023</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hands-on X SES with</td>
<td>0.489</td>
<td>0.042</td>
<td>&lt;0.001</td>
<td>0.570</td>
</tr>
<tr>
<td>Investigation X SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hands-on X SES with</td>
<td>0.510</td>
<td>0.044</td>
<td>&lt;0.001</td>
<td>0.637</td>
</tr>
<tr>
<td>Interaction X SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hands-on X SES with</td>
<td>0.536</td>
<td>0.046</td>
<td>&lt;0.001</td>
<td>0.634</td>
</tr>
<tr>
<td>Application X SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigation X SES with</td>
<td>0.481</td>
<td>0.038</td>
<td>&lt;0.001</td>
<td>0.563</td>
</tr>
<tr>
<td>Interaction X SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investigation X SES with</td>
<td>0.533</td>
<td>0.029</td>
<td>&lt;0.001</td>
<td>0.591</td>
</tr>
<tr>
<td>Application X SES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interaction X SES with</td>
<td>0.594</td>
<td>0.045</td>
<td>&lt;0.001</td>
<td>0.705</td>
</tr>
</tbody>
</table>

Taiwan. Similarly, the SEM model was applied to investigate whether any of four dimensions of reform-based science learning (i.e., hands-on activities, student investigation, interaction, and application) promoted education equity by minimizing the achievement gap for students from different SES backgrounds. Table 21 shows only the results for the newly-added four interaction terms as the other coefficients were almost the same as the previous SEM results. Similar to the results in the U.S. sample, after controlling for gender, self-efficacy, self-concept, SES, and four dimensions of science learning, there was no significant interaction effect between SES and four dimensions of science learning on science achievement. This indicated that the
SES effect on students’ science achievement was not conditioned on the frequencies of using reform-based learning activities. Therefore, none of the dimensions of reform-based science learning (i.e., hands-on activities, student investigation, interaction, and application) helped to minimize the achievement gap for students from different SES backgrounds in Taiwan.

Table 21.
The unstandardized and standardized estimates of interaction model for Taiwan

<table>
<thead>
<tr>
<th>Path</th>
<th>Unstandardized values</th>
<th>Standardized values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>S.E.</td>
</tr>
<tr>
<td>Hands-on X SES to Achievement</td>
<td>3.827</td>
<td>4.199</td>
</tr>
<tr>
<td>Investigation X SES to Achievement</td>
<td>1.348</td>
<td>2.417</td>
</tr>
<tr>
<td>Interaction X SES to Achievement</td>
<td>-0.820</td>
<td>2.432</td>
</tr>
<tr>
<td>Application X SES to Achievement</td>
<td>-3.077</td>
<td>2.944</td>
</tr>
</tbody>
</table>

Correlation

<table>
<thead>
<tr>
<th>Path</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
<th>Coefficient</th>
<th>S.E.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands-on X SES with Investigation X SES</td>
<td>0.429</td>
<td>0.055</td>
<td>&lt;0.001</td>
<td>0.614</td>
<td>0.035</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on X SES with Interaction X SES</td>
<td>0.387</td>
<td>0.063</td>
<td>&lt;0.001</td>
<td>0.591</td>
<td>0.049</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hands-on X SES with Application X SES</td>
<td>0.399</td>
<td>0.053</td>
<td>&lt;0.001</td>
<td>0.578</td>
<td>0.034</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Investigation X SES with Interaction X SES</td>
<td>0.419</td>
<td>0.054</td>
<td>&lt;0.001</td>
<td>0.558</td>
<td>0.041</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Investigation X SES with Application X SES</td>
<td>0.414</td>
<td>0.052</td>
<td>&lt;0.001</td>
<td>0.523</td>
<td>0.030</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Interaction X SES with Application X SES</td>
<td>0.433</td>
<td>0.060</td>
<td>&lt;0.001</td>
<td>0.584</td>
<td>0.044</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Summary

This chapter reports the results that address the main foci of the study (i.e., implementation, availability, and achievement gap). First, to examine how reform-based science learning has been implemented in the science classrooms, the results of LPA are reported. In the U.S., 10% of the students fell into the highest usage reform-based learning subgroup and 40% into the high usage reform-based learning subgroup; in total, approximately 50% of the students learned science through the reform-based learning activities in most lessons. In Taiwan, 7% of
the students fell into the highest usage reform-based learning subgroup and 28% into the high usage reform-based learning subgroup; in total, approximately 35% of the students learned science through the reform-based learning activities in most lessons.

Second, this chapter reports the results of LCR and SEM to examine whether reform-based science learning was equally available for all students regardless of their SES backgrounds in the U.S. and Taiwan. Table 22 summarized the results for the null hypotheses one and two. In the U.S., when examining the reform-based learning activities as a whole, students with different SES were equally likely to be in the four science learning subgroups. This indicated that the frequency of students who learned science through the reform-based learning activities had no association with students’ SES background. However, when individualizing each dimension of the reform-based learning activities, the higher the students’ SES background were, the higher the frequency of students learned science through hands-on activities, interactive teaching, and real-life applications. The results from two analyses did not agree with each other.

In Taiwan, when examining the reform-based learning activities as a whole, the students in the most usage reform-based science learning subgroup had lower SES backgrounds compared to the other three subgroups. When individualizing each dimension of the reform-based learning activities, students’ SES background had no association with the frequency of students who learned science through any dimension of the reform-based learning activities. The results from two analyses also did not agree with each other for Taiwan.

Finally, this chapter reports the results of regression mixture modeling and SEM to examine whether reform-based science learning helped minimize the achievement gap for students from different SES backgrounds in the U.S. and Taiwan. Table 22 summarized the results for the null hypotheses three and four. In the U.S., when examining the reform-based
science learning activities as a whole, the students in the four subgroups have the same SES-associated achievement gap. When individualizing each dimension of the reform-based science learning activities, the frequency of students who learned science through each dimension of the reform-based science learning activities has no association with SES-associated achievement gap.

In Taiwan, when examining the reform-based science learning activities as a whole, the students in the highest reform-based learning subgroup have most SES-associated achievement gap. When individualizing each dimension of the reform-based science learning activities, the frequency of students who learned science through each dimension of the reform-based science learning activities has no association with SES-associated achievement gap.

Table 22.

<table>
<thead>
<tr>
<th>Null hypothesis (H₀)</th>
<th>Decision</th>
<th>U.S.</th>
<th>Taiwan</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₀₁: No significant difference on SES comparing a specific subgroup to the most usage reform-based science learning subgroup</td>
<td>Accept</td>
<td>Reject</td>
<td>(H,M,L&gt;Highest)</td>
</tr>
<tr>
<td>H₀₂: No significant association between SES and frequency to learn science through each of four dimensions of reform-based science learning</td>
<td>Hands-on: Reject (+)</td>
<td>Accept</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Investigation: Accept</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction: Reject (+)</td>
<td>Accept</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Application: Reject (+)</td>
<td>Accept</td>
<td></td>
</tr>
<tr>
<td>H₀₃: No significant difference for the influences of students’ SES on their science achievement across different science learning subgroups</td>
<td>Accept</td>
<td>Reject</td>
<td>(Highest&gt;H,M,L)</td>
</tr>
<tr>
<td>H₀₄: No interaction effect between SES and each of four dimensions of reform-based science learning on achievement</td>
<td>Hands-on: Accept</td>
<td>Accept</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Investigation: Accept</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interaction: Accept</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Application: Accept</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. + indicated positive association
> indicated significant difference between two subgroups
CHAPTER FIVE

SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS

Chapter Five discusses the results of the current study. The first section summarizes the purposes, design, research question, and conclusions for this study. The second section discusses the results for each research question and the related research. Finally, the third section describes the recommendations for future study.

Summary of the Study

Responding to rapid economic development and social change, science education reform has been implemented around the world for the past two decades. However, the largely continual and steady effort for reform has had little effect on real classroom practice. In the U.S., teachers’ beliefs and values about learning and teaching, parental resistance, and limited teaching ability were thought to be the obstacles for changing classroom practices (Anderson, 2002). In Taiwan, examination tradition has been deeply rooted in the education system and played a major role in teaching practice and emphasis. Considering the obstacles the two countries faced when implementing science education reform, this study aimed to evaluate to what extent students learned science through reform-based learning in schools for the U.S. and Taiwan.

When the U.S. was working toward reform-based science teaching, the *NSES* (NRC, 1996) set a goal for the science education reform:

> The intent of the Standards can be expressed in a single phrase: Science standards for all students. The phrase embodies both excellence and equity. The Standards apply to all students, regardless of age, gender, cultural or ethnic background, disabilities, aspirations, or interest and motivation in science. (p. 2)

The science education reform not only focused on excellence, where the aim was to improve student achievement, but also emphasized equity, where the purpose was to reduce achievement gaps traditionally experienced by students from different backgrounds. In this
study, students’ socio-economics (SES) background was interest because SES was the strongest predictor for students’ achievement, and it was difficult to mitigate the impact of SES by any school factor (Coleman et al., 1966). In the literature, Lynch (2000) described equity issues as equality of inputs and outputs. Under equality of inputs, all students should have the same level of educational opportunity and resources regardless of their demographics or background. The United States has a decentralized education system based upon the federal Constitution, which grants power over education to the states and local authorities. The aim of decentralization is to give more power to local leaders and school officials, who supposedly know more about local educational situations than national officials, and who have more ability to lobby for resources from local community. However, education resources have been distributed unequally, and the school systems in low-SES communities are often under-resourced (Barbarin, 2008). Compared to the U.S., Taiwan's education policy would seem to help promote educational equity because the educational system in Taiwan is centralized. What teachers teach followed national curriculum, and therefore, every student receives the same curriculum at the same grade. Moreover, in order to ensure equal distribution of resources across schools, the central government provides additional funding and grants for schools of remote or mountain area. Every student should receive equal educational opportunities and resources (Chen, Crockett, Namikawa, Zilimu, & Lee, 2012). Based upon these different education systems, this study helped to evaluate whether science education reform in each country achieved its goal of equity. That is, it evaluated whether all students have the same level of chance to learn science through reform-based learning activities regardless of their SES background in the U.S. and Taiwan.

When the goal of equality of outputs ideal was achieved, all students should have the same level of success regardless of their demographics background. In the U.S., the science
achievement gap between students with high- and low-SES family background was a long-standing issue. The National Assessment of Educational Progress (NAEP) results showed that from 1996 to 2011, the achievement gaps between students from low-income and high-income families remained unchanged for fourth, eighth, and twelfth grade students (Grigg, Lauko, & Brockway, 2006; NCES, 2012). The similar situation also happened in Taiwan. Cheng and Hung (2006), addressing equity issues in Taiwan’s education system, showed that students entering top-tier universities were from high-SES families at disproportionately larger numbers than from low-SES families.

Regarding the causes of the SES-associated achievement gap, students from low-SES families were deficit in financial capital (e.g., household income), human capital (e.g., education), and social capital (e.g., social networks and connections) in comparison to their peers. Therefore, they had less access to educational resources or learning materials, less parental involvement with learning and development, and lack of successful peer models (Bradley & Corwyn, 2002). With low-SES students coming from family environments with so many disadvantages against successful learning, schools became an important avenue for providing a learning environment with more educational opportunity and resources to make their learning outcomes as successful as their peers. Reform-based science teaching and learning focused on engaging students in hands-on activities, student investigation, teacher-student interaction, and real-life applications (OECD, 2009). The reform-based learning activities provided opportunity for students to work on science materials, conduct science experiments, communicate scientific ideas with teachers and peers, and relate science knowledge to their everyday life. Through the opportunity to work on these learning activities, students from low-SES families may overcome their disadvantages of family environment and learn as well as their peers. This study evaluated
whether reform-based science teaching promoted education equity to help minimize the achievement gap for students from different SES backgrounds in the U.S. and Taiwan.

The main purpose of this study was to investigate whether reform-based science teaching promoted education equity through being available and beneficial for students from different SES backgrounds in the U.S. and Taiwan. The data were obtained from U.S. and Taiwan 15-year-old students who participated in the Program for International Student Assessment (PISA) 2006. The PISA’s definition of SES combined economic capital (e.g., material resources), cultural capital (non-material resources), and social capital (e.g., social connections). The PISA’s SES scores used parents’ occupational status, parents’ education level, and household possessions as indicators. The PISA 2006 is a nationally representative dataset and results can be generalized to the whole nations and can be used for a nation-wide evaluation of reform-based science teaching in the U.S. and Taiwan. This study answered the following three research questions:

1) How broadly has reform-based science learning been implemented in classrooms in the U.S. and Taiwan?

2) Did all students have the same opportunity to learn science through the reform-based learning activities regardless of their SES background in the U.S. and Taiwan?

3) Did students who had more opportunities to learn science through the reform-based learning activities have a smaller SES-associated achievement gap in the U.S. and Taiwan?

First, how reform-based science learning was implemented was investigated. A Latent Profile Analysis (LPA) was used to classify students into different science learning subgroups based on the 15 items concerning students’ reform-based learning activities in their science
classrooms both in the U.S. and Taiwan. The LPA belongs to a broad family of statistical models, known as finite mixture models, that are used to empirically derive latent subgroups or classes to represent an unknown overall population. In this study, different science learning subgroups represented different usage and foci of the learning activities that students experienced in their classroom. The optimal number of latent subgroups was determined by fit indices, classification accuracy, and the interpretability of each subgroup. The Akaike information criterion (AIC), Bayesian information criterion (BIC), adjusted BIC, Vuong–Lo–Mendell–Rubin (VLMR) test, and entropy were used to determine the number of subgroups. Subgroups were interpreted and named based on item-response probabilities of each item for the usage and foci of the students’ classroom learning activities for each subgroups. Latent subgroups’ prevalence were also estimated to understand what proportions of students belonged to each subgroup.

Second, this study investigated whether reform-based science learning activities were equally available for all students regardless of their SES background through two different analyses. The first analysis examined the reform-based science learning activities as a whole (15 items), and Latent Class Regression was used for testing whether students’ SES background was a predictor of their science learning subgroup membership. The second analysis individualized the reform-based science learning activities into four science learning dimensions: hands-on activities (4 items), student investigation (3 items), interaction (4 items), and application (4 items). Structural equation modeling (SEM) was used to test the relationship between students’ SES and the frequency of students who learned science through each of the four science learning dimensions. The fit indices of comparative fit index (CFI), root-mean-square error of approximation (RMSEA), and standardized root mean squared residual (SRMR) were used for evaluating the model fit.
The third research question was to examine the effects of reform-based science learning activities on the achievement gap among students from different SES backgrounds. This research question was also investigated through two different analyses. The first analysis examined the reform-based science learning activities as a whole and Regression Mixture Modeling was used. Regression mixture modeling allowed the effect of students’ SES on achievement to be different for the different latent subgroups of students. The regression mixture modeling was used to test whether students’ SES on their science achievement was different among different science learning subgroups. The second analysis was also individualized into each dimension of science learning and the four interaction terms between students’ SES and four dimensions (i.e., hands-on activities, student investigation, interaction, and application) were added to the SEM models. This study tested whether there were significant interaction effects.

Conclusions

On the basis of the data, assumptions, analyses, and limitations, the following conclusions can be made:

Implementation of Reform-based Science Learning

1) Students in both the U.S. and Taiwan had more opportunities to learn science through the activities of hands-on activities, interactive teaching, and real-life applications than student investigations.

2) Students in the U.S. had more opportunity to learn science through the reform-based learning activities than students in Taiwan. In the U.S., in total approximately 50 percent of the students learned science through the reform-based learning activities in most lessons. Only approximately 10 percent of the students seldom learned science through the reform-based learning activities. In Taiwan, whereas approximately 35 percent of the
students learned science through the reform-based learning activities in most lessons, approximately 17 percent of the students seldom learned science through the reform-based learning activities.

**Availability of Reform-based Science Learning**

3) This study found inconsistent results regarding the availability of reform-based science learning in the U.S. sample. When examining the reform-based learning activities as a whole, students with different SES were equally likely to be in the four science learning subgroups. The results of the study showed that the frequency of students learned science through the reform-based learning activities had no association with students’ SES background. When individualizing each dimension of the reform-based learning activities, higher SES students had more opportunity to learn science through hands-on activities, interactive teaching, and real-life applications than lower SES students.

4) This study found inconsistent results regarding the availability of reform-based science learning in Taiwan sample. When examining the reform-based learning activities as a whole, the students in the most usage reform-based science learning subgroup had lower SES backgrounds in comparison to the other three subgroups. When individualizing each dimension of the reform-based learning activities, students’ SES background had no association with the frequency of students who learned science through any dimension of the reform-based learning activities.

**Achievement Gap and Reform-based Science Learning**

5) In the U.S., when examining the reform-based science learning activities as a whole, there was no significant difference for the effect of students’ SES on their science achievement across four different science learning subgroups. Consistently, when
individualizing each dimension of the reform-based science learning activities, the frequency of students who learned science through each dimension of the reform-based science learning activities had no association with SES-associated achievement gap.

6) This study found inconsistent results regarding the availability of reform-based science learning in the Taiwan sample. In Taiwan, when examining the reform-based science learning activities as a whole, students’ SES had significantly more effect on students’ achievement in the highest reform-based usage subgroup than on the other three subgroups. However, when individualizing each dimension of the reform-based science learning activities, reform-based science learning activities had no association with SES-associated achievement gap.

**Discussions**

**Implementation of Reform-based Science Learning**

**United States.** The data of this study came from students’ self-reported data, and the results indicated that around 50 percent of the students learned science through the reform-based science learning activities in most of their lessons. The results were similar to the national teacher survey from Horizon Research, Inc. (Hudson, McMahon & Overstreet, 2002) indicating that 50 percent of the teachers had implemented the recommended practices from the national standards in their classroom at least to a moderate extent. However, the results were different from other observational studies (Simmons et al., 1999; Roehrig & Lull, 2004) indicating that only approximately 10-20 percent of teachers in their studies implemented reform-based instructions in their science teaching. The difference may due to different evaluation standards held by teachers, students, and education researchers. Teachers and students just evaluated themselves by whether or how frequently they learned or taught science through a specific
activity. On the other hand, education researchers usually used some rigorous observational protocols and analysis matrix, which may add more criterion and higher standards, when evaluating teaching practices. Therefore, what students’ and teachers’ perception of what they were doing in the classroom would be different from education researchers’ observation. Simmons et al. (1999) found that observations of teachers’ teaching practices contrasted with teacher beliefs that while teachers believed they taught as student-centered approaches, in reality, they taught in teacher-centered ways. Simmons’ study suggested that teachers may want to teach a more student-centered approach because they believed that was an effective teaching approach. However, due to constraints, they did not translate these beliefs into real classroom practice and did not discover the inconsistency between what they believed and what they practiced.

Even though the data of this current study was limited by coming from students’ perception, this study did find that there approximately 10 percent of the students seldom learned science through the reform-based learning activities. The reform-based activities provided learning environments where students actively engaged in hands-on activities, interpreted and explained data, negotiated understandings of the findings with other students, and connected their understandings to the real world. Although the results of this study did not find evidence that these 10 percent of students had lower science achievement, this study found that these students had lower self-efficacy and self-concept compared to other students with more opportunity to learn science through reform-based learning activities. Taraban et al. (2007) found that students who learned through reform-based learning activities gained more content knowledge and knowledge of process skills, developed critical thinking and problem-solving skills, and had more interest in learning science. Thus, in order to better prepare students for the
modern society, it would be optimal to let every student have the opportunity to learn science through reform-based activities.

**Taiwan.** This study found that only 35 percent of the students in Taiwan have the opportunity to learn science through the reform-based learning activities in most lessons and that approximately 17 percent of the students seldom learned science through the reform-based learning activities. The results confirmed what was documented in the Taiwanese education literatures. Wang and Lin (2009) reported that in a test-driven learning environment, middle school science teachers exclusively concentrated on the content knowledge covered in the textbook and tests. Therefore, the science teachers responded that parents and schools’ evaluation for the quality of a teacher was only based on the average scores of students’ science tests posted in front of a classroom. The students responded that the focus of learning in science was on drills or repetition in order to get a high score on the test, making them feel more self-confident. Neither teachers nor students perceived inquiry-learning and classroom interaction as an important components in science classroom (Wang & Lin, 2009).

The results of the current study revealed that the implementation of the curriculum reform in Taiwan did not attend to its goals to change classroom practice and environment. Gibbs (1992) concluded that most of reform policies did not come up with anticipated results in changing learning environment and practice, and reflection, hearing feedback, and policy modification were required. Fullan and Stiegelbauer (1991) found that many factors obstructed educational change, such as absence of communication of between the policy makers and practicing teachers, as well as not rewarding the implementation of change. In order to make sure that the goals of science education reform in Taiwan can be attended, continuing efforts should be required from researchers, policy-makers, curriculum designers, and school teachers. Given
the fact Taiwan society focuses on getting a high score in high school and college entrance examination, the reform of the school entrance examination is needed and should include assessing students’ inquiry-ability and science-process skills. Hopefully, science teachers would shift their teaching focus toward more reform-based learning activities with the intention of increasing students’ test scores.

The results of the current study also identified that the highest usage reform-based learning subgroup consisted of only 7 percent of students in Taiwan. This subgroup grabbed the author’s attention because the results for this subgroup were opposite to the author’s expectation. While the other three subgroups had no significant differences on students’ SES, achievement, and the SES-associated achievement gap, students in this highest-usage subgroup had the lowest SES background, the lowest science achievement, and the largest SES-associated achievement gap. This contradicted the author’s expectation because students in this highest-usage subgroup had the most opportunity to learn science through reform-based learning activities. It was not intuitive that there was a group of students who came from low-SES family backgrounds and had low science achievement, but, at the same time, had the most opportunity to learn science through reform-based science teaching. Based on the literature review (Chen et al., 2012; Hsueh, 2007; Chang, 2007), one of the possibilities was that these students come from the minority groups who live in the mountain or rural area in Taiwan. These minority groups were culturally, economically, and language disadvantaged. Their family had a low income and were busy earning a living. These low SES students’ parents in general did not participate and support children’s learning activities, cared less about children’s learning, and provided fewer educational resources (Tsai, 2006). Schools in the mountain or rural area were not easily accessible. These rural schools provided fewer professional development opportunities, and it
would be time-consuming for teachers to commute between home and school. Consequently, the majority of the teachers were not willing to work in these remote schools. Also, the schools in the rural and mountain area were short of teachers. In response to these schools’ needs, the Department of Education in Taiwan purposefully assigned newly graduated teachers to these rural or mountain area schools. These newly graduated teachers were just out of school with fresh knowledge and skills of reform-based science teaching. These recent graduates were trained in their teacher education program with the latest knowledge and were more willing to implement reform-based learning activities in the classroom. Moreover, these teachers in the remote schools usually received less pressure from parents and did not need to focus their teaching on entrance examinations. Therefore, teachers in the rural or mountain areas would be more likely to implement reform-based learning activities. In addition, in order to ensure equal distribution of resources across schools, the central government provided additional funding and grants for schools of remote or mountain areas. This was to ensure that every student received equal educational opportunities and resources (Chen, Crockett, Namikawa, Zilimu, & Lee, 2012). These were the possible reasons that this non-mainstream group of students had more opportunity to learn science through reform-based learning activities. Because the PISA 2006 Taiwan data did not supply students’ ethnicity information, it was unclear whether this group of students came from minority groups who lived in rural or mountain areas. However, the results of this current study raised the concern that even though these additional funding and grants for schools of remote or mountain areas, students in rural or mountain area still had worse science achievement. Researchers, policy-makers, and school teachers in Taiwan should pay more attention to this group of students to help them learn science as well as the other students.
Comparing to the U.S., students in Taiwan had less opportunity to learn science through the reform-based learning activities. However, students in Taiwan had higher PISA 2006 science assessment scores. This result did not indicate that reform-based learning activities could not improve students’ achievement. There were many reasons to explain the achievement differences among different countries when doing research on an international comparison study. One of the most common explanations was the time students spent to learn science inside and outside of school (Kobarg et al., 2011). Students in Taiwan usually received after-school instruction in cram schools. Cram schools were a unique educational product coming out of the test-driven learning environment, specifically in East Asian countries (Tsai & Kuo, 2008). In order to improve test scores and prepare for high school and college entrance examinations, students in Taiwan usually spent five to ten hours a week in cram schools. At home, students not only were required to finish homework assignments by school teachers but also by cram schools. Thus, students in Taiwan spent more time on leaning science then students in the U.S. The time students spending on leaning science was a prerequisite for students to develop a basic understanding of scientific concept (Kobarg et al., 2011).

**Availability of Reform-based Science Learning**

**United States.** The current study did not find that there was an association between students’ SES and their frequency of learning science through the reform-based science learning activities when examining the reform-based science learning activities as a whole. However, this study found that students with higher SES had a higher chance to learn science through hands-on activities, interactive teaching, and real-life applications when individualizing each dimension of the reform-based science learning activities. Peske and Haycock (2006) found that youths from poor families were exposed to an environment with fewer experienced teachers and less
academic support. Moreover, under a decentralized education system, schools’ education resources usually were constrained by the local community. Aikens and Barbarin (2008) found that the school systems in low-SES communities were often educationally under-resourced. Issues about less qualified teachers and educational resources were barriers or obstacles for implementing reform-based science teaching (Anderson, 2002). Thus, providing more educational resources and attracting more qualified teachers to schools in low-SES communities were needed to ensure the availability of reform-based science teaching for all students regardless of their SES.

**Taiwan.** Other than 7 percent of students with lower SES who had more opportunity to learn science through reform-based science learning activities, students’ SES had no relationship with the opportunity that students had to learn science through reform-based science learning activities in Taiwan. These results confirmed Chen et al.’s (2012) research that the Education Department controlled the implementation of curriculum, the distribution of education resources, and the allocation of qualified teachers. Therefore, every student received the same curriculum regardless of where students lived and who their teachers were, and every school had equal education resources regardless where the schools were and who was the leadership for school. Thus, whether or not students had opportunity to learn through reform-based science teaching tended to have no relationship with students’ SES background under a more centralized education system.

**SES Achievement Gap and Reform-based Science Learning**

Other than 7 percent of students with lower SES in Taiwan, either examining the reform-based science learning activities as a whole or as four different dimensions, the frequency of students who learned science through the reform-based science learning activities had no
association with SES-associated achievement gap both in the U.S. and Taiwan. Thus, the answer for the main focus of the current study that whether reform-based science teaching promoted education equity through decreasing SES-associated achievement gap. While most of the research towards reform-based science teaching were investigated at school level or district level (Marx et al., 2004), there was less research at the national level. This study used the data from the PISA 2006, which is the nationally representative data and the results can be generalized to the whole nations. The PISA 2006 supplied different perspectives and served as a reliable national evaluation for implementing reform-based science teaching in the U.S. and Taiwan. This study informed policy makers, researchers, and teachers that the results of the implementation of reform-based science teaching in classroom did not meet the expectation for the education equity. These results could be due to teacher qualification and trainings. It is possible that not all teachers were well prepared for conducting these reform-based learning activities, especially teaching these activities for students from different SES background. Several studies, which had successfully improved low-SES students learning through reform-based teaching, suggested that a curriculum should be carefully developed, assessments should be aligned with curriculum materials, and professional development should be designed to change teachers’ traditional practice (Geier et al., 2008). Other studies suggested that school science was usually disconnected from the multicultural contexts of students’ lives. It was important for teachers to identify intersections between students’ everyday knowledge and scientific practice and use these intersections to inform their reform-based science teaching practices (Warren et al., 2001).

Another reason for being unable to find the relationship between reform-based learning and students’ SES-associated achievement gap may due to the limitation of the research design.
The PISA student questionnaire asked about the learning activities in their classroom for the current school year, while student science achievement score was a result of many years of collective learning and life experiences. One year’s of teaching practice in school might not be enough to compensate low-SES students’ learning disadvantages from their family. Failing to control students’ prior science achievement in this study might be the result of failing to find the relationship between reform-based learning and students’ SES-associated achievement gap.

Recommendations for Future Study

Research Methodology

The issues of the different results from two different analyses in this study should be noted. The current study did not find an association between students’ SES and their frequency of learning science through the reform-based science learning activities when examining the reform-based science learning activities as a whole. When looking at each reform-based dimension, this study found that students’ SES had positive associations with three of four dimensions of the reform-based science learning activities. One difference for these two analyses was that one examined reform-based science learning activities as a whole and the other as an individual dimension. The other difference was that one analysis used latent profile analysis (LPA), which treated the latent construct, reform-based science learning, as a categorical variable, and the other analysis used structural equation modeling (SEM), which treated the latent construct as continuous variables. Whether the latent construct of reform-based science learning should be treated as a categorical variable or a continuous variables remained a question. Latent variable research suggested that complex multi-factor and multi-class models often revert to simpler forms when factors and classes were estimated within a single model framework (Nylund, 2007). Nylund argued that when multi-factor, multi-class real data were
examined using either latent class analysis or exploratory factor analysis, there was a tendency to overestimate the number of factors and classes. Thus, multi-factor, multi-class models, also known as factor mixture models (FMMs), were suggested for the future study to compare whether LPA (multi-class models), SEM (multi-factor models) or FMM (multi-factor, multi-class models) would be the best way to represent the data about classroom practices.

Another issue was related to the form of assessment used to evaluate the outcome of reform-based science teaching. It remained as a question whether science achievement can be measured by a paper and pencil test effectively when reformed-based learning activities emphasized on students’ learning process and inquiry skills. The traditional paper and pencil test may not be an appropriate measure of performance for these scientific skills. Although PISA 2006 claimed that their science assessment focused on assessing students’ three scientific competencies--identifying scientific issues, explaining scientific phenomena, and using scientific evidence (OECD, 2006)--some research critiqued PISA 2006 science assessment items were “standard, decontextualized process questions embedded in a brief, but unnecessary story” (Sadler & Zeidler, 2009, p.916). When taking the test, students only needed to recall scientific knowledge and facts and did not need to investigate the information provided in the question setting. Thus, whether PISA assessment items were aligned with research objectives, curriculum materials, and classroom practices would be an important issue for the future study.

As mentioned previously, this study was not able to control for students’ prior achievement using the PISA 2006. Future studies should consider controlling for students’ prior achievement when evaluating the outcome of reform-based teaching. In addition, this study only used a standardized test score as an indicator of achievement; multiple measures of achievement could be used in the future studies.
Research Direction

**School Factors.** The PISA 2006 supplied both student information and school information. The current study only investigated the student factors and found the association between students’ SES and the opportunity to learn science through reform-based learning activities in the U.S. This study suggested that lack of education resources for schools in low SES community might be the reason. The next step is to investigate the school level factors in the PISA 2006. Does lack of school resources restrain teachers from implementing reform-based learning activities? What other factors in the PISA data, such as school organization, school decision-making, school climate, and teacher and school autonomy, would support teachers to implement reform-based learning activities? What kinds of support do teachers need when implementing reform-based learning activities for low-SES students? These research questions help future researchers to understand how schools can support science education reform and help low SES students’ science learning.

**PISA 2015.** The Programme for International Student Assessment (PISA) in 2015 will have science as the major assessment domain again. After 10 more years’ effort for science education reform, it is interesting to compare reform-based science learning activities implemented in science classrooms between 2006 and 2015. Are reform-based science learning activities more common in real classroom practice? Is the quality of implementing reform-based science learning activities improved and do these activities successfully influencing students’ learning outcomes? Does reformed-based science learning help decrease students’ SES-
associated achievement gaps? All of these questions are important for the science education community as the community continues to improve science education around the world.

**Next generation Science Standard (NGSS).** At the time of PISA 2006, the U.S. had been using the *National Science Education Standards (NSES)* (NRC, 1996) and *Benchmarks for Science Literacy* (AAAS, 1993) to guide the development of the current science standards. While these two documents have proven to be both effective for students’ learning, they are around 15 years old. During these 15 years, major developments have taken place in the world of science and in our understanding of how students learn science effectively. Thus, the NRC, the National Science Teachers Association (NSTA), the AAAS, and Achieve have developed the *Next Generation Science Standards (NGSS)* (Achieve, 2013). The final version of the NGSS was released in April 2013. The NGSS focuses on science practices, which emphasize students engaging in scientific inquiry and coordinating both science knowledge and process skills simultaneously. The NGSS also includes engineering practices, which emphasize engaging students to apply science and mathematics to design solutions for problems as engineers’ routines. It will be interesting to investigate whether the curriculum with these engineering practices will influence students’ learning outcomes among different SES groups. It will also be interesting to investigate how teachers can help students with low-SES background engage in these engineering practices.
Reference


American Association for the Advancement of Science. (1989). *Science for All Americans*. Washington, DC.


VITA

Nai-En Tang was born on January 18, 1978, in Taipei, Taiwan (R.O.C.). After obtaining his B.A. in Physics at the National Cheng Kung University, he received a M.S. in Electrophysics at the National Chiao Tung University. After working as an engineer for two semiconductor companies, he served as a physics teacher for a one-year teaching practicum in a high school in Taipei. He is currently studying for a Ph.D in Teaching, Learning, and Curriculum with an emphasis on Science Education and a M.A. in Statistics at the University of Missouri.