
Theses and Dissertations

Fall 2013

Longitudinal analysis of standardized test scores of students in the science writing heuristic approach

Niphon Chanlen
University of Iowa

Copyright 2013 Niphon Chanlen

This dissertation is available at Iowa Research Online: <http://ir.uiowa.edu/etd/4953>

Recommended Citation

Chanlen, Niphon. "Longitudinal analysis of standardized test scores of students in the science writing heuristic approach." PhD (Doctor of Philosophy) thesis, University of Iowa, 2013.
<http://ir.uiowa.edu/etd/4953>.

Follow this and additional works at: <http://ir.uiowa.edu/etd>



Part of the [Science and Mathematics Education Commons](#)

LONGITUDINAL ANALYSIS OF STANDARDIZED TEST SCORES OF STUDENTS
IN THE SCIENCE WRITING HEURISTIC APPROACH

by
Niphon Chanlen

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Science Education
in the Graduate College of
The University of Iowa

December 2013

Thesis Supervisor: Professor Brian Hand

Copyright by
NIPHON CHANLEN
2013
All Rights Reserved

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Niphon Chanlen

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Science Education at the December 2013 graduation.

Thesis Committee: _____
Brian Hand, Thesis Supervisor

Deborah Linebarger

Catherine Welch

Soonhye Park

Ingrid Ukstins Peate

ACKNOWLEDGMENTS

I would like to express the deepest appreciation to my committee chair Professor Brian Hand, for the continuous support of my Ph.D study and research, for his patience, motivation, enthusiasm, immense knowledge. His continually and persuasively conveyed a spirit of adventure in regard to research and scholarship, and an excitement in regard to teaching. I could not have imagined having a better advisor and mentor for my dissertation and throughout my entire doctoral program. Without his supervision and constant help this dissertation would not have been possible. I would also like to thank my dissertation co-advisor Dr. Deborah Linebarger for her feedback and support. It would not have been possible without her insightful.

My sincere thanks also goes to my committee members Dr. Soonhye Park, Dr. Catherine Welch, and Dr. Ingrid Ukstins for their encouragement, collaboration, insightful comments, and challenging questions. In addition, a thank you to Pam Nystrom for the helpful collaboration.

I would like to thank the Institute for the Promotion of Teaching Science and Technology (IPST), Bangkok, Thailand for this great opportunity and its financial support.

I thank my fellow graduate students, colleagues, and friends in science education program for the stimulating discussions, support, and friendship.

Last but not the least, I would like to thank my family for unconditional love and spiritual support. The strength and power they gave me have carried me to this far. I would not have made it without them.

ABSTRACT

The purpose of this study was to examine the longitudinal impacts of the Science Writing Heuristic (SWH) approach on student science achievement measured by the Iowa Test of Basic Skills (ITBS). A number of studies have reported positive impact of an inquiry-based instruction on student achievement, critical thinking skills, reasoning skills, attitude toward science, etc. So far, studies have focused on exploring how an intervention affects student achievement using teacher/researcher-generated measurement. Only a few studies have attempted to explore the long-term impacts of an intervention on student science achievement measured by standardized tests.

The students' science and reading ITBS data was collected from 2000 to 2011 from a school district which had adopted the SWH approach as the main approach in science classrooms since 2002. The data consisted of 12,350 data points from 3,039 students. The multilevel model for change with discontinuity in elevation and slope technique was used to analyze changes in student science achievement growth trajectories prior and after adopting the SWH approach.

The results showed that the SWH approach positively impacted students by initially raising science achievement scores. The initial impact was maintained and gradually increased when students were continuously exposed to the SWH approach. Disadvantaged students who were at risk of having low science achievement had bigger benefits from experience with the SWH approach. As a result, existing problematic achievement gaps were narrowed down. Moreover, students who started experience with the SWH approach as early as elementary school seemed to have better science

achievement growth compared to students who started experiencing with the SWH approach only in high school.

The results found in this study not only confirmed the positive impacts of the SWH approach on student achievement, but also demonstrated additive impacts found when students had longitudinal experiences with the approach. By engaging in the argument-based classrooms where teachers value students' prior knowledge, encourage students to take control of their learning, and provide non-threatening environment for students to developing big ideas through negotiation, student's achievement can be enhanced. The results also started to shed some light on sustainability of the SWH approach within the school district.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER I INTRODUCTION.....	1
Inquiry in Science Education.....	1
The Effectiveness of Inquiry	2
Argument-Based Inquiry	3
Purpose of the Study.....	5
Research Questions of the Study.....	5
Overview of the Study.....	6
Index of Abbreviations	7
CHAPTER II LITERATURE REVIEW	9
Inquiry-Based Instructional Approach	9
The BSCS 5E Instruction Model.....	12
Modeling-Based Inquiry.....	15
Argument-Based Inquiry	18
Similarities among Inquiry-Based Instructional Models.....	22
Connecting to the Next Generation of Science Standard	24
Effectiveness of Inquiry-Based Instruction	26
Equity in Science Classroom.....	31
Beyond the Effectiveness of Inquiry	33
Impacts of Inquiry on the Standardized Test Achievement.....	34
The Effectiveness of the Science Writing Heuristic (SWH) Approach	38
Summary of Literatures	44
CHAPTER III METHODOLOGY	46
Context of the Study	46
The Multilevel Model for Change with Discontinuity in Elevation and Slope	48
Data.....	51
The Standardized Test	51
Data Collection	52
Variables	54
Data Organization.....	60
Creating a Longitudinal Data Set	60
Handling Missing Data.....	62
Descriptive Statistic Analysis	63
Data Analysis Procedure.....	70
Summary.....	74
CHAPTER IV RESULTS.....	75
Data Exploration.....	76
Science ITBS Achievement of Iowa Students.....	79
Science Achievement Model Building	82
Building Leve-1 Model	83
Building Level-2 Models.....	95
The Impact of Starting Level.....	127

Summary	134
CHAPTER V DISCUSSION AND IMPLICATIONS	136
Research Questions.....	136
Research Question I.....	136
Research Question II	143
Research Question III.....	148
Discussion.....	150
Implications	156
Moving on to the Next Generation of Science Standard	157
The Sustainability of the School-Wide Adoption of the SWH Approach	157
The Impact of the SWH Approach on Standardized Test	159
Implications for Future Research.....	160
Limitations.....	162
REFERENCES	164

LIST OF TABLES

Table 2.1 Summary of Five Phases of the BSCS 5E Instructional Model	13
Table 2.2 The SWH templates for Teacher and Student	21
Table 3.1 Number of Participating Students in Each Cohort (Number of students exposing to the SWH was highlighted)	55
Table 3.2 List of Variables Used in Multilevel Models Examining the Student's ITBS Growth Trajectories	60
Table 3.3 Example of a Person-period Data Set	61
Table 3.4 Example of a Person-level Data Set	61
Table 3.5 Descriptive Statics Showing Number of Students, Mean and Standard Deviation of Science ITBS Scores.....	64
Table 3.6 Average and Percentage of Data Point Available in Each Cohort	66
Table 3.7 Total Number and Percentage of Students Based on Student's Backgrounds	68
Table 4.1 Results of Chi Square Test of Independence Showing Number of Student and Standardized Residuals	78
Table 4.2 Weighted average science ITBS and gain scores of Iowa students from 2001-2011	80
Table 4.3 Average science ITBS achievement gaps of Iowa students from 2001- 2011	81
Table 4.4 Results of Level-1 Model Building	84
Table 4.5 Results of Level-2 Models Building.....	97
Table 4.6 Results of Final Model Centering at Grade 3 to 11	126
Table 4.7 Results of Models Exploring First Timing of Exposure.....	132

LIST OF FIGURES

Figure 2.1 The Three Spheres of Activity for Scientists and Engineers (NRC, 2012).....	25
Figure 3.1 Examples of Multilevel Models for Change with Discontinuity.....	50
Figure 4.1 Students’ Science Achievement Trajectories from Grade 3 to Grade 11, Before and After Exposure the SWH Approach.....	92
Figure 4.2 Students’ Science Achievement Trajectories from Grade 3 to Grade 11 before and after Exposure to the SWH Approach Compared to Iowa Normative Achievement.	94
Figure 4.3 Science Achievement Trajectories before and after Exposure to the SWH Approach for Male and Female Students.....	100
Figure 4.4 Science achievement scores trajectories before and after exposure to the SWH approach for male and female students compared to Iowa normative achievement	102
Figure 4.5 Science Achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students	105
Figure 4.6 Science Achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students Compared to Iowa Normative Achievement	106
Figure 4.7 Science Achievement Trajectories before and after Exposure to the SWH Approach for Low-SES and Non-low-SES Students.....	110
Figure 4.8 Science achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students Compared to Iowa Normative Achievement	112
Figure 4.9 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels.....	116
Figure 4.10 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Compared to Iowa Normative Population.....	117
Figure 4.11 Science Achievement Trajectories before and after exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Controlling for Gender, IEP, and SES Status.	124
Figure 4.12 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Compared to Iowa Normative Population Controlling for Gender, IEP, and SES Status.....	125

Figure 4.13 Science Achievement Trajectories for Different Timing of the First Exposure of the SWH Approach131

Figure 4.14 Science Achievement Trajectories for Different Timing of the First Exposure of the SWH Approach Compared to Iowa Normative Population133

CHAPTER I

INTRODUCTION

Inquiry in Science Education

The reform movement in science education has called for the educational system to prepare students to be scientifically literate citizens for the 21st century. The reforms have emphasized the needs for teachers and students to join together as a learning community focusing on learning science. Furthermore, the National Research Council (NRC, 1996) has emphasized inquiry as central to science teaching and learning. Through inquiry, students will actively develop their understanding of science by combining scientific knowledge with reasoning and thinking skills. To achieve this goal, long term commitment and support from educational systems are required. Moreover, the National Research Council recently developed the framework for K-12 science education (NRC, 2012) and purposed the Next Generation Science Standards (NRC, 2013).

Recognizing the strengths and weaknesses from previous research, the Framework and the NGSS highlight the need for developing greater depth and rigor in K-12 science schooling by serving as a guideline for developing sufficient knowledge and appreciation of science. The Framework suggested that all students need to experience a wide range of instructional approaches that focus on four strands of proficiency: 1) Knowing, using, and interpreting scientific explanations of the natural world. 2) Generating and evaluating scientific evidence and explanations. 3) Understanding the nature and development of scientific knowledge. 4) Participating productively in scientific practices and discourse. (NRC, 2012, p. 251)

The Effectiveness of Inquiry

Since 1980, a numbers of studies have demonstrated positive impacts of various inquiry-based instructions on teachers and students (Bredderman, 1983; Shymansky, Kyle, & Alport, 1983; Minner, Levy, & Century, 2010; Furtak, Seidel, Iverson, & Briggs, 2012; Schroeder, Scott, Tolson, Huang, & Lee, 2007). Engaging in an inquiry classroom does not only help students construct better understanding of scientific concepts, but also other related skills, i.e., science process, reasoning skills, critical thinking skills, creativity, etc. However, studies (Ruiz-Primo, Shavelson, Halmilton, & Klein, 2002; Tretter and Jones, 2003) usually reported relatively smaller or null impacts when assessing through the standardized achievement tests. In addition, inquiry offers learning opportunities and challenges for all students, especially students who are usually underserved in science education. Studies (Secker, 2002; Lee, Buxton, Lewis, and Relay, 2006) have demonstrated bigger impacts from disadvantaged students (low-socioeconomic status, racial minority student, students with disabilities, English language learners, and females) narrowing down problematic achievement gaps. Despite the time and effort from the science education community, long-term changes in science teaching and learning have not been well studied. Arguably, there is little evidence about the impact of the reforms on students' science learning over a long term, that is beyond a single year (Schroeder et al., 2007).

Argument-Based Inquiry

Argumentation is an important aspect of the language practice of science and seen as a core practice and goal for making students scientifically literate (Cavagnetto, 2010). The goal of learning science has shifted from replicating scientific terminology to engaging in the construction and communication of an in-depth scientific understanding through argumentation processes. Argumentation and analysis are essential features of science such that scientists need to be able to examine, review, and evaluate their own knowledge and ideas and critique those of others (NRC, 2012). The Framework for K-12 Science Education (NRC, 2012) highlights the critical role of argumentation by putting it at the center of scientific and engineering practices. Consequently, students and their peers must use critical thinking and argumentation skills in every step of the scientific and engineering practices. Students are required to identify, critique, analyze, and evaluate strengths and weaknesses of investigation or explanation, to develop or refine ideas, designs, and explanations. By involving students with argumentation in the science classrooms, they can construct their scientific understanding about phenomenon using evidence and general concepts of science, articulate their understanding as a process of constructing shared understanding about that phenomenon, and persuade others of these explanations by using the ideas of science to explicitly connect the evidence to knowledge claims (Berland & Reiser, 2009).

There are multiple approaches and techniques to teaching science within argument-based inquiry. Based on the review of literatures, Cavagnetto (2010) purposed threes main instructional approach in argument-based inquiry; immersion, structure, and

socioscientific approach. The Science Writing Heuristic (SWH) approach is considered as an immersion argument-based inquiry approach. It is a tool for generating activities to promote negotiation of meaning in laboratory activities through social negotiation between both teachers and students (Hand & Keys, 1999). Students in the SWH classroom are encouraged to develop arguments consisting of three components: question, claims, and evidence (Hand, 2008) providing a structure where students are required to conduct inquiry investigations by posing their own questions about the concept under review, collect data, construct claims based on evidence, find out what experts say, and reflect upon their arguments to examine how their ideas have changed.

In addition, there are number of studies exploring the impact of the SWH approach on student learning in science classrooms in various contexts, and the results have shown benefits of the approach for both teachers and students. The SWH students showed significant better gains in their science contents and attitude toward science compared to students in the traditional science classrooms. Narrowing problematic achievement gaps among disadvantaged students has reported when students engage in the SWH approach. Despite the positive impact of the approach on classroom-level measurements, the lack of evidence showing the effectiveness on students standardized achievement tests can fail to persuade teachers to shift from traditional to non-traditional in the science classroom (Kingir, Geban, & Gunel, 2012)

Purpose of the Study

Attempting to address some of the gaps found in literatures, this study aimed to gain understanding of the long-term impacts of the SWH approach on students' science achievement assessed by standardized test scores of a school district. The district was involved with a long term SWH professional development program. After completion of the professional development project, the school district decided to adopt the SWH approach as the main approach to the teaching and learning of science. The researcher examines and compares possible changes in students' science achievement trajectories between prior and after the adoption of the SWH approach.

Expanding from Omar's (2002), Gunel's (2006), and Cavagnetto's (2006) original research, the current study included data from all students in the school district starting in elementary and following them through high school. This study also gathered data across a longer period time frame including before the SWH professional development project and after the professional development ended when the support from the professional developers was removed. Given students' repeated scores across grade, more appropriate statistical analysis, a multilevel model for change, is used to track longitudinal growth of students' achievement.

Research Questions of the Study

Thus, to gain understanding about the longitudinal impacts of the SWH approach as an argument-based inquiry approach on students' science achievement, the specific research questions are posed as follows:

1. Do students have better growth in their science test scores when they experience the SWH approach in science classroom
 - a. compare to prior to the SWH exposure ?
 - b. compare to Iowa normative population?
2. Does the SWH approach have equal impact on students from different backgrounds?
 - a. Does the SWH approach have equal impact on female and male students?
 - b. Does the SWH approach have equal impact on low socioeconomic status (SES) and non-low SES students?
 - c. Does the SWH approach have equal impact on students who need Individual Educational Programs (IEP students) and non-IEP students?
 - d. Does the SWH approach have equal impact on students with different levels of reading achievement?
3. Does the grade level when students first start experiencing with the SWH approach have differential impact on student science ITBS achievement?

Overview of the Study

In this chapter, rationale for examining the impacts of the Science Writing Heuristic approach as argument-based inquiry on students' science achievement assessing by the Iowa Test of Basic Skills (ITBS). The specific research questions reflecting the purposes of study also are indicated.

Chapter Two: Literature Review, discusses the theoretical framework derived from the review of literatures corresponding to research questions. Three main inquiry-

based instructional models, their impacts on student learning, and their connection to the Next Generation Science Standards are discussed. Finally, this chapter also explores the challenges and gaps among literatures regarding the longitudinal impacts of the inquiry-based instructions on standardized tests.

Chapter Three: Design and Method, provides rationale for the use of the multilevel model to examine student achievement growth. Next, background information about the study, context of the participating school district, data structure, and descriptive statistic of participating students are described. Finally, the details of the related variables, statistical analysis techniques, and detailed procedures are also discussed.

Chapter Four: Result, presents the Science ITBS Achievement of Iowa Students and models predicting the students' science ITBS scores corresponding to the research questions. Logics and interpretation of the results of each step in the process of building final model were elaborated in great details.

Finally, Chapter 5: Discussion, presents main findings addressing three main research questions. The limitation and implications are also discusses, as well as suggestions for future research.

Index of Abbreviations

The study contains number of abbreviations normally used in educational settings which could be unfamiliar with readers outside educational fields. To accommodate readers, all important abbreviations used in this study are summarized below.

ANCOVA Analysis of Covariance

ELL	English Language Learner
ES	Effect Size
HLM	Hierarchical Linear Modeling
ICC	Intraclass Correlation Coefficient (ICC)
IEP	Individual Educational Program
ITBS	Iowa Test Basic of Skills
ITED	Iowa Test of Educational Development
MEAP	Michigan Educational Assessment Program
NGSS	Next Generation Science Standards
NRC	National Research Council
RTOP	Reform Teaching Observation Protocol
SD	Standard Deviation
SES	Socioeconomic Status
SWH	Science Writing Heuristic

CHAPTER II

LITERATURE REVIEW

The purpose of this chapter is to establish the theoretical framework derived from the review of literatures corresponding to the research questions. In order to provide a framework for answering the questions, the inquiry-based instructional models and their impacts on student learning are examined. The section begins with how inquiry is defined within science education communities and how it is translated into practice. Three main inquiry-based instructional models (BSCS 5E, modeling-based inquiry, and argument-based inquiry including the Science Writing Heuristic approach) and their impacts on students are explored. Furthermore, as we move forward to the next generation of science standards, the connections between multiple inquiry-based instructional approaches and the Next Generation Science Standards are discussed. Next, this chapter explores what research reveals regarding the impacts of inquiry-based instruction on student performance from a synthesis of major meta-analysis studies since 1980. Further, in this section, the impacts of the inquiry-based instructions contributing to equity in science classrooms is focused. Finally, this chapter also explores the challenges and gaps among literatures regarding the longitudinal impacts of the inquiry-based instructions on standardized tests.

Inquiry-Based Instructional Approach

Inquiry has become a very important concept in teaching and learning science since the idea of the learning cycle was introduced in 1960s. Inquiry is treated as the overarching goal of scientific literacy (NRC, 1996). However, the NRC does not

operationally define inquiry (Abd-El-Kalick, 2002). As explained in the National Science Education Standards (1996),

Inquiry is multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data, proposing answers, explanations, and predictions, and communicating the results. (p. 23)

The National Research Council, NRC (2000, p. 25) also provides five essential features derived from the abilities of inquiry; emphasizing question, evidence, and explanation within learning context. The five essential features of inquiry are listed below,

1. Learners are engaged by scientifically oriented questions.
2. Learners give priority to evidence, which allows them to develop and evaluated explanations that address scientifically oriented questions.
3. Learners formulate explanation from evidence to address scientifically oriented questions.
4. Learners evaluate their explanations in light of alternative explanations, particularly those reflecting scientific understanding.
5. Learners communicate and justify their proposed explanation.

These five features introduce various aspects of scientific practice which are crucial in school science as well. Practicing inquiry helps students to develop their deeper understanding of science as well as critical thinking and scientific reasoning

skills (NRC, 2000). Teachers are required to provide multi-opportunities for students to engage in investigation that incorporate these five essential features. However, in the teaching practice, these five features of inquiry are not equally well received (Asay & Orgill, 2010).

Asay and Orgill (2010) conducted an analysis of the Essential Features of Inquiry found in articles published in *The Science Teachers* journal from 1998-2007. The analysis demonstrated that within inquiry-based classroom, students often gathered evidence and participate in teacher-guided analysis of that evidence. However, students had few opportunities to generate scientifically oriented questions, create evidence-based explanations, connect explanations to accepted scientific concepts, or justify the results of their investigation to a larger group of peers. One possible explanation could be that teachers view inquiry as a process that student should participate rather than a vehicle for learning science (Asay & Orgill, 2010).

However, the National Research Council (1996) refers to scientific inquiry as “the diverse ways in which scientists study the natural world and propose explanation based on the evidence derived from their work” (p. 23), but NRC (1996) does not prescribe a single approach to teaching science that would guarantee success in student learning. Thus, various inquiry-based instructional models have been introduced within the science education community to help teachers translate theory into practice. Each approach provides instructional guidelines to help teachers implement inquiry-based science classrooms. In the following section, three major inquiry-based instructional models in science education; BSCS 5E model, model-based inquiry, and argument-based inquiry approach, are discussed in detail.

The BSCS 5E Instruction Model

First, the BSCS 5E Instruction Model, also known as 5Es, started in mid-1980 adapted from the Science Curriculum Improvement Study (SCIS) learning model (Karplus & Their, 1976). Since then the BSCS 5E Instruction Model has been intensively and widely used in the development of new curriculum materials and professional development programs. According to Bybee, Taylor, Gardner, Van Scotter, Carlson Powell, Westbrook, and Landes (2006), the 5Es model views learning as dynamic and interactive. Students interpret a phenomenon and internalize the interpretations about that phenomenon that makes sense to them. Then students continuously redefine, reorganize, expand, and change their initial concepts through interaction with their environment, other students, or both. To change or to improve their conceptions about any phenomenon, learners need to be challenged and showed that their current conceptions are incomplete or inadequate. Moreover, opportunity, time, and experience are required for learners to further develop a better conception. With 5Es instructional guidelines, learners' construction of knowledge can be assisted by using sequences of lesson designed to challenge current conceptions and provide time and opportunities for reconstruction to occur.

The 5E Instruction Model consists of 5 phases; engagement, exploration, explanation, elaboration, and evaluation. This model is specifically designed to give opportunities for students to challenge their current ideas and construct better understanding about a phenomenon. The description of each phrases of the BSCS 5E Instructional Model is summarized and provide in Table 2.1.

Table 2.1 Summary of Five Phases of the BSCS 5E Instructional Model

Phase	Summary
Engagement	The teacher or a curriculum task accesses the learners' prior knowledge and helps them become engaged in a new concept through the use of short activities that promote curiosity and elicit prior knowledge. The activity should make connections between past and present learning experiences, expose prior conceptions, and organize students' thinking toward the learning outcomes of current activities.
Exploration	Exploration experiences provide students with a common base of activities within which current concepts (i.e., misconceptions), processes, and skills are identified and conceptual change is facilitated. Learners may complete lab activities that help them use prior knowledge to generate new ideas, explore questions and possibilities, and design and conduct a preliminary investigation.
Explanation	The explanation phase focuses students' attention on a particular aspect of their engagement and exploration experiences and provides opportunities to demonstrate their conceptual understanding, process skills, or behaviors. This phase also provides opportunities for teachers to directly introduce a concept, process, or skill. Learners explain their understanding of the concept. An explanation from the teacher or the curriculum may guide them toward a deeper understanding, which is a critical part of this phase.
Elaboration	Teachers challenge and extend students' conceptual understanding and skills. Through new experiences, the students develop deeper and broader understanding, more information, and adequate skills. Students apply their understanding of the concept by conducting additional activities.
Evaluation	The evaluation phase encourages students to assess their understanding and abilities and provides opportunities for teachers to evaluate student progress toward achieving the educational objectives.

Bybee et al., (2006) conducted an extensive review of the effectiveness of the learning cycle; the SCIS Learning Cycle and the BSCS 5E Instruction Model, from various research varying in subject matters and grades. The composited results demonstrate the positive effect of the learning cycle on students' gain score in understanding science concepts, comparing to students taught using traditional approaches. The 5Es model not only helps students developing understanding of scientific concepts, but also consistently shows better results for helping students developing scientific inquiries abilities and general reasoning skills. In addition, students who are taught by using learning cycle approach consistently showed positive attitude toward science. Wilson, Taylor, Kowalski and Carlson (2010) argued that the 5Es model is effective than traditional instruction because it builds on students' prior knowledge, emphasizes deep understanding, recognizes the importance of metacognition, and encourage social aspect of learning. Besides helping students, Bybee et al., (2006) also found positive impacts of the 5Es learning cycle model on teachers as well. Teachers reported gains in their subject matter conceptual understanding, reason skills, and attitude toward science after implementing the approach.

In summary, the learning cycle is most effective when (Bybee et al., 2006);

- All three phases of the model must be included in instruction, and the exploration phase must precede the term introduction phase.
- The specific instructional format may be less important than including all phases of the model, but laboratory work (typical in the exploration phase) is more effective for many students, provided it is followed by discussion (term introduction).

- Finally, student attitudes toward science instruction are more positive when they are allowed to explore concepts through experimentation or other activities before discussing them (p. 27).

Although discussion and investigation already listed as crucial parts of the 5Es model, the findings start to highlight the critical functions of discussion and argumentation leading to success in the 5E instructional model. A more recent study by Wilson et al. (2010) explored the impacts of the 5E models on middle school students' argumentation. The results showed students who had experienced with the inquiry-based 5E models had statistically significant higher scores in claims, evidence, and reasoning than students in the control group.

Modeling-Based Inquiry

The second instructional model is the model-based inquiry. Scientists develop models and representation as ways to generate understanding about the natural world, as well as to make their thinking visible (NRC, 2000). Scientists create models in the forms of analogies, conceptual drawings, diagrams, graphs, physical constructions, or computer simulation in order to describe and understand the organization of systems. The National Science Education Standard (NSES) highlights the importance of models in the science classroom by explicitly specifying that, for secondary students, investigation “should culminate in the formulating an explanation or model” (NRC, 1996, p.175). The interconnection between scientific explanation and model is prominent. The model-based inquiry approach views model as key roles in the explanation of science (Gilbert, Boulter & Rutherford, 1998). Models can provide the basis for all five types of explanation

(intentional, descriptive, interpretive, causative, and predictive). They also can be appropriately adapted their complexity to different audiences e.g., model of white light for younger students is not complicated as for older students. Finally, developing models can support the production of explanation. In addition, modeling also provides crucial elements that help generate hypotheses for testing, act as referent in interpreting observations, and are themselves targets of revision (Windschitl, Thompson & Braaten, 2008).

Schwarz, Reiser, Davis, Kenyon, Achér, Fortus, and Krajcik (2009) suggested two critical dimensions of scientific modeling. The first dimension concerns the generative nature of models as a tool for explaining and predicting about scientific phenomena. Second, scientific models are dynamic. Models can be revised or changed reflecting changes in individual's understanding of a concept. However, the term models are not well defined. In order to clarify what consider as models, Harrison and Treagust (2000) purposed 10 categories of model used in scientific community and science education. The first group of models can be called analogical models which includes 1) scale model, 2) pedagogical analogical models, 3) iconic and symbolic models, 4) mathematical models, 5) theoretical models, 6) maps, diagrams and table, 7) concept-process models, and 8) simulations. These models could be a simplified or exaggerated representation of an object or process ranging from concrete to abstract models. The other groups are 9) mental models and 10) synthetic models.

In order to help teachers to successfully implement the modeling-based approach, general guideline for instruction framework in the modeling-based inquiry were introduced by various researchers (Schwarz et al., 2009, Windschitl et al., 2008).

According to Schwarz et al. (2009), teachers need to guide students through various process including; engaging with questions or problems often through materials or initial models, developing hypotheses about causal or relationship in the phenomenon; making systematic observations to test these hypotheses; creating models of the phenomena that would account for the observation; evaluating this model against standards of usefulness, predictive power, or explanatory adequacy; and finally, revising the model and applying it in new situations. In addition, Windschitl et al. (2008) also purposed a tentative guideline for modeling-based inquiry instruction similar to the guideline mentioned above but place more emphasis in using modeling to help generating questions for investigation and encouraging to generate their final arguments to support or refute claims about the explanatory process in the original model.

Various studies found that using modeling-based inquiry could help students understand scientific concepts better (Windschitl et al., 2008). Students also started to recognize the explanatory power of models by using models to explain a phenomena more, instead of using model as an illustrative tool. However, there are some critical aspects of successful modeling-based instruction. Asking students to build a model is not effective enough. First, students need an authentic reason for building a model other than a required task (Schwarz et al., 2009). Second, students revealed strongest ownership when they build their own models (Harrison & Treagust, 2000). Lastly, there needs to be active negotiation about models among teachers and students (Harrison & Treagust, 2000).

Argument-Based Inquiry

The third instructional model is the argument-based approach. Since 1990s, the goal of science education had shifted from remembering scientific terminology to be more focus on the ability to construct an understanding and communicate that understanding to boarder audiences through language. The needed shift had highlighted the importance of language in science classroom.

As Hand (2008) argued, “language is a critical to the construction of science knowledge, the debate and argument of science, and the dissemination of science knowledge” (p. 1). Argument-based inquiry focuses the importance of the use of language in science through argumentation. Driver, Newton, and Osborne (2000) underscored that “arguing is a human practice that is situated in specific social settings” (p. 290). The concept of argumentation is building on Vygostky’s (1978) learning theory and the notion that science is not possible without language including text, modes of representation, and talk (Lemke, 2004). Language is a fundamental aspect of science as inquiry as it drives the epistemic nature of science and captures the culture of science; both inform interpretation and knowledge construction (Ford, 2008). Moreover, language is also viewed as the primary tool for communication in science and making thinking visible (NRC, 2000). In this sense, students reflect on and develop their own scientific understanding by themselves or others through the practice of language. Teachers also use language as ways to assess and understand how students thinking.

Building on this position, teachers should create opportunities for students to engage in talking and challenging each other’s claims supported by evidence in small groups or in a whole class. Argument is an important aspect of the language in the

practice of science (Cavagnetto, 2010) and a fundamental tradition of science communities (Hand, 2008). According to Berland and Reise (2009), by involving students with argumentation in the science classroom, they are able to construct their scientific understanding about phenomenon using evidence and general concepts of science. Moreover, students are able to articulate their understanding as a process of constructing shared understanding about that phenomenon. Lastly, students can persuade others of these explanations by using the ideas of science to explicitly connect the evidence to knowledge claims. As results, first, students can learn in more meaningful ways. Moreover, student's communication skills, reasoning skills, and scientific literacy are developed. Finally, students' understandings of scientific culture and practice are enhanced (Jimenez-Aleixandre & Erduran, 2007).

There are multiple approaches and techniques in teaching science within argument-based inquiry. Cavagnetto (2010) had done an extensive review of literatures of argument-based inquiry where the focus is on generation and evaluation of scientific evidence and explanations. According to the results, there were three major argument intervention approaches used within science: an immersion for learning scientific argument (immersion approach), teaching the structure of argument (structure approach), and emphasizing the interaction of science and society (Socioscientific approach). Cavagnetto (2010) also critiqued the advantages and disadvantage of those three approaches arguing that the structure approach focuses on the communication of defense of knowledge claims but pays little attention to other elements of knowledge construction in science. Furthermore, the socioscientific approach focuses on exploring the social elements that influence science and vice versa. Although the community of science and

society are totally interrelated, science knowledge construction and socioscientific issues are quite different. Finally, the immersion approach is based on the concept that students need to be actively engaged in the process of argumentation as a means to learn about science argument. That is, argument was not considered as separated part from inquiry but was found throughout the inquiry processes as students generated questions, designed investigation, collected data, generated claims, reflected on their ideas. Consequently, argument can be used as a tool to develop understanding in both science principles and cultural practices of science.

The Science Writing Heuristic (SWH)

The Science Writing Heuristic (SWH) approach is considered as an immersion argument-based inquiry approach. The approach was originally introduced into classrooms as a writing-to-learn strategy highlighting the uses of non-traditional writing tasks that extend students' needs to engage with the demands of science (Prain & Hand, 1996). The SWH approach is conceptualized as a tool for generating activities to promote understanding of laboratory activities through negotiation involving both teacher and student (Hand & Keys, 1999). Table 2.2 provides an overview of the tentative guideline templates for teachers and student in the SWH classroom. It is a guideline designed to promote the negotiation of meaning in laboratory activities as well as a metacognitive support to prompt student reasoning about data.

Table 2.2 The SWH templates for Teacher and Student

<u>The Science Writing Heuristic, Part I</u>	<u>The Science Writing Heuristic, Part II</u>
A template for teacher-designed activities to promote laboratory understanding.	A template for student.
1. Exploration of pre-instruction understanding through individual or group concept mapping or working through a computer simulation.	1. Beginning ideas - What are my questions?
2. Pre-laboratory activities, including informal writing, making observations, brainstorming, and posing questions.	2. Tests - What did I do?
3. Participation in laboratory activity.	3. Observations - What did I see?
4. Negotiation phase I - writing personal meanings for laboratory activity. (For example, writing journals.)	4. Claims - What can I claim?
5. Negotiation phase II - sharing and comparing data interpretations in small groups. (For example, making a graph based on data contributed by all students in the class.)	5. Evidence - How do I know? Why am I making these claims?
6. Negotiation phase III - comparing science ideas to textbooks for other printed resources. (For example, writing group notes in response to focus questions.)	6. Reading - How do my ideas compare with other ideas?
7. Negotiation phase IV - individual reflection and writing. (For example, creating a presentation such as a poster or report for a larger audience.)	7. Reflection - How have my ideas changed?
8. Exploration of post-instruction understanding through concept mapping, group discussion, or writing a clear explanation.	

Over time, the SWH approach is evolving through research and is now focusing on meaning-negotiation and argumentation in science classroom. Building on the writing-to-learn framework (Prain & Hand, 1996), the SWH approach is now focusing on embedding science argument within the context of doing inquiry (Hand, 2008). The SWH approach promotes the implementation of authentic, meaning-seeking opportunities for students to construct their understanding around an argument (Hand, 2008). Negotiation of meaning is provided and encouraged across format for discussion and writing within science topics though out the SWH classrooms.

The SWH approach is based on a simple argument structure; questions, claims, and evidence. Students in the SWH classroom are required to conduct inquiry investigations by posing their own questions about the concept under review, collect data, construct claims based on evidence, find out what experts say, and reflect upon the their arguments to examine how their ideas have changed. Moreover, students are constantly required to negotiate meaning as individuals, in small groups and at the whole-class where importance is placed on public and private construction and critique of knowledge. The impacts of the SWH approach on students emerging from studies are discussed in more details later in this section.

Similarities among Inquiry-Based Instructional Models

According to descriptions and elements of these inquiry-based instructional approaches, the researcher argued that BSCS 5Es model, argument-based, and modeling-based inquiry have the similar core concepts. Although each approach has different focuses, they still show high degree of similarities. The 5Es model could be considered as

a general approach of inquiry-based instruction where students were encouraged to use their prior knowledge to develop scientifically sound questions; conduct an investigation to gather data; construct evidence based claim and explanation; share, communicate, convince their claims/explanations to teachers and peers; compare their existing claims/explanations to the current acceptable scientific knowledge; reflect what they have learned by making revision or changes of their claims/explanation. Students in modeling-based approach have to go through the same process but students are encouraged to use models as critical tools to generate question and construct claims and explanation in terms of their model. The same scientific inquiry processes mentioned above are crucial to the argument-based science classroom as well. However, the argument-based classroom place more emphasis on meaning negotiation throughout the processes. Moreover, teachers in argument-based inquiry are also encourage to incorporate models and representations to explain a phenomena.

Recently, NRC (2012) revised its direction of science standards in K-12 science education and provided a new framework for scientific activities based on over a decade of research. Consistent with the idea of the interconnection among major inquiry-based instructional models, the framework emphasizes the need for argumentation and modelling in the scientific inquiry process highlighting the overlapping core concepts among inquiry-based approaches and set up new expectations of activities in science classroom.

Connecting to the Next Generation of Science Standard

According to NRC (2012), the new standards were developed to ensure that All students have some appreciation of the beauty and wonder of science; possess sufficient knowledge of science and engineering to engage in public discussions on related issues; are careful consumers of scientific and technological information related to their everyday lives; are able to continue to learn about science outside school; and have the skills to enter careers of their choice, including (but not limited to) careers in science, engineering, and technology (p.1).

The new standards recognized that in the past, scientific concepts were presented as discrete facts with a focus on breadth over depth, and scientific practices were usually reduced into a single set of procedures. Moreover, other skills in scientific practices such as modeling, developing explanation, critique and evaluation (argumentation), and communication, were usually ignored or used without content. These skills wrongly became goals of classroom instruction. As a result, teachers purposefully taught students these skills, rather than used them as means to develop deeper understanding of scientific concepts (NRC, 2012).

The new science standards were developed to address and overcome these weaknesses by highlighting the importance of scientific and engineering practices, crosscutting concepts, and disciplinary core ideas (content). To reach these goals, the new standards focused on smaller but deeper set of scientific core ideas and the application of their content. The new science standards also aligned with cognitive demands with the English Language Arts and Mathematics Common Core State Standards. Most

importantly, engineering practices became as important as scientific practices in the new version of science standards. Teachers are encouraged to integrate engineering practice into science classroom in context.

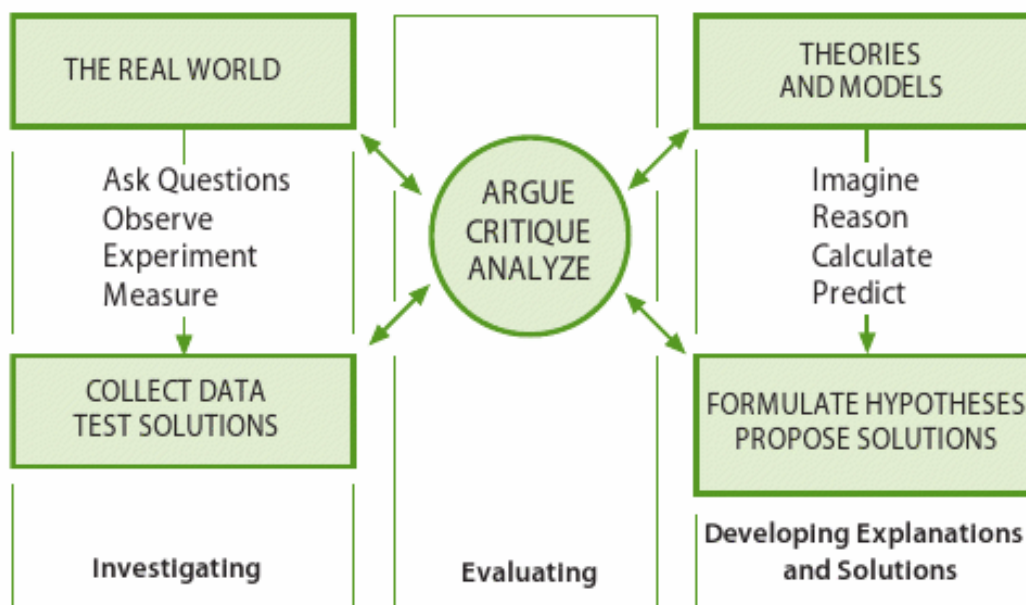


Figure 2.1 The Three Spheres of Activity for Scientists and Engineers (NRC, 2012)

In order to achieve these goals, NRC (2012) suggested a new framework called three spheres of activity for scientists and engineers to help develop understanding the practices of scientists and engineers. The framework (Figure 2.1) consisted of three main spheres; investigation, evaluation, and development of explanations and solutions. The left sphere; investigation, are activities related to scientific investigation procedures such as developing questions, planning and conducting experiment, and collecting data. The

right sphere; development explanations and solutions, are activities corresponding to developing models or establishing theories drawn from evidence. These activities often expand existing theories or generate new questions. Finally, the middle sphere; evaluation; are activities related to argumentation and critique. According to NRC (2012), the middle sphere constantly interacts with both left and right spheres, that is, students and their peers must use critical thinking and argumentation skills to identify strengths and weaknesses to develop or refine ideas, designs, and explanations. The connection among these spheres within this new framework highlights the importance and relationship of argumentation and the uses of models in the science classroom.

Effectiveness of Inquiry-Based Instruction

A large number of studies have been conducted in science education community to examine the effectiveness of inquiry-based science instructional approaches. Since many science classroom teaching techniques have been introduced in science education, the meta-analysis technique is very useful to compare the results across studies. Since 1980, there have been various meta-analysis studies published in science education synthesizing and comparing the effects across various inquiry-based science teaching strategies on student achievement. Mixed impacts of inquiry-based instruction in science classrooms have been reported. Both neutral and negative impacts have been found across different inquiry-based instructional strategies and contexts; however, the majority of studies reported positive impacts of inquiry on students.

First, Bredderman (1983) conducted a meta-analysis exploring the effectiveness of three major activity-based elementary science programs; Elementary Science Study

(ESS), Science-A Process Approach (SAPA), and the Science Curriculum Improvement Study (SCIS) which were widely adopted and researched at time compared to traditional classrooms. The data consisted of 57 studies published from 1967 to 1980 covering approximately 13,000 students from over 900 classrooms in various grad levels. The studies measured the effectiveness of the activity-based elementary science programs in nine different areas; science content, science process, creativity, language, mathematics, perception, affective, intelligence, and logical development. The results indicated that the effects of the activity-based program were generally positive on all outcomes with weighted mean effect size of 0.35. According to Bredderman (1983), the highest improvement could be expected from tests of science process, creativity, and intelligence with an effect size of 0.4 - 0.6. However, a small average effect size was observed in affective, perception, logical development, and language. While the impact of the approaches on the test in science content and mathematics area yielded very small effect with average effect size smaller than 0.1. Moreover, the impacts assessed by nationally standardized tests tended to reported lower or null effect size. The results also indicated that there was no influence of duration of treatments on effect sizes across studies. Bredderman (1983) suggested that the advantages developed initially from the activity-based experience were maintained but not accumulated over time; however, the advantage might be disappeared when the treatment was removed. This observation was made where a group of elementary students performed better in activity-based classroom but the advantage was lost when they were moved to a traditional classroom in middle school.

Second, Shymansky, Kyle, and Alport (1983) conducted a similar study during the same period of time of Bredderman's study (1983); however, Shymansky, Kyle, and Alport (1983) expanded the analysis to a wider range of contexts. There were 105 experimental studies included in this meta-analysis covering 45,000 students in 18 different student performance measures. The results showed similar trends as found in Bredderman's (1983) study; however, the average reported effect sizes of achievement tests were generally higher than those were reported in Bredderman's (1983). On average students who were exposed to a new science curriculum performed 63% better in performance measures compared to students in traditional science classrooms (Shymansky et al., 1983). The greatest gains were found in the areas of processing skill development and attitude toward science.

Since Bredderman (1983) and Shymasky et al. (1983) reported the synthesis of the effectiveness of new curriculum and activity-based science programs, there have been more studies about new and innovative science curriculums and approaches. Expanding from studies of Bredderman (1983) and Shymasky et al (1983), Minner, Levy, and Century (2010) conducted a meta-analysis to examine the impact of inquiry science instruction on K-12 student outcome with the timeframe of 1984 to 2002. Since the growing number of studies about inquiry-based instruction created confusion about the concept of inquiry within the science education community, three aspects of inquiry-based science instruction were purposed as a criterion framework to select literatures in this study. According to Minner et al. (2010), inquiry science instruction must have these following three aspects; "(1) the presence of science content, (2) student engagement with science content, and (3) student responsibility for learning, student active thinking,

or student motivation within at least one component of instruction” (p. 476). There were 138 studies included in this synthesis. The results indicated a clear and consistent trend showing the positive association of the inquiry-based instruction and the improvement of student content learning, although some studies showed no or negative impact. Moreover, hands-on experience and activities emphasizing on student active thinking or responsibility also affected student science conceptual learning; however, hands-on activities alone were not effective enough for conceptual change. Students need to be provided opportunities to engage with process of meaning developing through class discussion (Minner et al., 2010).

A more recent meta-analysis conducted by Furtak, Seidel, Iverson, and Briggs (2012) examining the effectiveness of inquiry-based teaching approach, particularly experimental and quasi-experimental studies, on student learning focusing on the cognitive domains of inquiry activity and degree of guidance given to students. There were 37 experimental and quasi-experimental studies published from 1996 to 2006 included in this study. Overall, the results revealed positive impacts of inquiry-based teaching reforms on student learning of science with average effect size of 0.50. Moreover, studies focusing on epistemic domain, procedural, and social domains showed the largest effect size compared to studies focusing on conceptual domain of inquiry. Furtak et al. (2012) argued that it was crucial for students to engage in generating, developing, and justifying explanations in science activities in order to effectively learn science. When comparing between teacher-led classroom instruction and student-led classroom instruction, teacher-led classroom condition where teachers actively guided student activities was more effective in helping student learn science with averaged effect

size of 0.65, while averaged effect size of student-led classroom conditions was only 0.25. The results confirmed the important role of the teacher the context of inquiry learning as also found in previous meta-analysis by Schroeder, Scott, Tolson, Huang, and Lee (2007).

Schroeder et al., (2007) explored the impacts of teacher's strategies in inquiry-based instruction on student learning across 61 experimental or quasi-experimental studies. The results indicated significant different effects of strategies adopted by teachers in inquiry-based classrooms. The classrooms where teachers focused on connecting learning to student's prior knowledge and real-world examples or providing opportunities for students to work with physical objects showed largest effect sizes. If students engage with classroom where their interests and prior knowledge are actively connected to instruction and have opportunities to experience collaborative scientific inquiry under the guidance of an effective teacher, achievement will be enhanced (Schroeder et al., 2007).

According to the results from multiple published meta-analyses since 1983 to 2012, there was strong evidence indicates the positive impacts of an inquiry-based instruction on student learning and other related skills. However, the recent meta-analysis started to reveal the differential effects studies. Teacher's strategies, types of studies, and types of assessment contributed to the variations of the impacts (Schroeder et al., 2007). One of critical results emerging from research was that the inquiry-based instruction did not equally impact all students even if they shared the same classroom experience (Secker, 2002). Disadvantage students who usually were at risk of having low achievement, had greater benefits from an inquiry-based classroom environment contributing to inequity in science classroom.

Equity in Science Classroom

Although the idea of science for all (AAAA, 1993) has long been put forward, achievement gaps by gender, race, and socioeconomic status (SES) still remain a critical issue in science education (Wilson et al., 2010). Students from low SES families, minority students, females are more at risk of low science achievement (Von Secker & Lissitz, 1999). However, the concept of inequality cannot be conceptualized on the basis of broad demographic characteristics. After all, individual students are not at risk of having low science achievement just because they are females, poor, or minority. Rather, individuals with low SES, minorities, and females are members of highly variable populations with heightened probabilities of an undesirable achievement outcome.

Secker and Lissitz (1999) gathered data from the 1990 High School Effectiveness Study (HSEs) and conducted HLM models to investigate the achievement gaps among students from different backgrounds. Secker and Lissitz (1999) demonstrated significant achievement gaps between disadvantage and non-disadvantage students. They found that an average science achievement of males was 0.288 standard deviations higher than that of females, average science achievement of minorities was 0.578 standard deviations lower than that of majority students, and average science achievement of students in low-SES families were significant lower than non-low SES students.

Further, obtaining data from the first follow-up of the National Education Longitudinal Study (NELS), Secker (2002) sought to examine the effect of inquiry-based instructional practice on academic excellence and equity. The data consisted of 4,377 tenth-grade students in 1,406 classes across 50 states. The results demonstrated the existing achievement gaps in classrooms; girls had 0.48 standard deviations lower science

achievement than boys and minority students also had 0.51 standard deviations lower achievement than majority students. For every quartile increase of SES, science achievement was predicted to be 0.30 standard deviations higher. After teachers implemented inquiry-based practices, Secker (2002) found that inquiry-based classroom implementation was associated with higher science achievement overall. However, the instructional techniques that teachers used do not affect all students equally. Minority and lower SES students had more benefits from classrooms emphasizing appropriate laboratory techniques or problem solving skills. Consequently, achievement gaps among students with different demographic profiles were reduced contributing to greater inequities in science classroom.

In addition, Lee, Buxton, Lewis, and LeRoy (2006) studied the impact hands-on based science instruction on elementary school students from different backgrounds focusing on English Language Learner (ELL) and low-SES students. In this study, teachers adopted the teacher-explicit to student-initiated continuum approach in which teachers provide extensive assistance at an early stage and gradually reduce assistance as students learn to take initiative and conduct inquiry on their own. The results demonstrated the enhanced abilities with the inquiry tasks, e.g. ability in controlling variables, and generating evidence to support their theories, of all students; however, the impacts were more prominent for students from disadvantaged backgrounds. Lee et al., (2006) suggested that it is critical for teacher to provide explicit instruction and scaffolding for the disadvantage students or those with limited science experience to develop inquiry abilities.

Emerging research has reported consistent trends with results showing students from disadvantage groups demonstrated greater gains in their learning from inquiry-based instruction. However, a study by Wilson, Taylor, Kowalski, and Carlson (2010) demonstrated different results. With a smaller sample size of fifty-eight students age between 14 to 16, the results showed significant greater impacts of the 5Es instructional model on all students compared to traditional classroom. And the effectiveness of inquiry was consistent across races, genders, and socioeconomic statuses. Consequently, while persistent achievement gaps in traditional classroom could widen over time, the inquiry-based instruction could prohibit the expansion of the existing gaps.

Beyond the Effectiveness of Inquiry

Beside studies exploring the effectiveness of inquiry-based instruction on student's learning, other significant findings have emerged from research. Lee et al., (2006) started to explore a cumulative impact of the inquiry-based instruction and found that when students participated in the science inquiry classroom over multiple years, the students experienced cumulative gains contradicting to Bredderman (1983)'s study which suggested that impact could be maintained but not accumulated over time. Moreover, Johnson, Fargo, and Kahle (2010) explored the impacts of a systemic reform efforts by conducting a longitudinal study following students learning performance and teachers changes for five years. The teachers and students were participants in a systemic reform professional program for three years. The results indicated continuous improvement of teachers during the professional development project and after the project was ended. The analysis of students' state assessment also demonstrated consistent higher scores of

participating students compare to the control group. The results demonstrated the sustainability of gains from a reform program of both teachers' changes and students' performance even after the program had ended (Johnson et al., 2010). However, studies focusing on systematic reform, long-term impacts of inquiry, or impacts of inquiry on standardized measurement are lacking from literature (Johnson et al., 2010; Lee et al., 2006; Schroeder et al, 2007).

Impacts of Inquiry on the Standardized Test Achievement

Standardized testing is being increasingly emphasized within school systems in the United States (Aydeniz & Sutherland, 2012). Scholars have split views toward the use of standardized testing in education reform. First, Greene and Winter (2003) argued for employing the standardized tests in educational reform because of their high validity and reliability. Moreover, standardized testing is also an effective system to ensure the minimum competencies of all students. While other scholars argued against the idea by stating that the results from one single standardized test are not enough to make high-stakes decisions (Aydeniz, 2007; Brickhouse, 2006, Madden, 2008). Aydeniz (2007) and Madden (2008) raised concerns about the emphasis on increasing standardized test scores could negatively impact the reform efforts which encourage science teachers to use assessments closely align with their instructional goals. Some teacher were forced to change their instruction to accommodate standardized testing though it contradicted to their personal beliefs about good teaching (Aydeniz & Sutherland, 2012).

While standardized tests are considered to be least sensitive of the inquiry approach, it is one of the most high-stake assessments used to evaluate students learning

by teachers, administrators, and policy-makers (Ruiz-Primo, Shavelson, Halmilton, & Klein, 2002). Most science reform initiatives have used pre/post assessment or teacher-generated assessment to evaluate the impact of implementation; however, these results showed little evidence of the impact on student performance on mandatory tests by states (Johnson et al., 2010). According to a meta-analysis from Schroeder et al. (2007), there were 15 (24.2%) studies examining the impact of non-traditional science classroom on students' achievement using some types of standardized tests during 1980 to 2004. Geier et al. (2008) argued that *“the lack of students-level distal standardized test data to demonstrate achievement gains from standards-based inquiry science curricula remains a weakness in the literature”* (p. 924).

Based on Snow's (1974) ideas of referent generality, Ruiz-Primo, Shavelson, Halmilton, and Klein (2002) developed the notion of multiplelevel-multifaceted assessment as a framework to provide a better picture of the extent of the effect that science instruction is having. Multilevel assessment refers to assessments that fall into the continuum of immediate and the distal of the enacted curriculum. Based on their instructional sensitivity, five types of achievement indicators are purposed as 1) immediate (science journals, classroom tests), 2) close (assessment from a slightly more advanced activities in the units, 3) proximal (new assessment with same concept, 4) distal (large-scale assessment from State/National curriculum framework, and 5) remote (standardized national science achievement tests). If science reform has an impact on students' achievement, these results could be found unevenly across different levels of assessments. Further, the multifaceted assessment refers to the notion that the assessments could measure different facets of knowledge; declarative, procedural, and

strategic knowledge. Different kinds of tests are need to assess different facets of knowledge. Ideally, a comprehensive multilevel-multifaceted assessment should contain immediate, proximal, and distal measures of declarative, procedural, and strategic knowledge.

Ruiz-Primo et al. (2002) argued that evaluating students' science achievements at different distance from science classroom curriculum would provide a better picture of the impacts of the educational reform than using only close or distal measures. In addition, the information on initial student status, post-instruction student performance, and reform implementation should be collected longitudinally to better estimate and interpret reform effects. Based on their pilot and the main study (Ruiz-Primo et al., 2002), the impacts of the inquiry-based instruction on students' performance and the magnitude of the impact was found biggest at immediate achievement indicator and decreased as the distance of the assessment from the curriculum implemented increases. Moreover, socioeconomic status and general ability have greater predictive power on achievement measures that are far from classroom instruction than immediate and close levels.

Further, Tretter and Jones (2003) conducted a study examining the impact of inquiry based science laboratory on students' achievement. Even though, the multilevel-multifaceted assessment framework was not used as a referential framework in this study, the core ideas were consistent with the multilevel-multifaceted assessment framework. The researchers collected data from three sources; classroom attendance, classroom grades, and physical science standardized test scores from 1,300 students in four years. The results were compared between treatment and control group. The results were consistent with the notion of multiplelevel-multifaceted assessment (Ruiz-Primo et al.,

2002). The results demonstrated higher students' engagement in inquiry classroom measured by classroom participation. Treatment-group students were also more likely to show up to take the standardized tests at the end of the course and had significantly higher classroom grades compared to control groups; however, there was no significant differences in their standardized test score. Tretter and Jones (2003) concluded that classroom assessments were more sensitive to classroom instruction than standardized assessments.

Another longitudinal study examined the effects of a project-based science curriculum called Center for Learning Technologies in Urban Schools (LeTUS) on students' Michigan Educational Assessment Program (MEAP) scores (Geier, Blumenfeld, Marx, & Krajcik, 2008). During three years of the project, there were 37 participating teachers from 18 schools involving with approximately 5,000 students. Though, this was a longitudinal study, the complexity of data did not allowed individual achievement to be tracked over years. The researchers decided to divide participating students into two cohorts and compared their science MEAP scores to other students in the Detroit Public Schools (DPS). Teachers who participated in Cohort I received intensive support from researchers and professional developers, while participating teachers in Cohort II received less support in classrooms as being an upscale phase of the project.

The results suggested that Cohort-I students who completed at least one LeTUS unit, during 7th or 8th grade, significantly outperformed their peers on their overall MEAP science scores and all sub scores. LeTUS students in Cohort II also significantly outperformed their peers in both science contents and science process skills measured by

MEAP; however, the effect sizes were smaller compared to Cohort I. Further analysis indicated that male students had relatively higher gains than females, consequently reducing the gender gap in science achievement scores.

While, a number of studies have reported on the impacts of inquiry-based instruction on students standardized tests at school or state level (e.g., Geier et al., 2008; Ruiz-Primo et al., 2002; Tretter & Jones, 2003). Geier et al. (2008) and Lee et al. (2006) agreed that the limited numbers of studies investigating how inquiry-based instruction influenced students' standardized achievement still remains a weakness in the literature. Moreover, the reported impacts of reform teaching could be contaminated by factors other than school performance such as, family, demographic, community influence (Lee et al., 2006). In order to effectively examine the impact of reform teaching, the impacts of student-level data are needed. However, these student-level data can be difficult to obtain.

The Effectiveness of the Science Writing Heuristic (SWH)

Approach

There are a number of studies exploring the impact of the SWH approach on student learning in science classrooms in various contexts, and the results often showed benefits of the approach for both teacher implementation and student achievement (Hand & Key, 1999) and critical thinking skills (Chanlen & Hand, 2011). Teachers recognized and appreciated the benefits from activities in the SWH approach (Wallance & Kang, 2004). Generally, inquiry-bases teaching is difficult to implement in the current school culture where teachers had strong beliefs that they had to present canonical concepts and

explanations in an efficient manners. Teachers in the traditional classrooms also heavily relied on lecture format in delivering the scientific content (Villanueva, Taylor, Therrien, & Hand, 2012). Moreover, teachers also face challenges to make appropriate modifications based upon the needs of the students; in part because of the lack of available instructional methodologies and resources in most classrooms.

In general, teachers believe that inquiry-based instruction could help students develop conceptual understanding and foster thinking and problem solving skills; however, students can be too immature and lazy prohibiting success in inquiry-based instruction (Wallance & Kang, 2004). After adopting the SWH approach, teachers had better perception of the inquiry, particularly the SWH approach (Wallance & Kang, 2004). At the early stage of implementing the SWH approach, teachers viewed writing in science as primarily an excellent assessment; however, teacher's views about writing task gradually changed. Perception of the writing tasks in science classroom was moved away from just being a tool for recording of past learning to a resource for facilitating conceptual development. Teachers also moved from heavily relying on scientific technical terms to more everyday science terms indicating deeper understandings about the concept (Hand, Prain, & Yore, 2001).

Further, Gunel (2006) studied the impact of teachers' implementation of the SWH approach on student standardized test scores over three-year period. The data was collected from six teachers (three high school teachers and three middle school teachers) who were participating in the SWH professional development project during 2003-2005 academic year. In this study, three criteria (dialogical interaction, controlling of knowledge, and unit preparation) were used to identify teachers' level of implementation.

Based on the criteria, teachers were classified either as low, medium and high implementer. The students' science Iowa Test of Educational Development (ITED) were collected for four years (2001-2005) where 2001-2002 science scores were used as a baseline prior to the project. In general, Gunel (2006) found that even though the SWH professional development positively affected and helped the participating teachers to gradually move from teacher-centered approach to student-center approach, the individual rate of progression was complex, not linear, and different. Some teachers tried to implement the approach in the first year ranking as medium implementer. However, he or she might feel uncomfortable in the second year and moved back to traditional classroom approach. Some teachers struggled with the approach in first two years; however, he or she could be able to move rapidly in the third year of implementation. In sum, two high school teachers shifted from medium to high implementers where another high school teacher never changed his implementation. The similar results appeared with middle school teachers where two teachers shifted from using traditional approach to being high/medium implementers. However, the remaining middle school teacher had not changed.

An analysis of students' science Iowa Test of Educational Development (ITED) scores were employed using the Analysis of Covariance (ANCOVA). Though not all cases were statistically significant, the general findings suggested that as teachers' implementation levels improved, their students' achievement on science ITED scores increased with mostly a small effect size. Large and medium effect sizes could be observed in two successful middle school teachers. Moreover, there were three out of

four incidents where the negative effects on students ITED were observed when participating teachers who regressed back to a low level implementation.

In addition, Cavagnetto (2006) conducted similar study to Gunel (2006) monitoring the progress of two elementary school teachers; Lisa and Jenny, who participated in a three year SWH professional development as teachers transitioned from traditional classroom approach to a more student-centered instructional approach. Cavagnetto (2006) collected data from the first two years of the project (2004-2006 academic year) focusing on the effects of two different student-centered approaches on classroom environment and student achievement in science and language arts. Students were placed into two groups; small-group and whole-class discussion groups. In addition, teacher implementation was measured by the Reform Teaching Observation Protocol (RTOP). As found in Gunel's (2006) study, teacher changes were complicated. The results indicated that Lisa moved from a teacher-controlled instruction to a student-centered and teacher-managed instruction, that is, while the instruction was centered on students, the end point was controlled by the teacher. Her RTOP scores shifted from 37.5 at the early stage to 88.9 at the end of the study where the higher score means teachers moved toward reformed teaching as identified in the National Science Education Standards (NRC, 1996) with a total possible score of 100. Jenny, like Lisa, moved away from teacher centered classroom toward reformed teaching but at a slower rate. Jenny's RTOP scores were listed at 40 at the beginning and moved up to 79 at the end of the study.

Furthermore, Cavagnetto (2006) also explored the impact of the student-centered approach on student achievement in both teacher-generated exam and standardized exam

(Iowa Tests of Basic Skills; ITBS). Comparing the participating students' science ITBS score with the national average score, the participating students scored about one-half grade level above the national average. However, using a two-way Analysis of covariance (ANCOVA) revealed no significant difference between students' science ITBS score in small group discussion and whole-class discussion approaches. In contrast, the analysis of a student end-of unit exam using Univariate analysis of variance (ANOVA) revealed differential effects of the instructional strategies. Cavagnetto (2006) found that students in the small group treatment scored better than the whole class treatment during unit two. The growth that occurred between pre and post- test also was significantly different between two teachers in unit two and three. Cavagnetto (2006) argued that the higher rate of growth correlated to teacher pedagogy during the units. In this case, Jenny allowed greater opportunities for student voice, thus, the better growth was observed.

Another study related to the use of writing-to-learn activities in the SWH approach was conducted by McDermott and Hand (2010). The study explored the impact of non-traditional writing tasks on students learning. After experience the non-traditional writing tasks, students recognized the benefits of the writing tasks to their scientific learning and understanding in science classrooms. In addition, students also recognized the important aspect of the non-traditional writing task having to write for different audiences as they needed to alter vocabulary, from technical terms used in class to everyday language (McDermott & Hand, 2010). The impact of non-traditional writing activities in the SWH approach on student performance on conceptual questions were

cumulative and increased over time when a student had multiple experiences with the activities (Hand, Hohenshell, & Prain, 2004).

Furthermore, Gunel, Hand, and Prain (2007) summarized the impacts of the SWH approach on student learning from previous published studies related to writing-to-learn strategies in science classrooms. The results showed that students in treatment groups showed significantly greater gains in both total test scores and conceptual questions over students in control groups. The small effects sizes were generally found in the conceptual questions with multiple-choice format. However, the impacts of the SWH approach were bigger when observed with total test scores. A more recent study in Turkey (Kingir, Geban, & Gunel, 2012) showed similar results. For ninth-grade Turkish students, the SWH approach contributed to better test performances by students compared to the traditional. Moreover, low and middle achievement students in the SWH group significantly outperformed students in the traditional group in their post-test scores.

Studies among the SWH approach also showed narrowing achievement gaps among advantaged and disadvantaged students. Implementing the argument-based inquiry approaches may be influential in closing the achievement gap by embedding writing and argumentation within the science inquiry activity (Grimberg & Hand, 2009, Cavagnetto, 2010). In order to achieve equitable learning for all, instruction have to values and respect students' experience and prior knowledge, provide adequate support and resource, and create non-threatening environment (Villanueva & Hand, 2011). While researchers focus on closing achievement gaps by gender, ethnicity, and socioeconomic status, Villanueva, Taylor, Therrien, and Hand (2012) also suggested that student success was also greatly increased for students who have difficulty with reading and writing skills

using the inquiry. The SWH approach provides opportunities for student with disabilities to learn science by experiencing evidence-based strategies incorporating the use of appropriate scaffolds and supports (Villanueva et al., 2012). The use of peer-assisted learning and student-generated explanation found in the SWH classroom also helps students with learning disability improving their science achievement (Therrien et al., 2011).

According to the evidence from various studies using the SWH approach presented in this section, the results consistently demonstrated the positive impacts of the approach on student learning and teacher implementation. However, there was not enough evidence showing the effectiveness of the approach, particularly on achievement levels and standardized tests, to persuade teachers to shift from traditional to non-traditional in the science classroom (Kingir, Geban, & Gunel, 2012).

Summary of Literatures

In summary, inquiry has been critically influential among the science education community for decades. Although various inquiry-based instructional model were introduced to help bridging the gap between theory and classroom practices, they were connected and overlapped. All approaches aimed to help students developing deeper understanding of scientific concepts as well as scientific reasoning and critical thinking skills. Research has studied the impacts of inquiry-based instruction on student achievement. Negative and neutral impacts have been reported. However, the majority of studies agreed that the inquiry-based instructional model positively impacted student learning, especially disadvantaged students.

As the science education community transitions to the Next Generation Science Standards, argumentation in science classroom is highlighted as a critical tool to develop better and deeper understanding of scientific concepts. The SWH approach is one of argument-based inquiry approaches promoting the uses of language practice in science classroom by engaging in negotiation scientific argumentation. Scholars have been showing that the SWH approach helps students develop deeper understanding of scientific concepts, critical thinking and reasoning skills, and positive attitude toward science.

However, most studies exploring the effectiveness of the inquiry-based intervention including the SWH approach were limited to the short-term studies. Moreover, studies about the impacts of inquiry on standardized tests have been neglected from science education community. In order to develop better understanding about the effectiveness of the inquiry-based intervention, the long-term impacts are needed to be explored.

CHAPTER III

METHODOLOGY

To address the research questions, a quantitative approach is used as the methodology. This chapter provides rationale for the use of the multilevel model to examine student achievement growth. Next, background information about the study, context of the participating school district, data structure, and descriptive statistics of participating students are described. Finally, the details of the related variables, statistical analysis techniques, and detailed procedures are also discussed.

Context of the Study

The purpose of this study is to examine the long-term effects of the Science Writing Heuristic (SWH) approach on the growth of students' Iowa Test Basic of Skills (ITBS) scores in one school district. The longitudinal science ITBS scores were collected from all students in the school district between 2000 and 2011. The science achievement trajectories prior and after the adoption of the SWH approach were examined and compared.

The school district is located in a moderate sized, federally designated impoverished town in rural Iowa. It was involved with a long term SWH professional development program with two phases. Beginning in 2002, high school and middle school teachers were trained in the SWH approach and the project ended in 2005. In 2005, the training expanded in the second phase to include elementary school teachers and ended in 2008. All participating teachers were trained during a five-day summer workshops and 4.5 one-day events during the school year across three years. In addition,

all participating teachers were regularly visited and supported by researchers and professional developers throughout the school year. After completion of the professional development project, the school district decided to adopt the SWH approach as the main approach in teaching and learning science.

Currently, the school district has approximately 2000 students who are predominantly Caucasians (95%). About 25%-30% are considered coming from low socioeconomic status families due to receipt of free or reduced lunch. There were also approximately 15%-20% of students with special needs identified by their Individual Educational Program (IEP). In total, 6 third-grade and 6 fourth-grade classes with one teacher who teaches every subject areas including science participated in the elementary school. In middle school, the class size was about 25 students with two science teachers in each grade. Each science teacher is responsible for half of the students at each grade level. There are 7 science teachers in high school who are responsible in one science area each. Although most teachers participated in the original professional development, there were new teachers who joined the school after the professional development. Because the school district adopted the SWH approach as the main approach in science teaching, new teachers were trained by the experienced teachers.

This study is an extension study of Omar's (2002), Gunel's (2006) and Cavagnetto's (2006) original researches. At early stage of the professional development, Omar (2002), Gunel (2006) and Cavagnetto (2006) had explored parts of the impacts of the SWH approach on teachers and students of the same school district of this study. Details of these studies are presented in chapter II. However, these studies were conducted with a limited number of participants. The current study expanded these

studies by including data from all students in the school district starting in elementary and following them through high school. This study also included data across a longer period time frame including before the SWH professional development project and after the professional development when the support from the professional developers was removed. Given students' repeated observations across grade, more appropriate statistical analysis, a multilevel model for change, was used to track longitudinal growth of students' achievement.

The Multilevel Model for Change with Discontinuity in Elevation and Slope

The multilevel model for change (Singer and Willett, 2003) is one of the statistical techniques associated with in Hierarchical Linear Modeling (HLM). HLM has increased in popularity and acceptance in the social science area for the analysis of longitudinal data, such as repeated measures within individuals, to assess changes in the growth rate of various outcomes. Traditional approaches including multivariate analysis of variance (MANOVA) or repeated measures design, have been critiqued as failing to recognize interactions between levels of data, an important characteristic of longitudinal data.

Singer and Willett (2003) suggested that the multilevel model for change is suitable for exploring individual longitudinal change that has 1) multiple waves of data, 2) an outcome whose values change systematically over time, and 3) a sensible metric for clocking time. Multilevel modeling also has an advantage over other approaches because it considers different growths curve for each subject (Hox, 2010). For these reasons, the

multilevel model for change was selected to examine the growth trajectories of science ITBS scores over time. Individual students could also have different growth trajectories based on variables associated with their exposure to the SWH approach. In order to explore changes in achievement trajectories before and after the school-wide SWH approach adoption, the multilevel model for change with discontinuous individual change (Singer & Willett, 2003) was used to estimate the impact of the SWH approach as an argument-based inquiry on students' achievement. Discontinuous individual change in multilevel modeling is employed when individual growth is expected to change at a certain time. Singer and Willett (2003) suggested three types of discontinuities which might occur in an individual's growth: discontinuity in elevation only, discontinuity in slope only, and discontinuity both elevation and slope, when an intervention is introduced to a classroom

Figure 3.1 illustrates examples of discontinuity models of student's achievement trajectory. Figure 3.1A shows a discontinuity in elevation model where students received an intervention at time = 2. As a result, the students' achievement trajectories are immediately bumped up and remain parallel to the regular achievement trajectory. In figure 3.1B, no immediate bumping of trajectory is found; however, students' achievement grows at a different rate after exposure to an intervention compared to regular trajectories without any exposure. Figure 3.1B is an example of discontinuity in slope only, not elevation. Figure 3.1C represented a new trajectory with discontinuity in both slope and elevation. There is an immediate bump after receiving an intervention with continuous growth at a different rate.

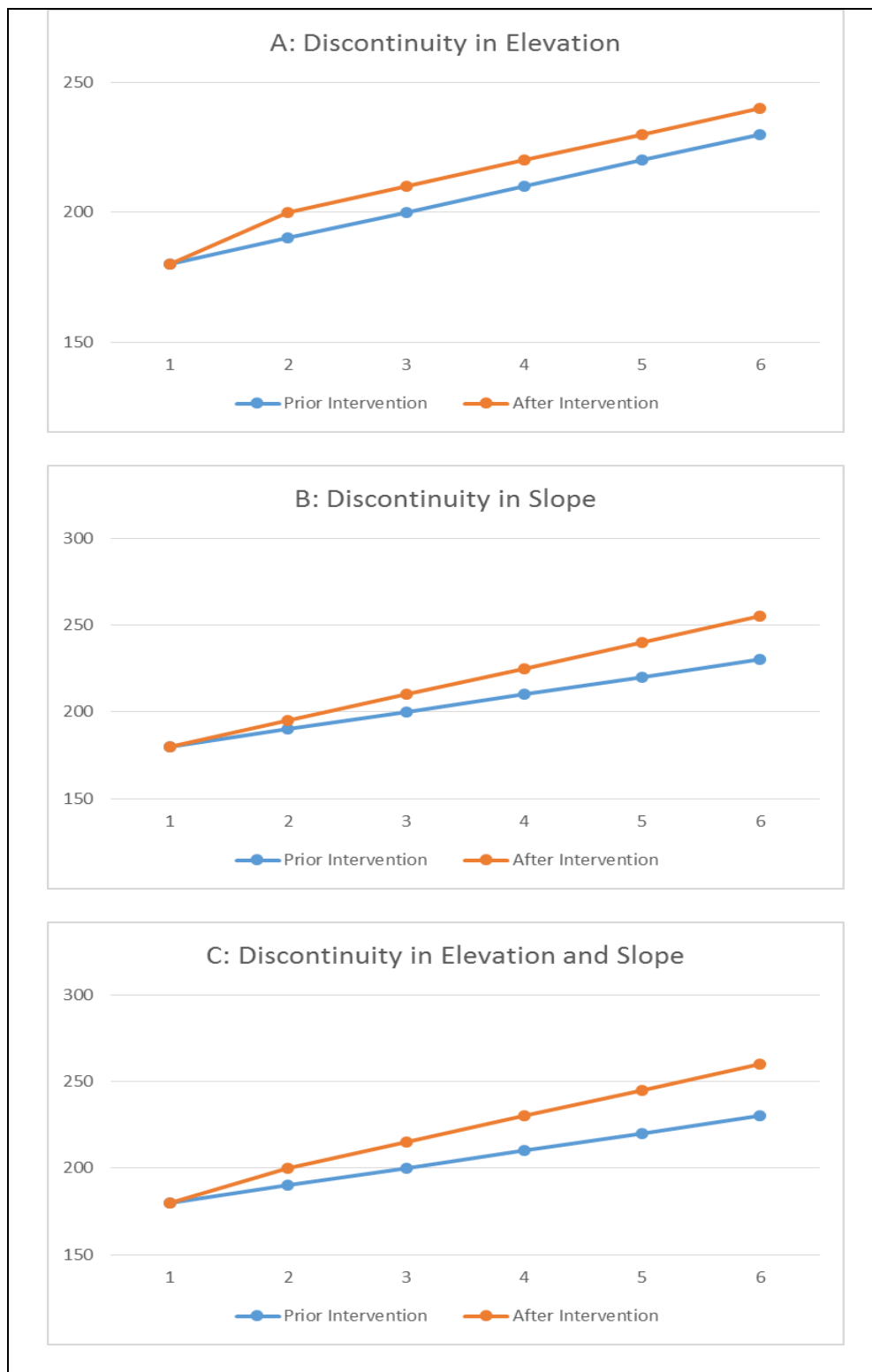


Figure 3.1 Examples of Multilevel Models for Change with Discontinuity

Adopting from Singer and Willett's (2003) the multilevel model for change with discontinuous individual change, in this study, I hypothesized that the effectiveness of the SWH approach, if any, would interrupt the smoothness of individual growth trajectories. When teachers started to adopt the SWH approach in their classrooms, three types of shifting might occur in an individual's growth: elevation only, slope only, and both elevation and slope.

A discontinuity in elevation only would be expected if a growth trajectory was bumped up upon receiving the SWH approach but there was no effect on subsequent rate of change. Discontinuity in slope only would be expected if the slope of growth scores was changed after receiving the SWH approach with no corresponding shift in elevation or mean score. In some cases, discontinuity in both elevation and slope might be found at the same time.

Data

The Standardized Test

The Iowa Tests of Basic Skills (ITBS). The ITBS tests measure educational achievement in 15 subject areas for kindergarten through grade 8 (Hoover, Dunbar, Frisbee, 2003). The test was developed by faculty and professional staff at the University of Iowa. The main purpose of the ITBS is to provide information to be used in designing effective classroom instruction that supports students' learning. The ITBS is administered at three different time frames during the academic school year: Fall, Mid-Year, and Spring. The scores are reported in three different scales: standard scores, grade equivalent scores, and percentile ranks. The internal-consistency reliability coefficients

calculated by Kuder-Richardson Formula 20 (K-R20) were reported ranging from 0.699 to 0.980, mostly in 0.8-0.9 range. The equivalent-forms reliabilities range from 0.64 to 0.86. These coefficients were calculated from scores in two forms of the complete battery tests.

Iowa Test of Educational Development (ITED). The Iowa Test of Educational Development or ITED is similar to the ITBS and it was developed by the University of Iowa. It measures achievement in nine areas including science and reading comprehension for students from grades nine to twelve. Studies (Rosemier, 1962; Loyd, Forsyth, Hoover, 1980) showed high correlation between the ITBS and ITED composites indicating these two tests measure similar constructs.

Data Collection

The longitudinal ITBS data were collected from two main sources. First, Iowa normative scores, showing Iowa students' achievement for the past ten years were collected from the Iowa testing program. Descriptive statistics (number of students, mean, and standard deviation) of mid-year ITBS and ITED scores for school years 2000-2001 to 2010-2011 in two areas (science and reading comprehension) for all students in Iowa were collected from the Iowa Testing Program, College of Education, University of Iowa. Descriptive statistics were also collected based on students' demographics; gender, socioeconomic status (SES) as indicated by participation in the free/reduced lunch program, and Individual Educational Program (IEP).

Second, individual student ITBS scores were collected directly from the school district. All individual students' ITBS scores in science and reading comprehension were

collected from the school for years corresponding to the Iowa norm scores; 2000-2001 to 2010-2011. Students' demographic information was also collected (gender, SES, and IEP.) Data for the ITBS from elementary students is missing because, prior to the 2004-2005 school year, science ITBS tests were not administered to the elementary school students.

In this study, students who started school in the same grade level at the same time were considered as a student cohort. The data were collected from a total of 19 cohorts of students. There were approximately 13,364 data points collected from all students. Note that information from 6th grade students could not be collected before 2009-2010 academic year because of the administrative structure of the school district. Prior to 2009, all students in the school district had to move to grade 6 in another school district.

Numbers of students in each Cohort were listed in Table 3.1. Highlighted numbers represented numbers of students exposing to the SWH approach. For example, Cohort eleven (see Table 3.1) were students who started first grade together in the 2000-2001 academic year. However, the science ITBS was not administered for elementary school students before the 2004-2005. Thus, science ITBS scores Cohort eleven were not available from grade 3 to 4. The science ITBS scores of 155 fifth-graders of Cohort eleven were collected in the 2004-2005. Because Cohort eleven students had to join another sixth grade at other school district in the 2005-2006, their ITBS scores were not accessible. The students in cohort eleven started engaging with the SWH approach at seventh grade in the 2006-2007 and continued experiencing with the SWH approach to eleventh grade in the 2010-2011 (the number of students: 175, 174, 198, 175, and 159, were highlighted, see Table 3.1).

Overall, students in the school district can be categorized into three different groups based on their experience with the SWH approach. Students in Cohort 1-3 had never experienced SWH throughout their school life. Students in Cohort 4-13 had mixed experience because they were in school while the SWH professional development took place. They had experience with traditional school science prior to the beginning of the project and their experience had changed after their teachers participated in the SWH professional development. Finally, students in Cohort 14-19 had experience with the SWH approach throughout their school experience.

Variables

Drawing from the review of literatures and research questions of this study, the following variables were identified and examined. In this study, two levels of data were identified. The level one data correspond to variables that could change every time students completed the ITBS including their scores, grade level, and quadratic term of grade level. In addition, experience with the SWH approach and number of consecutive years of experience with the SWH approach as level-1 variables are created as the indicator of continuity of elevation and slope, respectively. Level two data correspond to fixed variables for individuals that do not change over time, such as gender, ethnicity, etc.

Dependent variable: individual science ITBS scores are examine as the dependent variable.

Table 3.1 Number of Participating Students in Each Cohort (Number of students exposing to the SWH was highlighted)

Cohort	Year	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
1	Grade	11										
	N	201										
2	Grade	10	11									
	N	NA	207									
3	Grade	9	10	11								
	N	176	NA	191								
4	Grade	8	9	10	11							
	N	158	183	NA	194							
5	Grade	7	8	9	10	11						
	N	165	162	187	190	181						
6	Grade	6	7	8	9	10	11					
	N	NA	123	130	159	152	178					
7	Grade	5	6	7	8	9	10	11				
	N	NA	NA	179	203	220	221	200				
8	Grade	4	5	6	7	8	9	10	11			
	N	NA	NA	NA	NA	203	223	217	189			
9	Grade	3	4	5	6	7	8	9	10	11		
	N	NA	NA	NA	NA	182	201	205	182	179		
10	Grade	3	4	5	6	7	8	9	10	11		
	N	62	60	142	NA	180	174	165	168	155		

Table 3.1. Continued

Cohort	Year	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11
11	Grade			3	4	5	6	7	8	9	10	11
	N			NA	NA	155	NA	175	174	198	175	159
12	Grade				3	4	5	6	7	8	9	10
	N					156	155	NA	178	176	204	164
13	Grade					3	4	5	6	7	8	9
	N					97	159	165	NA	175	173	165
14	Grade						3	4	5	6	7	8
	N						89	158	158	NA	183	170
15	Grade							3	4	5	6	7
	N							90	170	175	173	170
16	Grade								3	4	5	6
	N								83	150	144	138
17	Grade									3	4	5
	N									158	163	149
18	Grade										3	4
	N										134	122
19	Grade											3
	N											135

Note: number of students exposing to the SWH was highlighted

Level-1 variables: level-1 variables correspond to data that could change every time students completed the ITBS including their scores. Four variables are listed as follow;

First, GRADE; variable indicates the grade level of students when they took the test. According to the construct of the ITBS test, the standard scores directly relate to student grade level. The standard scores generally improve from younger grades to higher grades indicating growth in achievement.

Second, GRADE^2 ; variable corresponds to the quadratic term of GRADE ($\text{GRADE}^2 = \text{GRADE} * \text{GRADE}$). According to the reported average ITBS standard scores, the growth scores are not equal from grade to grade. In general, students in younger grades tend to have bigger gains from year to year compared to students in higher grades. The quadratic term of GRADE will allow the researcher to explore non-linear growth in the science ITBS scores.

Third, SWH; the variable indicates whether students had experience with the SWH approach at particular grade level. The SWH variable explored discontinuity in elevation of mean of the science standard score growth trajectories.

Finally, CONSWH; the variable indicating the number of consecutive year that students were exposed to the SWH approach. The CONSWH variable models potential discontinuity in slope.

Level-2 variables; level two variables correspond to fixed data for individuals that do not change over time. Five level-two variables are listed as follow;

First, GENDER: dichotomous variable represents student gender.

Second, SES: dichotomous variable represents socioeconomic status indicated by school lunch plan that a student received each year. In general, individual's lunch plan could change annually. In this data set, students who received free/reduced lunch plan more than half (ie, average > 0.5) of their school experience were considered low-socioeconomic status. Students who never received free/reduced lunch plan or receive free/reduced lunch plan less than half (ie, average < 0.5) of their school experience were not considered low-socioeconomic status.

Third, IEP: dichotomous variable represents student's learning difficulty or disability indicated by the presence of the Individualized Educational Program (IEP). In general, Individualized Educational Program status could change annually. In this data set, students who were identified with IEP more than half (ie, average > 0.5) of their school experience were considered as having a disability or learning difficulty. Students who were never identified with IEP, or having less than half (ie, average < 0.5) of their school experience were considered to not have a disability or learning difficulty.

Fourth, InREAD; variable represents level of an individual's ITBS reading achievement. Since the SWH approach is argument-based inquiry instruction involving all linguistic skills i.e., reading, writing, talking, and listening, a student's linguistic skills was identified as a variable of interest. In this study, the earliest obtained ITBS reading achievement scores for each student were divided into four levels based on their reading percentile rank scores. One of the important interpretations of students' ITBS scores is that there are certain cut scores used to classify students into three levels of proficiency based on their percentile rank scores. Students who score under the 41st percentile are considered below proficient. Students who score above the 41st percentile are considered

proficient. And students who score above the 91st percentile are considered highly proficient.

For this study, the 26th percentile rank was introduced as the third cut score creating 4 achievement levels with approximately equal number of students in highly proficient level and extremely below proficient level. Using, 26th, 41st, and 91st percentile ranks as the cut score, four levels of proficiency were identified as flowing; students who score above the 91st percentile were considered high proficient. Students who score above the 41st percentile were considered proficient. Students who score above the 26th percentile were considered below proficient. And students who score below the 26th percentile were considered extremely below proficient.

Finally, StartLev; variable represents school level where students started experiencing with the SWH approach where 0 = students never experienced with the SWH approach, 1 = students started experiencing with the SWH approach in high school (grade 9-11), 2 = students started experiencing with the SWH approach in middle school (grade 6-8), 3 = students started experiencing with the SWH approach in elementary school (grade 3-5), and 4 = students started experiencing with the SWH approach in early elementary school.

Dependent, level-1, and leve-2 variables including their values corresponding to the research questions are summarized in table 3.2.

Table 3.2 List of Variables Used in Multilevel Models Examining the Student's ITBS Growth Trajectories

Dependent variables	
SCIENCE	Science ITBS standard score
Level-1 variables	
GRADE	Grade when student took the test
GRADE ²	GRADE*GRADE (quadratic term of GRADE)
SWH	Exploring discontinuity in elevation: dichotomous variable where 1= students experienced with the SWH approach in that particular year and 0 = students experienced with traditional classroom
CONSWH	Exploring discontinuity in slope: number of consecutive year that students had exposure to the SWH approach where 0 = before exposed to the SWH approach, 1= year first year of the SWH approach exposure, 2 = second year of the SWH approach exposure, 3 = third year of the SWH approach exposure and so on
Level two variables	
GENDER	Dichotomous variable where 1= Female and 0= Male
SES	Dichotomous variable where 1= students with identified free/reduced lunch plan and 0= students with no identified free/reduced lunch plan
IEP	Dichotomous variable where 1= students with identified individual educational program and 0= students with no identified individual educational program
InREAD	Levels of individual's ITBS reading achievement where 1= extremely below proficient level, 2= below proficient level, 3= proficient level, 4= high level.
StartLev	School levels where students started experiencing with the SWH approach where 0 = students never experienced with the SWH approach, 1 = high school (grade 9-11), 2 = middle school (grade 6-8), 3 = elementary school (grade 3-5), and 4 = kindergarten school (K-2)

Data Organization

Creating a Longitudinal Data Set

Before conducting the multilevel model for change analysis, two data sets were created. First, a person-period data set contain multiple record for each person where each record contained data for each measurement occasion. The person-period data set

contains all level-one variables that detailed value changes for each time students took the test. Table 3.3 displays a sample person-period data set from three students.

Table 3.3 Example of a Person Period Data Set

Cohort	ID	GRADE	GRADE²	SWH	CONSWH	SCIENCE
4	447723	3	9	0	0	193
4	447723	4	16	1	1	205
4	447723	5	25	1	2	220
4	447723	7	49	1	3	244
4	447723	8	64	1	4	260
7	129990	5	25	0	0	223
7	129990	6	36	1	1	230
7	129990	7	49	1	2	242
2	474972	4	16	0	0	198
2	474972	5	25	0	0	211
2	474972	6	36	1	1	219
2	474972	7	49	1	2	223
2	474972	8	64	1	3	230

Second, person-level data set contain data for level-two variables was created.

There is only one record for each person in this data set. Table 3.4 displays a sample person-level data set for the same three students listed in Table 3.3.

Table 3.4 Example of a Person-level Data Set

Cohort	ID	GENDER	IEP	SES	InREAD	StartLev
4	447723	1	1	0	2	
7	129990	0	0	0	3	
2	474972	1	0	0	1	

Handling Missing Data

Time-varying missing data: this study used existing longitudinal data in a natural school setting, thus, missing data were expected. Two types of missing data at level 1 were found; missing at random and systematic missing data. Level-1 missing data usually have no threat to validity and reliability of the models. However, special attention should be paid to the systematic missing data. Although the ITBS test was mandatory test for all students, students sometimes miss the test in various years. The researcher considered this type of missing data as missing at random (MAR) (Schafer, 1997) which did not create any bias for the analysis. Singer and Willett (2003) also argued that the analysis of longitudinal data with MAR still permits valid generalizations of multilevel models for change.

A more systematic missing data found in this dataset was sixth-grade missing data. Prior to 2009, all students in the participating school district joined sixth grade class in a nearby school district; therefore, I did not have access to approximately 600 sixth-grade ITBS scores for all students prior to 2009. To handle this systematic missing data, a simple imputation technique was used by calculating a mean score using fifth and seventh grade science scores. The analysis of the imputed database reproduced trends found in the analysis of the non-imputed database. Thus, the non-imputed database was used in the analysis process.

Level-2 missing data: HLM2 assumes complete data, thus all level-2 cases with missing data were deleted.

Descriptive Statistic Analysis

The original database contained student's ITBS scores across 19 cohorts consisting of 3,914 students and 13,364 data points. To conduct reliable and efficient multilevel modeling for change, at least three waves of data per person is required. Thus, cohort one, two, and nineteen and individual students with fewer than three waves of data were excluded from the dataset. Moreover, outliers identified as having standard scores exceeded 3.00 ($Z > 3.00$) were also removed (Osborne & Overbay, 2004). Finally, there were 14 students cohorts consisting of 2,740 students and 11,727 data points used in the analysis. The number of students, their average science ITBS scores, and the standard deviation for each cohort of the final dataset used were presented in Table 3.5. Numbers of students in each cell were approximately equal, except elementary school level in early year.

Due to the complexity of the data structure, available data points were varied based on cohort and individual student. Table 3.6 shows the details of number of data points collecting from each cohort. For example, for Cohort 11, students' science ITBS scores were available from 5th to 11th grade, except 6th grade, totally 6 data points. On average, approximately 4.7 (SD=1.53) data points or about 78.5% were collected for Cohort 11. According to the table 3.6, overall, over 80% data points were collected indicating high percentage data points for each students available for the analysis.

Table 3.5 Descriptive Statics Showing Number of Students, Mean and Standard Deviation of Science ITBS Scores

Grade	Cohort	4	5	6	7	8	9	10
11	N	184	178	153	191	180	167	145
	Mean	288.07	302.83	304.25	298.1	304.98	297.22	303.96
	SD	47.74	38.76	43.70	36.29	42.10	40.65	42.27
10	N		189	148	213	207	182	163
	Mean	N/A	282.12	287.14	279.41	286.08	285.88	285.06
	SD		41.89	40.59	40.98	35.85	40.81	40.78
9	N	182	188	157	218	215	199	169
	Mean	275.45	275.99	276.46	274.83	279.41	284.52	277.51
	SD	42.61	40.78	42.91	38.03	35.57	35.84	39.43
8	N	158	161	128	200	189	198	171
	Mean	258.78	261.33	260.61	265.74	266.33	266.65	263.21
	SD	39.95	40.38	38.92	35.65	39.15	34.17	35.78
7	N		162	123	178		181	161
	Mean	N/A	259.65	255.07	240.08	N/A	264.8	252.82
	SD		39.88	38.49	38.11		34.92	36.84
6	N							
	Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SD							
5	N							147
	Mean	N/A	N/A	N/A	N/A	N/A	N/A	227.69
	SD							28.19
4	N							
	Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SD							
3	N							
	Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	SD							

Table 3.5. Continued

Grade	Cohort	11	12	13	14	15	16	17
11	N	156						
	Mean	303.27	N/A	N/A	N/A	N/A	N/A	N/A
	SD	42.14						
10	N	164	164					
	Mean	286.17	294.08	N/A	N/A	N/A	N/A	N/A
	SD	39.83	36.26					
9	N	188	186	165				
	Mean	279.19	281.85	277.68	N/A	N/A	N/A	N/A
	SD	38.62	35.22	33.39				
8	N	171	171	173	168			
	Mean	267.25	266.55	273.17	266.44	N/A	N/A	N/A
	SD	33.34	34.95	31.54	34.66			
7	N	173	172	170	176	170		
	Mean	256.12	256.22	259.46	259.15	258.11	N/A	N/A
	SD	36.41	31.47	33.48	32.37	33.15		
6	N					170	137	
	Mean	N/A	N/A	N/A	N/A	246.08	247.77	N/A
	SD					29.22	35.25	
5	N	147	152	159	154	171	141	149
	Mean	227.69	229.74	228.57	221.57	229.41	234.72	230.03
	SD	28.19	28.25	29.83	31.45	26.31	30.07	28.72
4	N		147	151	154	166	141	158
	Mean	N/A	206.05	211.4	208.27	212.83	212.92	213.28
	SD		24.00	27.72	27.35	26.68	24.51	25.66
3	N			86	85	86	81	150
	Mean	N/A	N/A	190.92	189.58	192.38	193.77	192.97
	SD			21.84	21.77	18.97	22.08	18.19

Table 3.6 Average and Percentage of Data Point Available in Each Cohort

Cohort		Science	Cohort		Science
4	Range	3	11	Range	3-6
	Average	3 (100%)		Average	4.71 (78.5%)
	SD	0		SD	1.53
5	Range	3-5	12	Range	3-6
	Average	4.58 (91.6%)		Average	4.56 (76.0%)
	SD	0.78		SD	1.68
6	Range	3-5	13	Range	3-6
	Average	4.41 (88.2%)		Average	4.51 (75.2%)
	SD	0.84		SD	1.42
7	Range	3-5	14	Range	3-5
	Average	4.46 (89.2%)		Average	3.95 (79.0%)
	SD	0.89		SD	1.09
8	Range	3-4	15	Range	3-5
	Average	3.49 (87.2%)		Average	4.04 (80.8%)
	SD	0.84		SD	1.04
9	Range	3-5	16	Range	3-4
	Average	4.04 (80.8%)		Average	3.38 (84.5%)
	SD	1.33		SD	0.64
10	Range	3-5	17	Range	3
	Average	4.05 (81.0%)		Average	3 (100%)
	SD	1.29		SD	0

There were totally 2,740 students from 14 cohorts included in this study. The analysis of descriptive statistics showed that there were approximately equal proportion of males (51.0%) and females (49.0%). In general, there were 13.2% of students were identified with IEP status. The proportions of IEP student were approximately stable across all cohorts. The majority of students in this school district were Caucasians (94.8%). Although the population of students in the school district were stable, there were

slight changes in student demographics. First, Caucasians students were counted for more than 95% in early cohorts while the proportion of Caucasian students started to decrease to approximately 90% in later cohorts.

In addition, there were 27.8% of students who required a free/reduced lunch program. The proportions of students from low-income family tended to increase over years. Overall, the majority of students ($61.0\% + 13.6\% = 74.6\%$) were proficient in their reading comprehension ITBS tests. There were 12.2% of students who performed extremely below proficient level in their reading comprehension ITBS tests. Table 3.7 provided extensive descriptive statistics information of participating students including details of each cohort.

Table 3.7 Total Number and Percentage of Students Based on Student’s Backgrounds

Cohort	Gender		Ethnicity		IEP		Low SES		Initial Reading Achievement Level				Start Level		
	Total		Male	Female	White	Others	No	Yes	No	Yes	1	2		3	4
4	192	N	92	100	177	15	176	16	158	34	34	23	105	30	1
		%	47.9	52.1	92.2	7.8	91.7	8.3	82.3	17.7	17.7	12	54.7	15.6	
5	161	N	79	82	160	1	146	15	142	19	25	20	93	23	1
		%	49.1	50.9	99.4	0.6	90.7	9.3	88.2	11.8	15.5	12.4	57.8	14.3	
6	225	N	117	109	221	4	194	30	191	33	44	32	125	23	1
		%	51.8	48.2	98.7	1.3	86.6	13.4	85.3	14.7	19.6	14.3	55.8	10.3	
7	225	N	95	130	217	8	192	33	172	53	35	26	141	23	2
		%	42.2	57.8	96.4	3.6	85.3	14.7	76.4	23.6	15.6	11.6	62.7	10.2	
8	224	N	118	106	217	7	190	34	164	60	28	33	137	26	2
		%	52.7	47.3	96.9	3.1	84.8	15.2	73.2	26.8	12.5	14.7	61.2	11.6	
9	202	N	100	102	194	8	174	28	138	64	25	32	125	20	2
		%	49.5	50.5	96	4	86.1	13.9	68.3	31.7	12.4	15.8	61.9	9.9	
10	195	N	106	89	187	9	164	31	147	48	33	25	121	16	2
		%	54.4	45.6	95.9	4.1	84.1	15.9	75.4	24.6	16.9	12.8	62.1	8.2	
11	213	N	108	105	207	9	189	27	161	55	25	28	128	32	2
		%	50.9	49.1	95.8	4.2	87.5	12.5	74.5	25.5	11.7	13.1	60.1	15	
12	222	N	123	99	206	16	199	23	141	81	26	18	146	32	3
		%	55.4	44.6	92.8	7.2	89.6	10.4	63.5	36.5	11.7	8.1	65.8	14.4	

Table 3.7 Continued

Cohort	Gender		Ethnicity		IEP		Low SES		Initial Reading Achievement Level				Start Level		
	Total		Male	Female	White	Others	No	Yes	No	Yes	1	2		3	4
13	199	N	105	94	186	13	175	24	131	68	29	26	117	27	3
		%	52.8	47.2	93.5	6.5	87.9	12.1	65.8	34.2	14.6	13.1	58.8	13.6	
14	186	N	100	86	174	12	156	30	121	65	21	20	114	31	3
		%	53.8	46.2	93.5	6.5	83.9	16.1	65.1	34.9	11.3	10.8	61.3	16.7	
15	189	N	92	97	175	14	162	27	112	77	17	23	118	31	4
		%	48.7	51.3	92.6	7.4	85.7	14.3	59.3	40.7	9	12.2	62.4	16.4	
16	148	N	82	66	139	9	121	27	97	51	10	10	98	30	4
		%	55.4	44.6	93.9	6.1	81.8	18.2	65.5	34.5	6.8	6.8	66.2	20.3	
17	159	N	81	78	142	17	142	17	105	54	9	19	103	28	4
		%	50.9	49.1	89.3	10.7	89.3	10.7	66	34	5.7	11.9	64.8	17.6	
Total	2740	N	1398	1343	2602	142	2380	362	1980	762	361	335	1671	372	
		%	51	49	94.8	5.2	86.8	13	72.2	27.8	13.2	12.2	61	13.6	

Data Analysis Procedure

The simple model of change in students' achievement can be described as:

Level 1 (repeated measures level) model:

$$\text{SCIENCE}_{ti} = \beta_{0i} + \beta_{1i} \text{TIME}_{ti} + \varepsilon_{ti}$$

where SCIENCE_{ti} is an individual student's ITBS score at TIME_{ti} .

β_{0i} is the expected estimation of the SCIENCE score for the i^{th} individual at TIME zero.

β_{1i} is the average annual change in estimation of the SCIENCE score for the i^{th} individual over time.

ε_{ti} is the within-individual random variation.

Level 2 (individual-level) models;

$$\beta_{0i} = \gamma_{00} + U_{0i}$$

$$\beta_{1i} = \gamma_{10} + U_{1i}$$

where γ_{00} is population average true initial status for nonparticipants.

U_{0i} is the difference in population average true initial status between participants and nonparticipants.

γ_{10} is population average annual rate of true change for nonparticipants.

And U_{1i} is difference in population average annual rate of true change between participants and nonparticipants.

To answer the first and second research questions whether the SWH approach had impacts on students' longitudinal science ITBS achievement, and whether the SWH

approach equally impacted different groups of students (research question one and two); first, building on simple model of change, level-one model was tested and developed from the following variables; GRADE, GRADE², SWH, and CONSWH. Second, the following level-two model variables; GENDER, IEP, SES, and InREAD, were add into model separately to test the impacts of the SWH approach on each variable. Third, GENDER, IEP, SES, and InREAD, were added and test simultaneously. Finally, the final model was developed based on previous results.

Level-1 model:
$$\text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}_{ti}) + \pi_{2i} * (\text{GRADE}_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

Level-2 model:
$$\pi_{0i} = \beta_{00} + \beta_{01} * (\text{GENDER}_i) + \beta_{02} * (\text{IEP}_i) + \beta_{03} * (\text{SES}_i) + \beta_{04} * (\text{InREAD}_i) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (\text{GENDER}_i) + \beta_{12} * (\text{IEP}_i) + \beta_{13} * (\text{SES}_i) + \beta_{14} * (\text{InREAD}_i)$$

$$\pi_{2i} = \beta_{20} + \beta_{21} * (\text{GENDER}_i) + \beta_{22} * (\text{IEP}_i) + \beta_{23} * (\text{SES}_i) + \beta_{24} * (\text{InREAD}_i)$$

$$\pi_{3i} = \beta_{30} + \beta_{31} * (\text{GENDER}_i) + \beta_{32} * (\text{IEP}_i) + \beta_{33} * (\text{SES}_i) + \beta_{34} * (\text{InREAD}_i) + r_{3i}$$

$$\pi_{4i} = \beta_{40} + \beta_{41} * (\text{GENDER}_i) + \beta_{42} * (\text{IEP}_i) + \beta_{43} * (\text{SES}_i) + \beta_{44} * (\text{InREAD}_i) + r_{4i}$$

To evaluate the validity and reliability of multilevel models, after multilevel models for change were fitted, the residual files for level-1 and level-2 models were produced to check the fit and distributional assumptions of the models.

Moreover, even though multilevel modeling is considered as an appropriate method for the longitudinal data structure such as science achievement scores, the intraclass correlation coefficient (ICC) and the design effect were used to confirm the need for using a multilevel modeling from unconditional means model (Peugh, 2010).

Intraclass correlation coefficient (ICC) represents the proportion of the variance in outcome among students:

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2}$$

Spybrook, Raudenbush, Liu, Congdon, and Martinez (2006) stated that ICC values that exceed 0.40 are common in longitudinal social research studies.

The design effect quantifies the effect of independence violation on standard error estimate and is an estimate of the multiplier that needs to be applied to standard errors to correct for the negative bias results from nested data. The design effect is computed by:

$$Design\ Effect = 1 + (n_c - 1)ICC;$$

where n_c is average number of observation per student

Note that some researchers believe that design effect estimates greater than 2.0 indicate a need for multilevel modeling (Muthen, 1991, 1994), however, it was not set in stone (Peugh, 2010).

Furthermore, the *proportion of variance accounted for* each variable was computed to report the size of the impacts of the SWH approach on students' science achievement. Tymms (2004) argued that the measures of *proportion of variance accounted for* and the *intra-class correlation* combine with a direct interpretation of the coefficients are enough to provide a clear picture of the size of the impact. However, since there are also growing numbers of interest in the use of effect size as found in

traditional experimental design, the standardized effect size (ES) was also calculated by the following formula (Raudenbush & Xiao-Feng, 2001);

$$ES = \beta_p / \sqrt{\tau_{pp}}$$

where β_p is the group difference on polynomial trend p and $\sqrt{\tau_{pp}}$ is the population standard deviation of the polynomial trend of interest. Cohen (1988) suggested the values of effect size (Cohen's f) of 0.10, 0.25, and 0.40 represent small, medium, and large effect size, respectively.

Finally, to answer that questions of how students in the participating school district performed in their science ITBS test comparing to all students in Iowa, descriptive statistic of Iowa norms were developed, including sub-populations. Achievement gaps were calculated based on their predicted achievement scores. The Iowa norms were compared to predicted scores, and predicted achievement gap of the participating school district. Graphs were created to help visualize the comparison.

Furthermore, to answer the third research question whether the grade level that students started exposing to the SWH approach had impact on their science ITBS achievement (research question 3); first, level-one model was tested and developed from the following variables; GRADE, GRADE². Second, StartLev was added into level-two model to test its impact. Finally, StartLev was added into level-two model to test its impact again but this time GENDER, IEP, SES, and InREAD, were add into level-two model as covariate (grand-center).

Level-1 model: $SCIENCES_{ti} = \pi_{0i} + \pi_{1i}*(GRADE_{ti}) + \pi_{2i}*(GRADE^2_{ti}) + e_{ti}$

Level-2 model: $\pi_{0i} = \beta_{00} + \beta_{01}*(GENDER_i) + \beta_{02}*(IEP_i) + \beta_{03}*(SES_i) + \beta_{04}*(StartLev_i) + \beta_{05}*(InREAD_i) + r_{0i}$

$$\begin{aligned}\pi_{1i} &= \beta_{10} + \beta_{11}*(GENDER_i) + \beta_{12}*(IEP_i) + \beta_{13}*(SES_i) \\ &+ \beta_{14}*(StartLev_i) + \beta_{15}*(InREAD_i) \\ \pi_{2i} &= \beta_{20} + \beta_{21}*(GENDER_i) + \beta_{22}*(IEP_i) + \beta_{23}*(SES_i) \\ &+ \beta_{24}*(StartLev_i) + \beta_{25}*(InREAD_i)\end{aligned}$$

To evaluate the validity and reliability of multilevel models, after multilevel models for change were fitted, the residual files for level-1 and level-2 models were produced to check the fit and distributional assumptions of the models. Moreover, the intraclass correlation coefficient (ICC) and the design effect were used to confirm the need for using a multilevel modeling from unconditional means model (Peugh, 2010). To analyze the size of the impacts, the *proportion of variance accounted for* each variable and the standardized effect size (ES) (Raudenbush & Xiao-Feng, 2001) was computed. Finally, to answer that questions of how students in the participating school district performed in their science ITBS test comparing to all students in Iowa, descriptive statistic of Iowa norms were developed, including sub-populations. Achievement gaps were calculated based on their predicted achievement scores. The Iowa norms were compared to predicted scores, and predicted achievement gap of the participating school district. Graphs were created to help visualize the comparison.

Summary

To explore the longitudinal impacts of the SWH approach, this study utilized a quantitative approach with the multilevel model for change technique as the methodology. This chapter elaborated the context of the study and rationale for the use of the multilevel model as the tool to determine student achievement growth. The details of

variables and analysis procedure were provided. The next chapter will present statistical results and their interpretation.

CHAPTER IV

RESULTS

To address the research questions, the results are presented in 4 sections. Section 1: Data Exploration, explores the relationship among parameters using Pearson correlation coefficient (r) and Chi Square test of independence (χ^2). This section provides general ideas of how one variable changes when the value of the other variable changes. Section 2: Science ITBS Achievement of Iowa Students, shows average achievement trajectories of all Iowa students during the same period of this study. Average achievement of IEP, low-SES, and males and females students are also presented. Section 2 serves as background information in model building processes and its interpretation in later section. Moreover, section 3: Science Achievement Model Building, demonstrates multi-model building process; level-1, level-2, and final model, explaining science achievement scores and its interpretation. Logics and results of each step in the process of building final model are elaborated in great details. The predicted science achievement trajectories resulting from each model are also compared with average Iowa ITBS science scores. Finally, Section 4 is The Impact of Starting Time focusing on results and their interpretations of models exploring the impact of ages that students first started to expose to the SWH approach.

Data Exploration

First, Pearson's correlation coefficients were computed to determine the relationship between science and reading achievement scores. The result suggested a

strong positive correlation between science and reading achievement scores, $r = 0.864$, $n = 11951$, $p < .001$.

Second, the Chi-square test of independence was used to determine whether there were significant associations between two categorical variables; GENDER, IEP, SES, and InREAD. The results were presented in Table 4.1.

Was there any association between a student's gender and individual's IEP status?

There was a significant association between a student's gender and IEP status, $\chi^2 (1) = 32.59$, $p < .001$. Male students were more likely to be identified with an IEP (64.1%) than female students.

Was there any association between a student's gender and SES?

There was no significant association between a student's gender and SES status, $\chi^2 (1) = 0.00$, $p = .994$. There was approximately the same percentage of male (51%) and female (49%) identified with low SES,

Was there any association between a student's IEP and SES?

There was a significant association between a student's SES and IEP status, $\chi^2 (1) = 148.20$, $p < .001$. Students who were from low-SES families were more likely to be identified with an IEP (51.2%). Students who were not from low SES family were less likely to be identified with an IEP (48.8%).

Was there any association between a student's gender and initial reading achievement level?

There was a significant association between a student's gender and initial reading achievement level, $\chi^2 (3) = 9.89$, $p = .020$. Male students were more likely to have low

initial reading achievement level (55.8%) compared to female students (44.2%). In contrast, female students were more likely to have proficient initial reading achievement level (51.1%) compared to male students (48.9%).

Table 4.1 Results of Chi Square Test of Independence Showing Number of Student and Standardized Residuals

		Gender		SES		IEP	
		Male	Female	non-Low	Low	No	Yes
SES	non-Low	1205	1160				
		(0)	(0)				
	Low	435	419				
		(0)	(0)				
IEP	No	1376	1431	2164	643		
		(-1.4)	(1.5)	(2.2 ^a)	(-3.7 ^b)		
	Yes	264	148	202	212		
		(3.7 ^b)	(-3.8 ^b)	(-5.9 ^b)	(9.7 ^b)		
InREAD	Low	236	187	239	185	213	211
		(1.4)	(-1.4)	(-4.1 ^b)	(6.8 ^b)	(-8.1 ^b)	(21.2 ^b)
	Medium	206	171	244	133	289	88
		(1.0)	(-1.0)	(-2.0 ^a)	(3.3 ^b)	(-2.2 ^a)	(5.7 ^b)
	Proficient	966	1011	1498	480	1870	108
		(-1.3)	(1.3)	(1.2)	(-2.0 ^a)	(3.5 ^b)	(-9.2 ^b)
	High	232	210	385	57	435	7
(0.5)		(-0.5)	(3.3 ^b)	(-5.6 ^b)	(2.5 ^a)	(-6.6 ^b)	

Note. ^a Standardized residuals is significant at 0.05 level

^b Standardized residuals is significant at 0.01 level

Was there any association between a student's IEP status and initial reading achievement level?

There was a significant association between a student's IEP status and initial reading achievement level, $\chi^2(3) = 699.39, p < .001$. Students with IEP were more likely to have low (49.8%) and medium (23.3%) initial reading achievement level compared to non-IEP students (50.2% and 76.7%, respectively) and less likely to have proficient

(5.5%) and high (1.6%) initial reading achievement level. In contrast, students with non-IEP were more likely to have proficient (94.5%) and high (98.4%) initial reading achievement level compared to IEP students (5.5% and 1.6%, respectively).

Was there any association between a student's SES and initial reading achievement level?

There was a significant association between a student's SES and initial reading achievement level, $\chi^2(3) = 125.73, p < .001$. Students from low SES families were more likely to have low (43.6%) and medium (35.3%) initial reading achievement level compared to students from non-low SES families (56.4% and 64.7%, respectively) and less likely to have proficient (24.3%) and high (12.9%) initial reading achievement level. In contrast, students from non-low income families were more likely to have proficient (75.7%) and high (87.1%) initial reading achievement level compared to students from low SES families (24.3% and 12.9%, respectively).

Science ITBS Achievement of Iowa Students

The mid-year science ITBS scores were collected and composited from 2001-2011. The overall weighted average science ITBS scores and sub-population have been computed and presented in Table 4.2. In general, the Iowa normative science ITBS average scores showed non-linear growth. The gain scores were higher at younger grades but gradually decreased at higher grades. For example, the average science ITBS score at fourth grade was 209.31, an increase of 19.96 (209.31-189.35) points over the third grade average score. While, the average science ITBS score at tenth grade was 277.29, an increase of 5.47 (277.29-271.82) points over the ninth grade average score.

Though, the growth trajectory of the Iowa normative population was not linear, the growth trajectory was very smooth. However, there was observable interrupted of the smoothness of the growth trajectory in grade ten where the averaged science score was lower than it ought to be. This could be explained by the fact that the ITBS tests are not mandatory tests for schools, the tests are administrated to certain voluntary schools. Thus, the population of students who took the tests at tenth grade was changed, as a result, the average science score was changed as well.

Table 4.2 Weighted average science ITBS and gain scores of Iowa students from 2001-2011

Grade	3	4	5	6	7	8	9	10	11
Iowa	189.35	209.31	227.31	239.1	254.52	265.08	271.82	277.29	293.5
<i>Gain score</i>		19.96	18	11.79	15.42	10.56	6.74	5.47	16.21
Male	189.74	210.54	229.05	240.48	255.52	265.74	270.76	275.99	291.97
<i>Gain score</i>		20.8	18.51	11.43	15.04	10.22	5.02	5.23	15.98
Female	188.96	208.03	225.48	237.65	253.49	264.39	272.94	278.66	295.1
<i>Gain score</i>		19.07	17.45	12.17	15.84	10.9	8.55	5.72	16.44
Non lowSES	193.24	214.13	233.07	245.45	260.84	271.36	278.96	283.51	298.69
<i>Gain score</i>		20.89	18.94	12.38	15.39	10.52	7.6	4.55	15.18
Low SES	181.92	199.75	215.95	225.92	240.33	250.27	254.86	261.49	277.03
<i>Gain score</i>		17.83	16.2	9.97	14.41	9.94	4.59	6.63	15.54
Non IEP	190.98	211.73	230.58	243.15	259.2	270.3	277.5	283.08	299.32
<i>Gain score</i>		20.75	18.85	12.57	16.05	11.1	7.2	5.58	16.24
IEP	175.7	191.63	204.9	211.92	223.15	231.05	233.74	238.1	249.37
<i>Gain score</i>		15.93	13.27	7.02	11.23	7.9	2.69	4.36	11.27

The same non-linear growth trends were found across all sub-populations. The biggest gains were found at elementary school, then the gain generally decreased as students moved to higher grades. Male and female students had similar gain scores across years; however, male students tended to have bigger gains at younger ages. Then female students started to have bigger gain than boys at sixth grade. The science achievement gap between male and female students was found to be small at third grade when males scored higher than females. The achievement gap of male student increased reaching the largest gap at fifth grade (3.57 points) then the gap narrowed. At ninth grade the achievement gap was reversed, girls had better science achievement than boys, a trend that continued to eleventh grade.

Table 4.3 Average science ITBS achievement gaps of Iowa students from 2001-2011

Grade	3	4	5	6	7	8	9	10	11
Male	189.74	210.54	229.05	240.48	255.52	265.74	270.76	275.99	291.97
Female	188.96	208.03	225.48	237.65	253.49	264.39	272.94	278.66	295.1
<i>Gender gap</i>	<i>0.78</i>	<i>2.51</i>	<i>3.57</i>	<i>2.83</i>	<i>2.03</i>	<i>1.35</i>	<i>-2.18</i>	<i>-2.67</i>	<i>-3.13</i>
NonlowSES	193.24	214.13	233.07	245.45	260.84	271.36	278.96	283.51	298.69
Low SES	181.92	199.75	215.95	225.92	240.33	250.27	254.86	261.49	277.03
<i>SES gap</i>	<i>11.32</i>	<i>14.38</i>	<i>17.12</i>	<i>19.53</i>	<i>20.51</i>	<i>21.09</i>	<i>24.1</i>	<i>22.02</i>	<i>21.66</i>
Non IEP	190.98	211.73	230.58	243.15	259.20	270.30	277.5	283.08	299.32
IEP	175.7	191.63	204.9	211.92	223.15	231.05	233.74	238.1	249.37
<i>IEP gap</i>	<i>15.28</i>	<i>20.1</i>	<i>25.68</i>	<i>31.23</i>	<i>36.05</i>	<i>39.25</i>	<i>43.76</i>	<i>44.98</i>	<i>49.95</i>

The same non-linear growth can be found in low-SES and IEP students as well.

The year-to-year gain scores of these disadvantaged groups were much smaller than other

students. This phenomenon gradually created wider achievement gaps as students moved to older grade levels. The science achievement gap between low-SES students and non-low-SES students was clearly existing in third grade where students from low-income family had science ITBS score 11.32 points lower than other students. The achievement gap became wider as students moved to higher grades. The science achievement gap reached its highest point at ninth grade where low-SES students' average science score was 24.1 points lower than other students.

The science achievement gap between students with and without an IEP found in ITBS test was the largest compared to achievement gap between genders or SES status. At third grade, students with an IEP clearly disadvantaged compared to non-IEP students, with 15.28 points lower. The IEP achievement gap was increasing as students moved to higher grade levels. At eleventh grade, students with an IEP scored at 249.37, 49.95 points lower than non-IEP students.

Science Achievement Model Building

This section explains the model building procedures and their interpretation. The model building processes started with the unconditional model where there were no predictors included. Then level-1 predictors were added and then evaluated. After level-1 model was established, each level-2 variable was added to explore impacts. Finally, the final model is developed.

Building Leve-1 Model

Model A: Unconditional Means Model

The unconditional means model describes the change in each student's science achievement scores over time as a flat line with a slope of zero. The unconditional equation shown in Equation 4.1 included only an intercept estimate; there were no predictor variables.

$$\begin{aligned}
 \text{Level-1 Model:} \quad & \text{SCIENCE}_{ti} = \pi_{0i} + e_{ti} \\
 \text{Level-2 Model:} \quad & \pi_{0i} = \beta_{00} + r_{0i} \\
 \text{Mixed Model:} \quad & \text{SCIENCE}_{ti} = \beta_{00} + r_{0i} + e_{ti} \qquad (4.1)
 \end{aligned}$$

The estimated parameters are provided in Table 4.4. The results showed a significant grand-mean science achievement score, $\beta_{00} = 260.59$, $p < .001$. The level-1 variance component showed differences between each student's observed and predicted science achievement scores over time, $\sigma^2 = 837.31$. Students' mean science scores (the average score across all collected assessments) were significantly varied around the grand mean, $\tau_{00} = 1385.87$, $p < .001$. Further, the intraclass correlation (ICC) calculations showed that 62% ($1385.87/[1385.87+837.31] = 0.62$) of science achievement variation occurred across students. The design effect statistic also computed (Design Effect = $1 + [3.84-1]0.62 = 2.76$). The value of ICC coefficient over 0.40 and the value of design effect over 2.0 are indicative of the appropriateness of multilevel model (Spybrook, Raudenbush, Liu, Congdon, & Martinez, 2008)

Table 4.4 Results of Level-1 Model Building

			Model					
			A	B	C	D	E	F
			Fixed Effect					
Composite model	Intercept	B_{00}	^a 260.59	^a 202.72	^a 192.91	^a 189.57	^a 193.11	^a 190.82
		SE	0.72	0.72	0.74	1.32	0.68	1.34
Grade	Intercept	B_{10}		^a 12.21	^a 18.26	^a 18.57	^a 17.85	^a 18.08
		SE		0.14	0.36	0.36	0.35	0.36
Grade ²	Intercept	B_{20}			^a -0.67	^a -0.72	^a -0.70	^a -0.72
		SE			0.04	0.04	0.04	0.04
SWH	Intercept	B_{30}				^a 3.72		^a 2.56
		SE				1.14		1.25
ConSWH	Intercept	B_{40}					^a 1.07	^a 0.86
		SE					0.27	0.30
			Random Effect					
Random Effect	Intercept	r_0	1385.87	853.88	861.77	1129.67	784.48	1130.78
		SWH				700.47		840.46
		ConSWH					20.45	18.00
		level-1	σ^2	837.31	438.70	423.46	376.75	383.68
Deviance Estimate		124381	116844	116532	116155	116325	115918	
parameter		2	2	2	4	4	7	

Note. ^a significant at 0.01 level ^b significant at 0.05 level

Model B: Linear Growth Model

Adding GRADE as a level-1 time predictor to the unconditional means model allowed the changes in each student's science achievement score over time to be modeled with a straight line with a non-zero slope. Since the science achievement scores were collected from third grade, the GRADE variable was centered at third grade. Centering at

third grade allows for the interpreting of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCE}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}-3) + e_{ti}$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10}$$

$$\text{Mixed Model: } \text{SCIENCE}_{ti} = \beta_{00} + \beta_{10} * \text{GRADE}-3_{ti} + r_{0i} + e_{ti} \quad (4.2)$$

The multilevel model was estimated and the results are presented in the second column of Table 4.4. The results demonstrated a significant grand-mean science achievement score at grade 3 ($\beta_{00} = 202.75, p < .001$) that was increasing 12.21 points per grade level ($\beta_{10} = 12.21, p < .001$). Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 853.88, p < .001$). The level-1 variance component showed differences between each student's observed and predicted science achievement scores over time, $\sigma^2 = 438.69$. By adding grade as a linear time variable to the level-1 model, the level-1 residual variance was decreased about $([837.31-438.69]/837.31)$ 48%. Consequently, using time-varying-level GRADE as a predictor of science achievement reduced the within students variance by 48%.

The difference in deviance statistics between unconditional means model (Model A) and Model B, $124380.59-116843.88=7536.71$, is distributed as a Chi-square with one degree of freedom. The likelihood ratio test was significant, $\chi^2(1) = 7536.71, p < .001$, demonstrating that predicting science achievement scores with time-varying grade (Model B) was significantly better fit to the data than predicting science achievement scores with the unconditional means model (Model A).

Model C: Quadratic Growth Model

The effect of quadratic growth was tested in Model C by adding $GRADE^2$ to the level-1 model as a level-1 variable presenting in Equation 4.3. Since the science achievement scores were collected from third grade and to be consistent with Model B, the $GRADE$ and $GRADE^2$ variables were centered at third grade to allow interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } SCIENCE_{ti} = \pi_{0i} + \pi_{1i}*(GRADE-3_{ti}) + \pi_{2i}*(GRADE-3)_{ti}^2 + e_{ti}$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + r_{0i}$$

$$\pi_{1i} = \beta_{10}$$

$$\pi_{2i} = \beta_{20}$$

$$\text{Mixed Model: } SCIENCE_{ti} = \beta_{00} + \beta_{10}*GRADE-3_{ti} + \beta_{20}*(GRADE-3)_{ti}^2 + r_{0i} + e_{ti} \quad (4.3)$$

The multilevel model was estimated results presenting in Table 4.4. The results demonstrated a significant grand-mean science achievement score at grade 3 ($\beta_{00} = 192.91$, $p < .001$). In quadratic growth model, there is no constant common slope. The β_{10} represents the instantaneous rate of change at one specific moment, when $GRADE-3 = 0$. The results showed that at grade three the rate of change was significant ($\beta_{10} = 18.26$, $p < .001$). In addition, the β_{20} represents the curvature parameter explaining the changing rate of change. The results showed that the curvature parameter was also significant ($\beta_{20} = -0.67$, $p < .001$). Because β_{10} was positive, the trajectory initially increased by 18.26 at third grade; however, β_{20} was negative indicating that the magnitude of gain scores diminished over time. Further, the variance component estimate showed a significant variation in science achievement scores at grade 3 ($\tau_{00} = 861.76$, $p < .001$). The level-1 variance component showed differences between each student's observed and predicted

science achievement scores over time, $\sigma^2 = 423.46$. By adding the quadratic term of GRADE to the level-1 model, the level-1 residual variance decreased by $([438.69 - 423.46] / 438.69) 3\%$, reducing the within student variance by 3%.

The difference in deviance statistics between Model B and Model C, $116843.88 - 116532.4 = 311.45$, is distributed as a Chi-square with one degree of freedom. The likelihood ratio test was significant, $\chi^2(1) = 311.45, p < .001$, indicating that the science achievement trajectories were better predicted with the quadratic change trajectory when compared to linear-change trajectory.

Model D: Discontinuity in Elevation, not Slope

Model D aimed to test the immediate shift in elevation but no shift in slope. That is, the elevation of science achievement scores trajectory jumps after students were exposed to the SWH approach but the slope and the change of slope before and after exposing to the SWH approach remains the same. The SWH was introduced to the model as a random effect allowing the effect of exposure to the SWH approach to vary between students. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were centered at third grade to allow interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCE}_{ii} = \pi_{0i} + \pi_{1i} * (\text{GRADE} - 3_{ii}) + \pi_{2i} * (\text{GRADE} - 3_{ii})^2 + \pi_{3i} * (\text{SWH}_{ii}) + e_{ii}$$

$$\begin{aligned} \text{Level-2 Model: } \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} \\ \pi_{2i} &= \beta_{20} \\ \pi_{3i} &= \beta_{30} + r_{3i} \end{aligned}$$

Mixed Model:
$$SCIENCE_{ti} = \beta_{00} + \beta_{10}*(GRADE-3_{ti}) + \beta_{20}*(GRADE-3_{ti})^2 + \beta_{30}*SWH_{ti} + r_{0i} + r_{3i}*SWH_{ti} + e_{ti} \quad (4.4)$$

The multilevel model was estimated with the results presented in Table 4.4. The grand-mean science achievement score at grade 3 was significant ($\beta_{00} = 189.57, p < .001$); the rate of change was significant ($\beta_{10} = 18.57, p < .001$); and the curvature parameter was also significant ($\beta_{20} = -0.72, p < .001$). β_{10} was positive indicating that the trajectory initially increased by 18.57 at third grade. Over time, because β_{20} is negative, the slope diminished. The elevation of science achievement associated with exposure to the SWH approach was significant ($\beta_{30} = 3.72, p = .001$), a small effect (Cohen's $f = 0.19 = 3.72/\sqrt{376.75}$).

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 1129.67, p < .01$) and significant slope variance ($\tau_{11} = 700.47, p < .001$). The level-1 variance component showed differences between each student's observed and predicted science achievement scores over time, $\sigma^2 = 376.75$. By adding SWH exposure to the level-1 model, the level-1 residual variance decreased by $([423.46-376.75]/ 423.46) 11\%$.

The difference in deviance statistics between Model C and Model D, $116532.4 - 116154.9 = 377.5$, is distributed as a chi-square with three degrees of freedom. The likelihood ratio test was significant, $\chi^2(3) = 377.5, p < .001$, demonstrating that the science achievement trajectories were better predicted with SWH exposure in the model.

Model E: Discontinuity in Slope, not Elevation

Model E tested the shift in slope after exposure to the SWH approach without a corresponding shift in elevation. This means that the elevation of science achievement scores hypothesized not to change before and after exposure to the SWH approach, while the slope and the change of slope after exposure to the SWH approach was hypothesized to change. CONSWH was introduced to the model as a random effect allowing the effect of consecutive years of exposure to the SWH approach to be varied between students. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² were centered at third grade to allow interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \quad \text{SCIENCE}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}-3_{ti}) + \pi_{2i} * (\text{GRADE}-3_{ti})^2 + \pi_{3i} * (\text{CONSWH}_{ti}) + e_{ti}$$

$$\begin{aligned} \text{Level-2 Model: } \quad \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} \\ \pi_{2i} &= \beta_{20} \\ \pi_{3i} &= \beta_{30} + r_{3i} \end{aligned}$$

$$\text{Mixed Model: } \quad \text{SCIENCE}_{ti} = \beta_{00} + \beta_{10} * (\text{GRADE}-3_{ti}) + \beta_{20} * (\text{GRADE}-3_{ti})^2 + \beta_{30} * \text{CONSWH}_{ti} + r_{0i} + r_{3i} * \text{CONSWH}_{ti} + e_{ti} \quad (4.5)$$

The multilevel model was estimated and the results were presented Table 4.4. The grand-mean science achievement score at grade 3 was significant ($\beta_{00} = 193.11, p < .001$); the rate of change was significant ($\beta_{10} = 17.85, p < .001$); the curvature parameter was significant ($\beta_{20} = -0.70, p < .001$). Because β_{10} was positive, the trajectory initially rose by 17.85 at third grade. β_{20} was negative indicating that with each subsequent grade, the slope diminished by 0.70 points. The shift of slope associated with the SWH exposure

was significant ($\beta_{30} = 1.07, p < .001$). After students experienced the SWH approach, rate of change in science scores increased an addition 1.07 points compared to before exposure to the SWH approach although the effect was trivial (Cohen's $f = 0.05 = 1.07/\sqrt{383.68}$).

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 784.48, p < .001$) and significant slope variance ($\tau_{11} = 20.45, p < .001$). The level-1 variance component showed differences between each student's observed and predicted science achievement scores over time, $\sigma^2 = 383.68$. By adding the CONSWH to the level-1 model, the level-1 residual variance decreased by $([423.46-383.68]/ 423.46)$ 9%.

The difference in deviance statistics between Model C and Model E, $116532.4-116325.1 = 207.4$, is distributed as a chi-square with three degree of freedom. The likelihood ratio test was significant, $\chi^2(3) = 207.4, p < .01$, indicating that the science achievement trajectories were better predicted by adding CONSWH, the consecutive exposure to the SWH approach to the model.

Model F: Discontinuity in both Elevation and Slope

Model F tested the shift in both trajectory elevation and slope simultaneously after students were exposed to the SWH approach. Both SWH and CONSWH were introduced to the model as random effects allowing the effect of exposure to the SWH approach to be varied between students. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were centered at third grade.

Level-1 Model:
$$SCIENCE_{ti} = \pi_{0i} + \pi_{1i}*(GRADE-3_{ti}) + \pi_{2i}*(GRADE-3_{ti})^2 + \pi_{3i}*(SWH_{ti}) + \pi_{4i}*(CONSWH_{ti}) + e_{ti}$$

Level-2 Model:
$$\begin{aligned} \pi_{0i} &= \beta_{00} + r_{0i} \\ \pi_{1i} &= \beta_{10} \\ \pi_{2i} &= \beta_{20} \\ \pi_{3i} &= \beta_{30} + r_{3i} \\ \pi_{4i} &= \beta_{40} + r_{4i} \end{aligned}$$

Mixed Model:
$$SCIENCE_{ti} = \beta_{00} + \beta_{10}*GRADE-3_{ti} + \beta_{20}*GRADE-3_{ti}^2 + \beta_{30}*SWH_{ti} + \beta_{40}*CONSWH_{ti} + r_{0i} + r_{3i}*SWH_{ti} + r_{4i}*CONSWH_{ti} + e_{ti} \quad (4.6)$$

The multilevel model was estimated with the results presenting in Table 4.4. The grand-mean science achievement score at grade 3 was significant ($\beta_{00} = 190.82, p < .001$); the rate of change was significant ($\beta_{10} = 18.08, p < .001$); and the curvature parameter was also significant ($\beta_{20} = -0.72, p < .001$). Because β_{10} was positive, the trajectory rose by 18.08 points per grade level; however, the negative curvature parameter (β_{20}) indicated that the slope diminished with each additional grade. The elevation of science achievement scores of 2.56 points was also significant ($\beta_{30} = 2.56, p < .001$); regardless of student's individual characteristics, after students experienced the SWH approach, their science achievement scores improved by 2.56 points, a small effect (Cohen's $f = 0.14 = 2.56/\sqrt{342.32}$). The shift in slope of 0.86 points after students exposed to the SWH approach was also significant ($\beta_{40} = 0.86, p < .001$) although this effect was trivial (Cohen's $f = 0.05 = 0.86/\sqrt{342.32}$). Figure 4.1 illustrates the comparison between predicted science achievement trajectories before and after exposure to the SWH approach. Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 1130.74, p < .001$). The level-1 variance component showed differences between each student's observed and predicted science achievement scores over time, $\sigma^2 = 342.32$. The variance component calculations showed

that 48% ($1130.78/[1130.74+840.46+18.00+342.32]=0.48$) of science achievement variation occurred across students; 36% ($840.45/[1130.74+840.46+18.00+342.32]=0.36$) of science achievement variation occurred within exposure to the SWH approach; only 0.7% ($18.00/[1130.74+840.46+18.00+342.32]=0.007$) of science achievement variation occurred for consecutive exposure to the SWH approach; and 14% of science achievement variation occurred across a grade level. By adding SWH and CONSWH to the level-1 model, the level-1 residual variance decreased by ($[423.46-342.32]/423.46$) 19%, indicating that adding SWH and CONSWH with both shift in slope and elevation as predictors of science achievement scores reduced the within student variance by 19%.

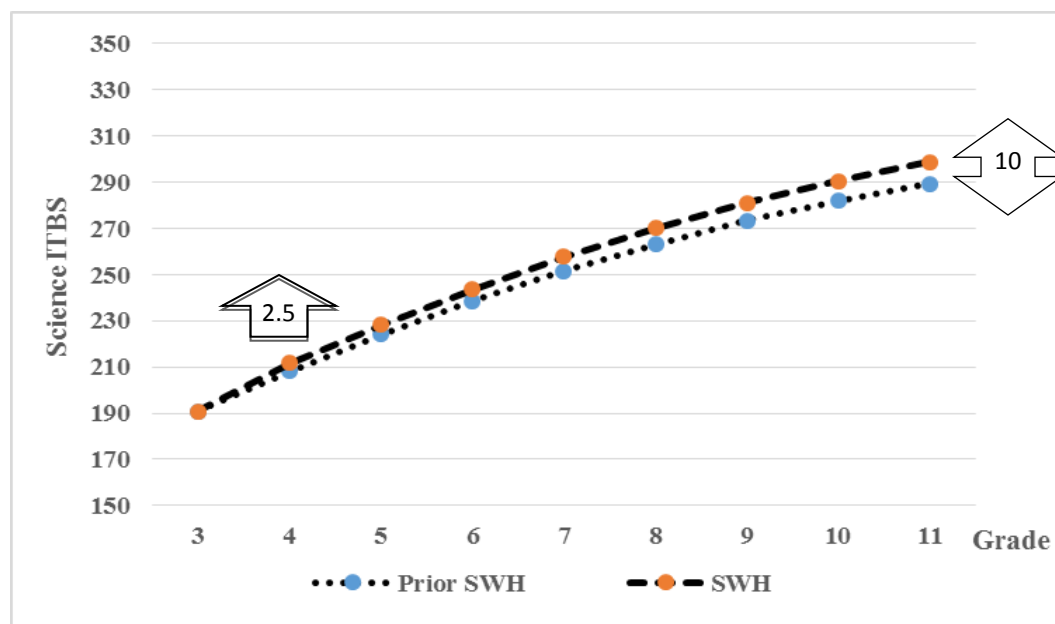
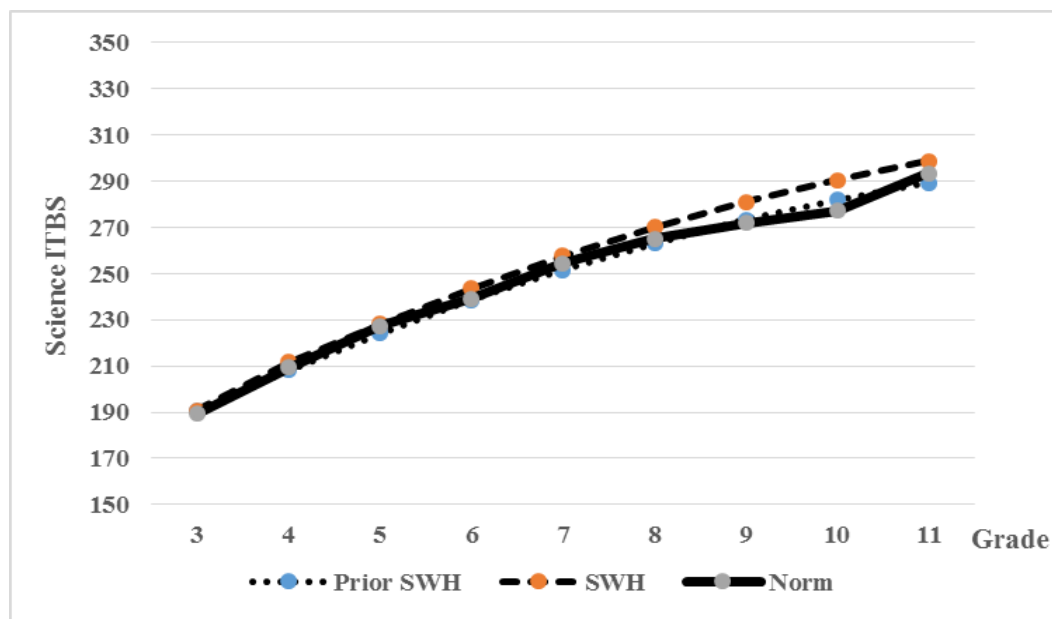


Figure 4.1 Students' Science Achievement Trajectories from Grade 3 to Grade 11, Before and After Exposure the SWH Approach

The difference in deviance statistics between Model C and Model F, $116532.4 - 115917.7 = 614.7$, is distributed as a chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(5) = 207.4, p < .01$, demonstrated that the science achievement score trajectory could be better predicted by adding SWH and CONSWH. The decreased deviance statistics by simultaneously testing both SWH and CONSWH (Model F) indicated that Model F was the best fitting model. Thus, Model F consisting of GRADE, GRADE², SWH, and CONSWH was set as the foundation level-1 model. By adding all level-1 predictors together, the model explained $([837.31 - 342.32] / 837.31)$ 59% of the level-1 residual variance.

Summary of Level-1 Model

When compare the predicted science achievement scores from the multilevel model to average scores of Iowa students, the predicted achievement growth trajectories were non-linear, consistent with the Iowa normative achievement growth trajectory described in Section 2. The predicted achievement trajectory of students before experiencing any SWH was similar to the normative of Iowa students, while the predicted achievement trajectory of students after experiencing the SWH approach was higher. The gain score analysis (Figure 4.2) showed that the predicted gain scores were higher at younger grades and then gradually decreased as students moved to higher grades. Moreover, the predicted gain scores of students who had experience in the SWH classrooms were consistently higher than predicted scores for students in traditional classrooms.



Grade	3	4	5	6	7	8	9	10	11
Prior SWH	190.83	208.19	224.1	238.57	251.59	263.16	273.3	281.98	289.38
Gain score		17.36	15.91	14.47	13.02	11.57	10.14	8.68	7.4
SWH	190.83	211.61	228.39	243.72	257.61	270.05	281.05	290.6	298.7
Gain score		20.78	16.78	15.33	13.89	12.44	11	9.55	8.1
Norm	189.35	209.31	227.31	239.1	254.52	265.08	271.82	277.29	293.5
Gain score		19.96	18	11.79	15.42	10.56	6.74	5.47	16.21

Figure 4.2 Students' Science Achievement Trajectories from Grade 3 to Grade 11 before and after Exposure to the SWH Approach Compared to Iowa Normative Achievement.

Building Level-2 Models

Model G: Effect of Gender

Building from Model F, a binary level-2 predictor variable, gender (i.e., female =1, male = 0) was added to the level-2 model to explain the intercept and slope variance in science achievement score. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² were centered at third grade to allow interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}-3_{ti}) + \pi_{2i} * (\text{GRADE}-3_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

$$\begin{aligned} \text{Level-2 Model: } \pi_{0i} &= \beta_{00} + \beta_{01} * (\text{GENDER}_i) + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} * (\text{GENDER}_i) \\ \pi_{2i} &= \beta_{20} + \beta_{21} * (\text{GENDER}_i) \\ \pi_{3i} &= \beta_{30} + \beta_{31} * (\text{GENDER}_i) + r_{3i} \\ \pi_{4i} &= \beta_{40} + \beta_{41} * (\text{GENDER}_i) + r_{4i} \end{aligned}$$

$$\begin{aligned} \text{Mixed Model: } \text{SCIENCES}_{ti} &= \beta_{00} + \beta_{01} * \text{GENDER}_i + \beta_{10} * \text{GRADE}-3_{ti} + \\ &\beta_{11} * \text{GENDER}_i * \text{GRADE}-3_{ti} + \beta_{20} * \text{GRADE}-3_{ti}^2 + \beta_{21} * \text{GENDER}_i * \\ &\text{GRADE}-3_{ti}^2 + \beta_{30} * \text{SWH}_{ti} + \beta_{31} * \text{GENDER}_i * \text{SWH}_{ti} + \\ &\beta_{40} * \text{CONSWH}_{ti} + \beta_{41} * \text{GENDER}_i * \text{CONSWH}_{ti} + r_{0i} + r_{3i} * \text{SWH}_{ti} + \\ &r_{4i} * \text{CONSWH}_{ti} + e_{ti} \end{aligned} \quad (4.7)$$

The multilevel model was estimated with results presenting in Table 4.5. In order to allow easier interpretation of results, the value of gender (0 = male, 1 = female) was substituted into Equation 4.7.

$$\begin{aligned} \text{Males : } \text{SCIENCES}_{ti} &= 189.67 + 19.07 * \text{GRADE}-3_{ti} - 0.83 * \text{GRADE}-3_{ti}^2 + \\ &3.63 * \text{SWH}_{ti} + 0.64 * \text{CONSWH}_{ti} \end{aligned} \quad (4.8)$$

Females :
$$\text{SCIENCES}_{i} = 191.99 + 17.08 * \text{GRADE-3}_{i} - 0.61 * \text{GRADE-3}_{i}^2 + 1.47 * \text{SWH}_{i} + 1.07 * \text{CONSWH2}_{i} \quad (4.9)$$

Male. The grand-mean science achievement score for male students in grade 3 was significant ($\beta_{00} = 189.66, p < .001$). The results showed that at grade three both the rate of change for male student ($\beta_{10} = 19.07, p < .001$) and the curvature parameter were also significant ($\beta_{20} = -0.83, p < .001$). The β_{10} was positive indicated that the trajectory rose by 19.07 each successive grade; however, a negative β_{20} curvature parameter, indicated reduction in the average slope of 0.83 points at each grade. The elevation of science achievement scores for male student of 3.63 points was also significant ($\beta_{30} = 3.63, p = .046$). Male students who experienced the SWH approach had science achievement scores that were 3.63 points higher than their male peers in traditional classrooms. The shift in the slope of science achievement trajectory of 0.64 points exposure to the SWH was not significant ($\beta_{40} = 0.64, p = .137$).

Females. The results showed that female students' grand-mean science achievement score in grade 3 were not significant different compared to male students ($\beta_{01} = 2.33, p = 0.38$). At grade three, the rate of change for female students was significantly lower than male students ($\beta_{11} = -1.99, p = .005$) and the curvature parameter was significant higher ($\beta_{21} = 0.22, p = .008$). Female students' trajectory rose by 17.08 points (significantly lower than male students at third grade); however, their rated of change increased with each successive grade by 0.22 points.

Table 4.5 Results of Level-2 Models Building

			Model F	Model G	Model H	Model I	Model J	Model K
			Level-1	Gender	IEP	SES	InREAD	Simultaneous
			Fixed effect					
Composite model	Intercept	B_{00}	**190.82 (1.34)	**189.66 (1.96)	**193.59 (1.37)	**194.99 (1.47)	**130.50 (3.97)	**134.72 (5.07)
	Gender	B_{01}		2.33 (2.67)				-0.52 (2.29)
	IEP	B_{02}			** -29.51 (4.34)			-4.18 (4.57)
	SES	B_{03}				** -21.13 (3.71)		** -6.36 (3.20)
	InREAD	B_{04}					**21.53 (1.37)	**20.69 (1.58)
GRADE	Intercept	B_{10}	***18.08 (3.57)	**19.07 (0.51)	**19.08 (0.36)	**18.37 (0.41)	**7.29 (1.29)	**10.93 (1.53)
	Gender	B_{11}		** -1.99 (0.71)				** -2.38 (0.65)
	IEP	B_{12}			** -8.26 (1.19)			** -3.91 (1.23)
	SES	B_{13}				** -2.27 (0.85)		-0.38 (0.76)
	InREAD	B_{14}					**4.10 (0.43)	**3.39 (0.47)

Table 4.5 Continue

			Model F	Model G	Model H	Model I	Model J	Model K
			Level-1	Gender	IEP	SES	InREAD	Simultaneous
GRADE ²	Intercept	<i>B₂₀</i>	** -0.72 (0.04)	** -0.83 (0.06)	** -0.82 (1.27)	** -0.78 (0.05)	** 0.42 (0.15)	0.04 (0.18)
	Gender	<i>B₂₁</i>		** 0.22 (0.08)				** 0.23 (0.08)
	IEP	<i>B₂₂</i>			** 0.64 (0.14)			0.28 (0.15)
	SES	<i>B₂₃</i>				0.17 (0.10)		0.11 (0.10)
	InREAD	<i>B₂₄</i>					** -0.41 (0.05)	** -0.34 (0.06)
	Intercept	<i>B₃₀</i>	** 2.56 (1.25)	* 3.63 (1.82)	1.95 (1.27)	1.55 (1.33)	** 20.99 (3.73)	** 19.40 (4.48)
SWH	Gender	<i>B₃₁</i>		-2.16 (2.50)				-0.21 (2.22)
	IEP	<i>B₃₂</i>			** 12.82 (4.21)			4.07 (4.56)
	SES	<i>B₃₃</i>				** 10.73 (3.65)		2.96 (3.21)
	InREAD	<i>B₃₄</i>					** -6.67 (1.29)	** -6.24 (1.51)
	Intercept	<i>B₄₀</i>	** 0.86 (0.30)	0.64 (0.43)	** 0.86 (0.31)	** 1.45 (0.34)	-1.48 (0.97)	-1.42 (1.17)
ConSWH	Gender	<i>B₄₁</i>		0.43 (0.60)				0.79 (0.54)
	IEP	<i>B₄₂</i>			1.58 (0.90)			1.40 (0.98)
	SES	<i>B₄₃</i>				-0.11 (0.75)		-1.30 (0.67)
	InREAD	<i>B₄₄</i>					0.60 (0.33)	0.59 (0.36)

Table 4.5 Continue

		Model F	Model G	Model H	Model I	Model J	Model K
		Level-1	Gender	IEP	SES	InREAD	Simultaneous
		Random effect					
Intercept	r_0	1130.781	1130.74	928.04	1061.22	447.99	420.91
SWH	r_3	840.46	838.50	848.83	805.38	603.28	606.41
ConSWH	r_4	18.00	17.72	17.14	17.83	17.83	16.81
level-1	σ^2	342.32	342.35	343.32	342.63	344.53	345.74
Deviance		115917	115905	115378	115724	113798	113651
estimate parameters		7	7	7	7	7	7

Note. * significant at 0.05 level ** significant at 0.01 level

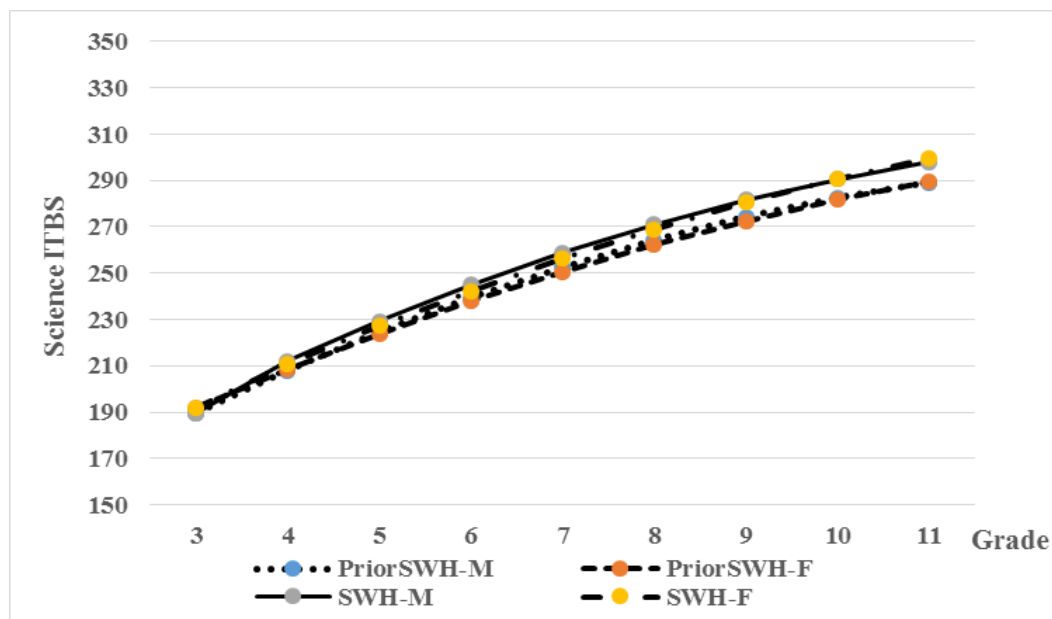


Figure 4.3 Science Achievement Trajectories before and after Exposure to the SWH Approach for Male and Female Students

Female students' science achievement trajectory demonstrated no significant difference in the shift of elevation ($\beta_{31} = -2.16, p = .386$) or slope ($\beta_{41} = 0.43, p = .473$) compared to male students' trajectories. Figure 4.3 illustrates the comparison between predicted science achievement trajectories before and after exposure to the SWH approach of male and female students.

The effect size analysis showed that when controlling for student gender, the standardized mean difference between before and after adopting the SWH approach was small (Cohen's $f = 0.20 = 3.63/\sqrt{342.35}$) and the change of slope was trivial (Cohen's $f = 0.03 = 0.64/\sqrt{342.35}$). The effect of gender on the change of elevation was also trivial

(Cohen's $f = 0.07 = 2.16/\sqrt{838.5}$), while the effect of gender on the change of slope was small (Cohen's $f = 0.10 = 2.16/\sqrt{18}$).

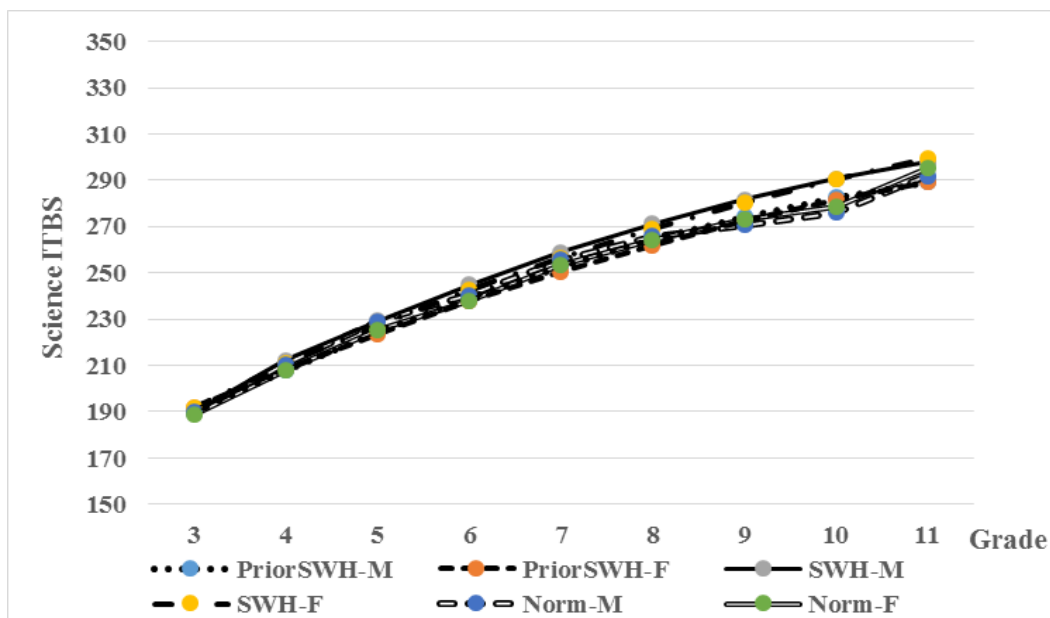
Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 1130.74$, $p < 0.001$). The level-1 variance component showing differences between each student's observed and predicted science achievement scores over time remained unchanged, $\sigma^2 = 342.35$. By adding gender to the level-2 model, the level-2 residual variance decreased approximately $([1989.24 - 1986.96]/1989.24) 0.1\%$. Gender contributed little to variation of science achievement score between grade levels.

The difference in deviance statistics between Model G and Model F, $119174.8 - 119146 = 28.8$, is distributed as a Chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(5) = 28.8$, $p < .01$, indicating that the science achievement trajectories were better predicted by adding the level-2 predictor; GENDER into Model F.

Gender Summary. The comparison between predicted science achievement and Iowa normative data revealed that the predicted science achievement growth trajectories of male and female students in this study were similar to these found in the Iowa normative population (see Figure 4.4). After experience the SWH approach, the participating students were expected to have greater achievement scores than before experience with the SWH approach. The analysis of the predicted achievement gap between genders also showed similar trends with Iowa's normative trend. In general, male students performed better than female students in elementary school, middle school

and early high school; however, the trend reversed at the end of high school.

Furthermore, the predicted achievement gaps in the participating school district were generally smaller than the achievement gap present in the Iowa normative population.



Grade	3	4	5	6	7	8	9	10	11
Prior SWH									
Male	189.67	207.9	224.48	239.4	252.67	264.27	274.22	282.5	289.1
Female	192	208.46	223.71	237.72	250.51	262.07	272.41	281.52	289.59
Gap	-2.33	-0.56	0.77	1.68	2.16	2.2	1.81	0.98	-0.49
SWH									
Male	189.67	212.18	229.4	244.96	258.87	271.11	281.7	290.63	297.85
Female	192	211.01	227.32	242.41	256.28	268.92	280.33	290.52	299.62
Gap	-2.33	1.17	2.08	2.55	2.59	2.19	1.37	0.11	-1.77
Norm									
Male	189.74	210.54	229.05	240.48	255.52	265.74	270.76	275.99	291.97
Female	188.96	208.03	225.48	237.65	253.49	264.39	272.94	278.66	295.1
Gap	0.78	2.51	3.57	2.83	2.03	1.35	-2.18	-2.67	-3.13

Figure 4.4 Science achievement scores trajectories before and after exposure to the SWH approach for male and female students compared to Iowa normative achievement

Model H: Effect of Individual Educational Program (IEP)

Building from Model F, a binary level-2 predictor variable, IEP (i.e., students who need individual education programs = 1, students who do not need individual education programs = 0) was added to the level-2 model to explain intercept and slope variance in science achievement scores. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² were centered at third grade to allow interpretation of the intercept as the average science achievement score for students in grade three.

Level-1 Model:
$$SCIENCES_{ti} = \pi_{0i} + \pi_{1i}*(GRADE-3_{ti}) + \pi_{2i}*(GRADE-3_{ti})^2 + \pi_{3i}*(SWH_{ti}) + \pi_{4i}*(CONSWH_{ti}) + e_{ti}$$

Level-2 Model:
$$\begin{aligned} \pi_{0i} &= \beta_{00} + \beta_{01}*(IEP_i) + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11}*(IEP_i) \\ \pi_{2i} &= \beta_{20} + \beta_{21}*(IEP_i) \\ \pi_{3i} &= \beta_{30} + \beta_{31}*(IEP_i) + r_{3i} \\ \pi_{4i} &= \beta_{40} + \beta_{41}*(IEP_i) + r_{4i} \end{aligned}$$

Mixed Model:
$$\begin{aligned} SCIENCES_{ti} &= \beta_{00} + \beta_{01}*IEP_i + \beta_{10}*GRADE-3_{ti} + \\ &\beta_{11}*IEP_i*GRADE-3_{ti} + \beta_{20}*GRADE-3_{ti}^2 + \beta_{21}*IEP_i*GRADE-3_{ti}^2 + \\ &\beta_{30}*SWH_{ti} + \beta_{31}*IEP_i*SWH_{ti} + \beta_{40}*CONSWH_{ti} + \\ &\beta_{41}*IEP_i*CONSWH_{ti} + r_{0i} + r_{3i}*SWH_{ti} + r_{4i}*CONSWH_{ti} + e_{ti} \end{aligned} \quad (4.10)$$

The multilevel model was estimated with results presenting in Table 4.4. In order to allow easier interpretation of results, the value of IEP (students who need individual education program =1, students who do not need individual education program = 0) was substituted into Equation 4.10.

Non-IEP student:
$$SCIENCES_{ti} = 193.59 + 19.08*GRADE-3_{ti} - 0.82*GRADE-3_{ti}^2 + 1.95*SWH_{ti} + 0.86*CONSWH_{ti} \quad (4.11)$$

IEP student:
$$SCIENCE_{i} = 164.08 + 10.81 * GRADE-3_{i} - 0.17 * GRADE-3_{i}^2 + 14.76 * SWH_{i} + 2.44 * CONSWH_{i} \quad (4.12)$$

Non-IEP Students. The results grand-mean science achievement score for Non-IEP students in grade 3 ($\beta_{00} = 193.59, p < .001$) was significant; the rate of change for Non-IEP student was significant ($\beta_{10} = 19.08, p < .001$); and the curvature parameter was significant ($\beta_{20} = -0.82, p < .001$). The rate of change for each successive grade, β_{10} , rose by 19.08 points. However, the negative curvature parameter indicated that the rate of change declines by 0.82 points per grade. The elevation of science achievement scores for non-IEP students was not significant ($\beta_{30} = 1.95, p = .125$); that is, after non-IEP students experienced the SWH approach, their science achievement scores increased by non-significant 1.95 points. The shift of the slope of science achievement of 0.86 points after non-IEP students' exposure to the SWH was significant ($\beta_{40} = 0.86, p = .006$).

IEP Students. The results showed that students with an IEP had significant lower grand-mean science achievement scores in grade 3 ($\beta_{01} = -29.51, p < .001$) when compared with non-IEP students. At grade three, the rate of change for IEP students was significant lower than non-IEP students ($\beta_{11} = -8.26, p < .001$) while, the curvature parameter was significantly higher ($\beta_{21} = 0.64, p < .001$). IEP students' trajectory rose by 10.81 points per grade level with additional acceleration of 0.64 points per grade. IEP students' science achievement trajectory had a significantly better shift of elevation compared to non-IEP student's trajectory ($\beta_{31} = 12.82, p = .002$) after exposure to the SWH approach, while the shift in slope was not significantly different ($\beta_{41} = 1.58, p = .079$). Figure 4.5 illustrates the comparison between predicted science achievement trajectories before and after exposure to the SWH approach for IEP and non-IEP students.

The effect size analysis showed that when controlling for student's IEP status, the standardized mean difference between before and after adopting the SWH approach was small (Cohen's $f = 0.10 = 1.95/\sqrt{343.32}$) and the change of slope was trivial (Cohen's $f = 0.05 = 0.86/\sqrt{343.32}$). The effect of student IEP status on the change of elevation was large (Cohen's $f = 0.44 = 12.82/\sqrt{848.83}$), while the effect of student IEP status on the change of slope was medium in size (Cohen's $f = 0.38 = 1.58/\sqrt{17.14}$).

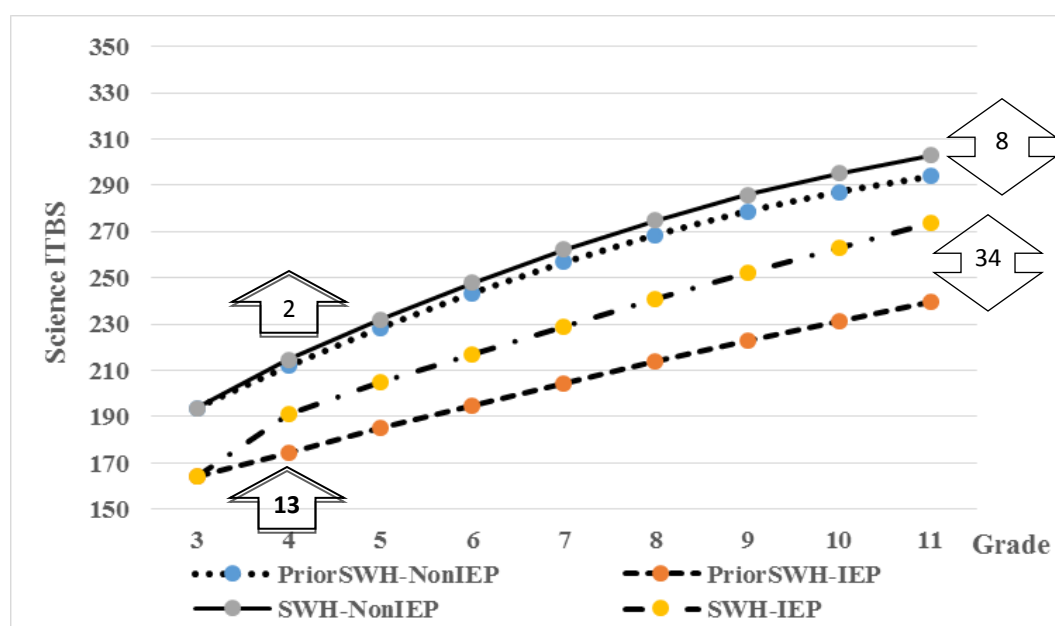
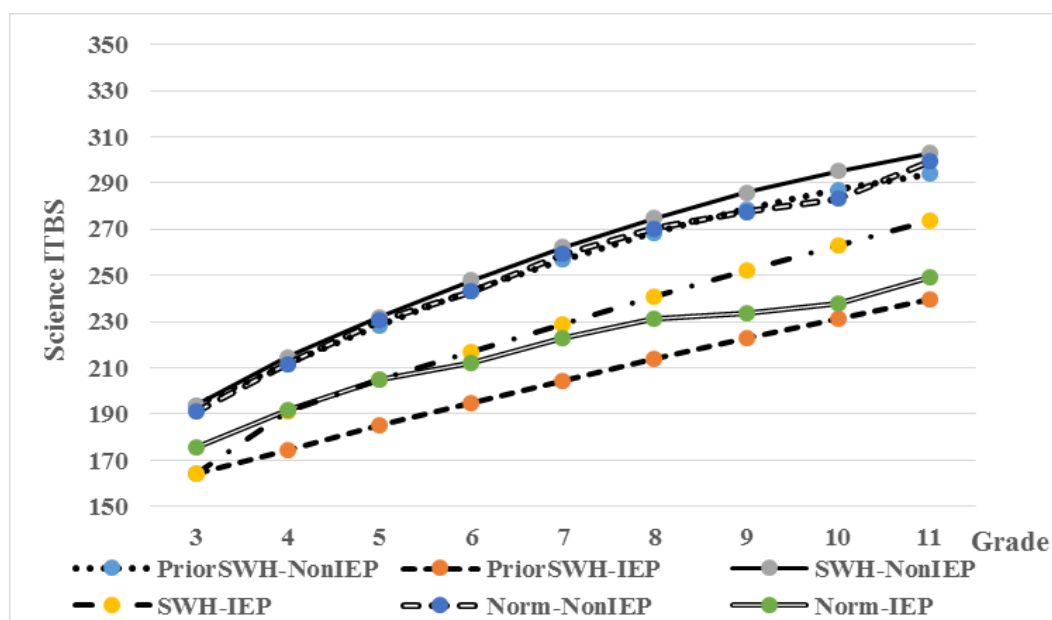


Figure 4.5 Science Achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students



Grade	3	4	5	6	7	8	9	10	11
Prior SWH									
NonIEP	193.59	211.86	228.49	243.5	256.87	268.61	278.73	287.21	294.06
IEP	164.08	174.72	185.02	194.97	204.58	213.84	222.76	231.33	239.55
Gap	29.51	37.14	43.47	48.53	52.29	54.77	55.97	55.88	54.51
SWH									
NonIEP	193.59	214.67	232.17	248.03	262.27	274.88	285.85	295.19	302.91
IEP	164.08	191.13	204.66	217.06	229.1	240.8	252.15	263.1	273.83
Gap	29.51	23.54	27.51	30.97	33.17	34.08	33.7	32.09	29.08
Norm									
Iowa	189.35	209.31	227.31	239.1	254.52	265.08	271.82	277.29	293.5
NonIEP	190.98	211.73	230.58	243.15	259.2	270.3	277.5	283.08	299.32
IEP	175.7	191.63	204.9	211.92	223.15	231.05	233.74	238.1	249.37
Gap	15.28	20.1	25.68	31.23	36.05	39.25	43.76	44.98	49.95

Figure 4.6 Science Achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students Compared to Iowa Normative Achievement

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 928.04$, $p < .001$). The level-1 variance component showing differences between each student's observed and predicted science achievement scores over time remained unchanged, $\sigma^2 = 343.31$. By adding IEP to the level-2 model, the intercept variance is decreased approximately ($[(1130.78-928.04)/1130.78]$) 18%.

The difference in deviance statistics between Model H and Model F, $115917.7 - 115377.6 = 540.10$, is distributed as a Chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(5) = 540.10$, $p < .01$, indicating that the science achievement trajectories were better predicted by adding the level-2 predictor IEP into Model F.

IEP Status Summary. The predicted science achievement scores of students in the participating school district prior to experiencing the SWH approach were approximately close to the Iowa normative for non-IEP students; however, the achievement scores of IEP students were expected to be smaller than the Iowa norms (see Figure 4.6). After students had experience with the SWH approach, the expected scores were higher than Iowa normative populations for both IEP and non-IEP students. Further, the IEP students experienced a bigger jump in their predicted achievement. In addition, the analysis of the achievement gap demonstrated that before adopting the SWH approach, the achievement gaps between IEP and non-IEP students were larger than Iowa normative populations. However, after adopting the SWH approach the predicted achievement gap tended to be stable in high school and not widen as found in the Iowa norms. More importantly, the predicted achievement gap of students after experiencing

the SWH approach were bigger than gap of Iowa norm at elementary school level while the achievement gap were smaller than the Iowa norm gaps as students moved to the middle school and high school levels.

Model J: Effect of Socioeconomic Status (SES)

Building from Model F, a binary level-2 predictor variable, SES (i.e., students who had free/reduced-lunch = 1, students who did not had free/reduced-lunch = 0) was added to the Level-2 model to explain intercept and slope variance in science achievement scores. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were centered at third grade allowing for interpreting of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}-3_{ti}) + \pi_{2i} * (\text{GRADE}-3_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

$$\begin{aligned} \text{Level-2 Model: } \pi_{0i} &= \beta_{00} + \beta_{01} * (\text{SES}_i) + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} * (\text{SES}_i) \\ \pi_{2i} &= \beta_{20} + \beta_{21} * (\text{SES}_i) \\ \pi_{3i} &= \beta_{30} + \beta_{31} * (\text{SES}_i) + r_{3i} \\ \pi_{4i} &= \beta_{40} + \beta_{41} * (\text{SES}_i) + r_{4i} \end{aligned}$$

$$\begin{aligned} \text{Mixed Model: } \text{SCIENCES}_{ti} &= \beta_{00} + \beta_{01} * \text{SES}_i + \beta_{10} * \text{GRADE}-3_{ti} + \\ &\beta_{11} * \text{SES}_i * \text{GRADE}-3_{ti} + \beta_{20} * \text{GRADE}-3_{ti}^2 + \beta_{21} * \text{SES}_i * \text{GRADE}- \\ &3_{ti}^2 + \beta_{30} * \text{SWH}_{ti} + \beta_{31} * \text{SES}_i * \text{SWH}_{ti} + \beta_{40} * \text{CONSWH}_{ti} + \\ &\beta_{41} * \text{SES}_i * \text{CONSWH}_{ti} + r_{0i} + r_{3i} * \text{SWH}_{ti} + r_{4i} * \text{CONSWH}_{ti} + e_{ti} \end{aligned} \quad (4.13)$$

The multilevel model was estimated with results presenting in Table 4.4. In order to allow easier interpretation of results, the value of SES (students who had free/reduced-lunch = 1, students who did not had free/reduced-lunch = 0) was substituted into Equation 4.69.

Non-low-SES:
$$SCIENCE_{ii} = 194.99 + 18.37*GRADE-3_{ii} - 0.79*GRADE-3_{ii}^2 + 1.55*SWH_{ii} + 1.45*CONSWH_{ii} \quad (4.14)$$

Low-SES:
$$SCIENCE_{ii} = 173.82 + 16.10*GRADE-3_{ii} - 0.61*GRADE-3_{ii}^2 + 12.28*SWH_{ii} + 1.34*CONSWH_{ii} \quad (4.15)$$

Non-Low-SES. The grand-mean science achievement score for non-low SES students in grade 3 was significant ($\beta_{00} = 194.99, p < .001$); the rate of change for non-low SES students was significant ($\beta_{10} = 18.37, p < .001$); and the curvature parameter was also significant ($\beta_{20} = -0.78, p < .001$). The rate of change for each successive grade, β_{10} , rose by 18.37 points.

However, the negative curvature parameter indicates that this rate of change declines by 0.78 points per grade. The elevation of science achievement scores for non-low-SES students was not significant ($\beta_{30} = 1.55, p = .243$), that is, after non-low-SES students experienced the SWH approach, their science achievement scores increased by non-significant 1.55 points. The shift of the slope of science achievement of 1.45 points after non-low-SES students' exposure to the SWH was significant ($\beta_{40} = 1.45, p < .001$).

Low-SES Students. The results showed that low-SES students had significant lower grand-mean science achievement scores in grade 3 ($\beta_{01} = -21.13, p < .001$) when compared to non-low-SES students. At grade three, the rate of change for low-SES students was significant lower than non-low-SES students ($\beta_{11} = -2.27, p = .007$); however, the curvature parameter was not significantly higher ($\beta_{21} = 0.17, p = .079$). Low-SES students' trajectory rose by 16.10 points per grade level with additional non-significant acceleration of 0.17 points per grade. Low-SES students' science achievement trajectory had a significantly better shift elevation compared to non-low-SES student's trajectory ($\beta_{31} = 10.73, p = .003$) after exposure to the SWH approach, while the shift in slope was not significantly different ($\beta_{41} = -0.11, p = 0.886$). Figure 4.7 illustrates the

comparison between predicted science achievement score trajectories before and after exposure to the SWH approach for low-SES and non-low-SES students.

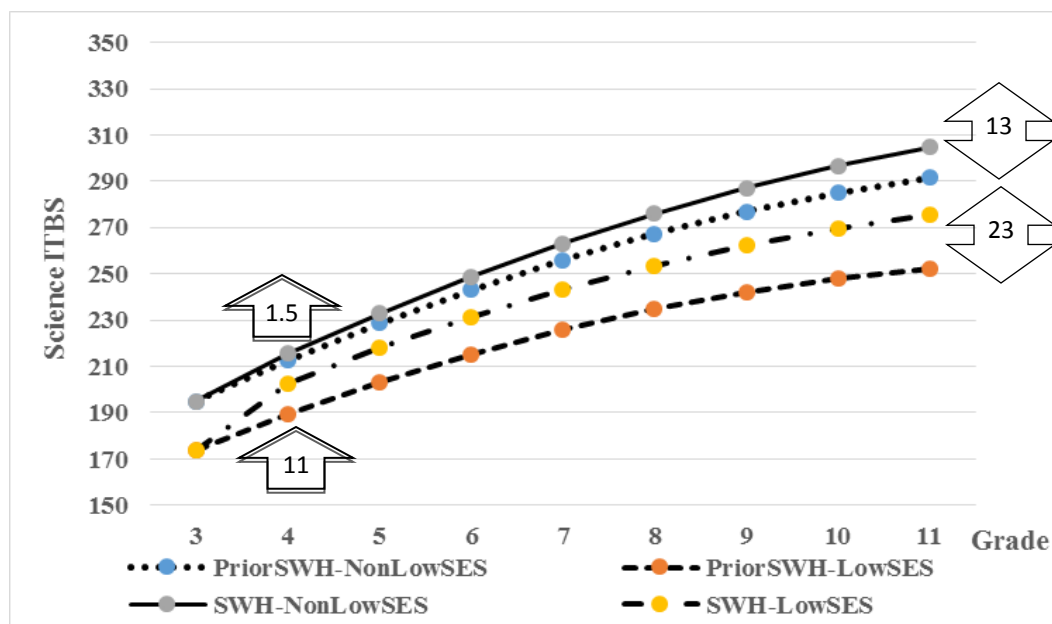


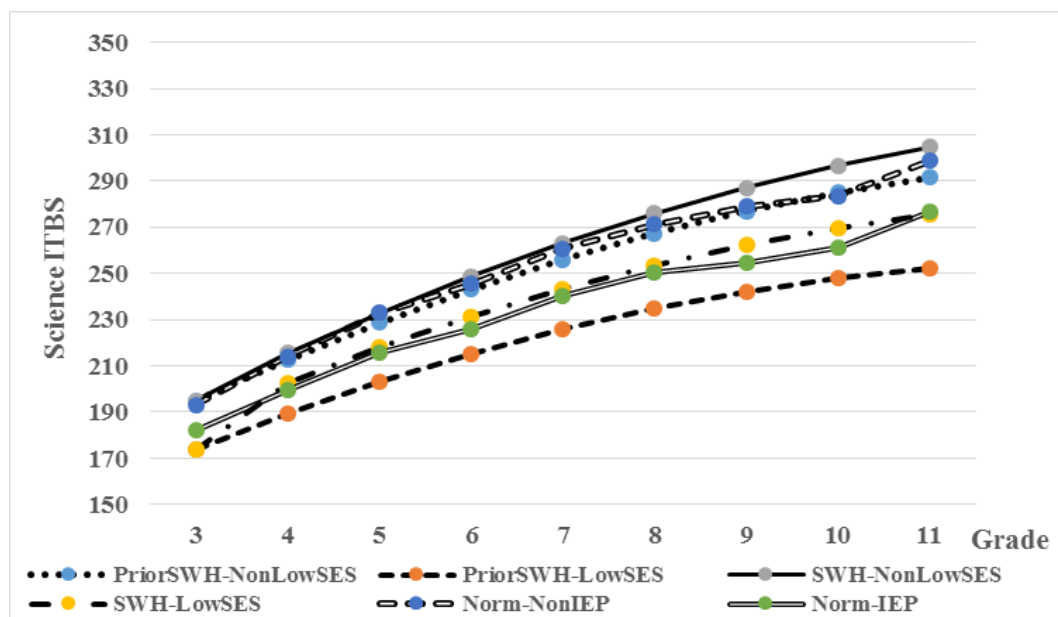
Figure 4.7 Science Achievement Trajectories before and after Exposure to the SWH Approach for Low-SES and Non-low-SES Students

The effect size analysis showed that when controlling for students' SES status, the standardized mean difference between before and after adopting the SWH approach was trivial (Cohen's $f = 0.08 = 1.55/\sqrt{342.63}$) and the change of slope was also trivial (Cohen's $f = 0.08 = 1.45/\sqrt{342.63}$). The effect of student SES status on the change of elevation was medium in size (Cohen's $f = 0.39 = 10.73/\sqrt{805.38}$), while the effect of student IEP status on the change of slope was trivial (Cohen's $f = 0.02 = 0.11/\sqrt{17.83}$).

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 1061.22, p < .001$). The level-1 variance component showing differences between each student's observed and predicted science achievement scores over time slightly changed, $\sigma^2 = 342.62$. By adding SES to the level-2 model, the intercept variance was decreased approximately $([1130.78-1061.22]/1130.78)$ 6%.

The difference in deviance statistics between Model M and Model F, $115917.7-115723.8 = 193.9$, is distributed as a Chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(5) = 193.9, p < .01$, indicating that the science achievement trajectories were better predicted by adding the level-2 predictor SES into Model F.

SES Status Summary. The predicted science achievement growth trajectories were non-linear and consistent with the predicted overall and Iowa normative trajectories. Average scores increased at faster rate in lower grades than in higher grades. The predicted average achievement scores for low-SES students prior to exposure to the SWH approach were lower than the Iowa normative achievement. However, after students had experience with the SWH approach, the predicted achievement scores were increased and higher than Iowa normative. The predicted average achievement scores for non-low-SES students prior to SWH experience were roughly close to the Iowa normative achievement. When non-low-SES students had experience with the SWH approach, their predicted score also increased to higher level than Iowa normative achievement; however, the rising scores were not as big as the one found for low-SES students.



Grade	3	4	5	6	7	8	9	10	11
Prior SWH									
NonLowSES	194.99	212.57	228.59	243.03	255.9	267.21	276.94	285.09	291.68
LowSES	173.86	189.17	202.92	215.09	225.7	234.73	242.19	248.08	252.4
<i>Gap</i>	<i>21.13</i>	<i>23.4</i>	<i>25.67</i>	<i>27.94</i>	<i>30.2</i>	<i>32.48</i>	<i>34.75</i>	<i>37.01</i>	<i>39.28</i>
SWH									
NonLowSES	194.99	215.57	233.03	248.92	263.24	275.99	287.17	296.78	304.81
LowSES	173.86	202.79	217.88	231.39	243.34	253.71	262.51	269.74	275.4
<i>Gap</i>	<i>21.13</i>	<i>12.78</i>	<i>15.15</i>	<i>17.53</i>	<i>19.9</i>	<i>22.28</i>	<i>24.66</i>	<i>27.04</i>	<i>29.41</i>
Norm									
NonlowSES	193.24	214.13	233.07	245.45	260.84	271.36	278.96	283.51	298.69
LowSES	181.92	199.75	215.95	225.92	240.33	250.27	254.86	261.49	277.03
<i>Gap</i>	<i>11.32</i>	<i>14.38</i>	<i>17.12</i>	<i>19.53</i>	<i>20.51</i>	<i>21.09</i>	<i>24.1</i>	<i>22.02</i>	<i>21.66</i>

Figure 4.8 Science achievement Trajectories before and after Exposure to the SWH Approach for IEP and Non-IEP Students Compared to Iowa Normative Achievement

Moreover, the predicted achievement gaps were found to widen over time. At third grade, the predicted achievement gaps for the participating school district were approximately 10 points bigger than the achievement gap found in the Iowa normative populations. By ninth grade, where the achievement gap of Iowa students reached its highest point, the predicted achievement gap of low-SES students without the SWH exposure widened and reached 15 points bigger than the Iowa normative scores. However, the predicted achievement gap of the low-SES students who received the SWH was approximately equal to the Iowa norms indicating the SWH exposure effectively closing the achievement gap.

Model J: Effect of Initial Reading Achievement Level

Building from Model F, a level-2 predictor variable, INREAD (Low=1, Medium=2, Proficient= 3, and High= 4) was added to the level-2 model to explain intercept and slope variance in science achievement scores. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were centered at third grade to allow for interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE} - 3_{ti}) + \pi_{2i} * (\text{GRADE} - 3_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

$$\begin{aligned} \text{Level-2 Model: } \pi_{0i} &= \beta_{00} + \beta_{01} * (\text{INREAD}_i) + r_{0i} \\ \pi_{1i} &= \beta_{10} + \beta_{11} * (\text{INREAD}_i) \\ \pi_{2i} &= \beta_{20} + \beta_{21} * (\text{INREAD}_i) \\ \pi_{3i} &= \beta_{30} + \beta_{31} * (\text{INREAD}_i) + r_{3i} \\ \pi_{4i} &= \beta_{40} + \beta_{41} * (\text{INREAD}_i) + r_{4i} \end{aligned}$$

Mixed Model:

$$\begin{aligned} \text{SCIENCES}_{ti} = & \beta_{00} + \beta_{01} * \text{INREAD}_i + \beta_{10} * \text{GRADE-3}_{ti} + \\ & \beta_{11} * \text{INREAD}_i * \text{GRADE-3}_{ti} + \beta_{20} * \text{GRADE-3}_{ti}^2 + \beta_{21} * \text{INREAD}_i * \\ & \text{GRADE-3}_{ti}^2 + \beta_{30} * \text{SWH}_{ti} + \beta_{31} * \text{INREAD}_i * \text{SWH}_{ti} + \\ & \beta_{40} * \text{CONSWH}_{ti} + \beta_{41} * \text{INREAD}_i * \text{CONSWH}_{ti} + r_{0i} + r_{3i} * \text{SWH}_{ti} + \\ & r_{4i} * \text{CONSWH}_{ti} + e_{ti} \end{aligned} \quad (4.16)$$

The multilevel model was estimated with results presenting in Table 4.5. In order to allow easier interpretation of results, the values of InREAD (Low=1, Medium=2, Proficient= 3, and High= 4) are substituted into Equation 4.16.

Extremely low initial reading achievement:

$$\text{SCIENCES}_{ii} = 152.03 + 11.40 * \text{GRADE-3}_{ii} + 14.32 * \text{SWH}_{ii} - 0.87 * \text{CONSWH}_{ii} \quad (4.17)$$

Low initial reading achievement:

$$\begin{aligned} \text{SCIENCES}_{ii} = & 173.56 + 15.50 * \text{GRADE-3}_{ii} - 0.40 * \text{GRADE-3}_{ii}^2 + 7.65 * \text{SWH}_{ii} \\ & + 0.26 * \text{CONSWH}_{ii} \end{aligned} \quad (4.18)$$

Proficient initial reading achievement:

$$\begin{aligned} \text{SCIENCES}_{ii} = & 195.10 + 19.60 * \text{GRADE-3}_{ii} - 0.82 * \text{GRADE-3}_{ii}^2 + 0.98 * \text{SWH}_{ii} \\ & + 0.33 * \text{CONSWH}_{ii} \end{aligned} \quad (4.19)$$

High initial reading achievement:

$$\begin{aligned} \text{SCIENCES}_{ii} = & 216.63 + 23.70 * \text{GRADE-3}_{ii} - 1.23 * \text{GRADE-3}_{ii}^2 - 5.69 * \text{SWH}_{ii} + \\ & 0.94 * \text{CONSWH}_{ii} \end{aligned} \quad (4.20)$$

The intercept for science achievement scores in grade 3 was significant ($\beta_{00} = 130.50, p < .001$) and significantly varied among initial reading achievement levels ($\beta_{01} = 21.53, p < .001$). Students with extremely low initial reading achievement have grand-mean science achievement scores at 152.03 and students with low initial reading achievement have predicted grand-mean science achievement scores at 173.46, 21.53 points higher than those with extremely low initial reading achievement. At grade three, the rate of change was significantly varied among levels of student's reading achievement ($\beta_{11} = 4.10, p < .001$). Students with extremely low initial reading achievement had rates

of change equal to 11.39, 4.10 points smaller than students with low initial reading achievement. The curvature parameter at grade three was also significant ($\beta_{20} = 0.42, p < .001$) and significant varied among levels of initial reading achievement. Equation 4.17 showed that student with extremely low initial reading achievement had a positive but very small curvature parameter ($0.42 - 0.41 = 0.01$). For students with extremely low initial reading achievement, the trajectory increased by 11.40 points per grade with an additional acceleration of 0.01 points per grade. The curvature parameter was negative for students with high initial reading achievement indicating that their rate of change declined by 1.23 points at each successive grade.

The elevation of science achievement scores for students who exposed to the SWH approach was significant ($\beta_{30} = 20.99, p < .001$) and also significantly varied among levels of initial reading achievement ($\beta_{31} = -6.67, p < .001$). The elevation of students with extremely low initial reading achievement was highest with 14.32 points and decreased as the initial reading achievement levels got higher. Students with high initial reading achievement had negative elevation (-5.69) indicating that inclined in their science achievement scores after the SWH exposure. Figure 4.9 illustrates the comparison between predicted science achievement score trajectories before and after exposure to the SWH approach for students with different initial reading achievement levels.

The effect size analysis showed that when controlling for student initial reading achievement level, the standardized mean different between before and after adopting the SWH approach was large (Cohen's $f = 1.13 = 20.99/\sqrt{344.53}$) while the change of slope was trivial (Cohen's $f = 0.08 = 1.48/\sqrt{344.53}$). The effect of student's initial reading achievement level on the change of elevation was medium in size (Cohen's $f = 0.27 =$

6.67/ $\sqrt{603.28}$) and the effect of student's initial reading achievement level on the change of slope was small (Cohen's $f = 0.14 = 0.60/\sqrt{17.83}$).

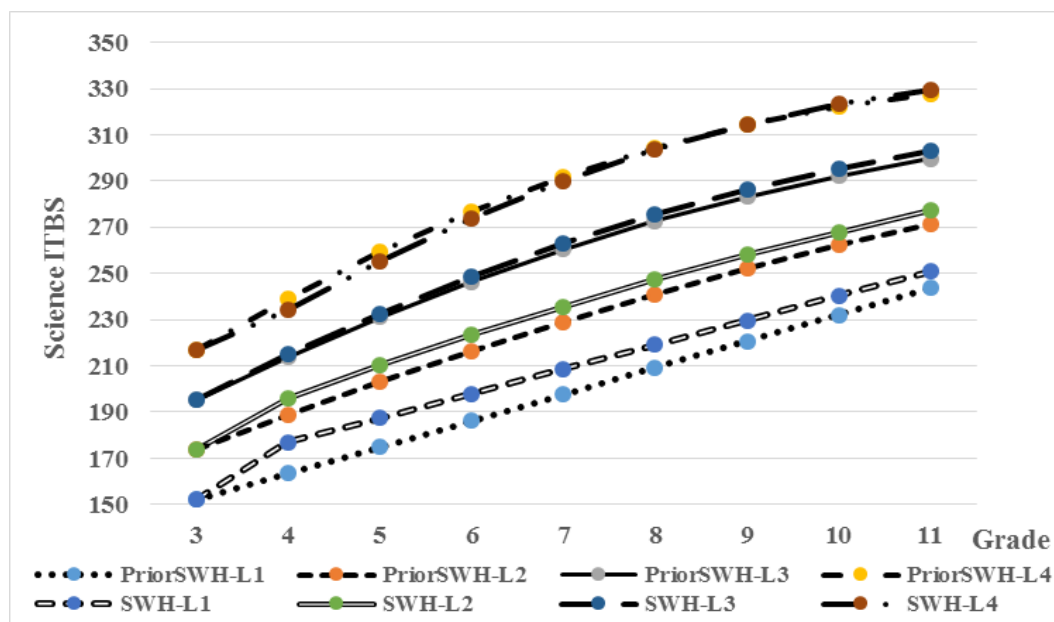
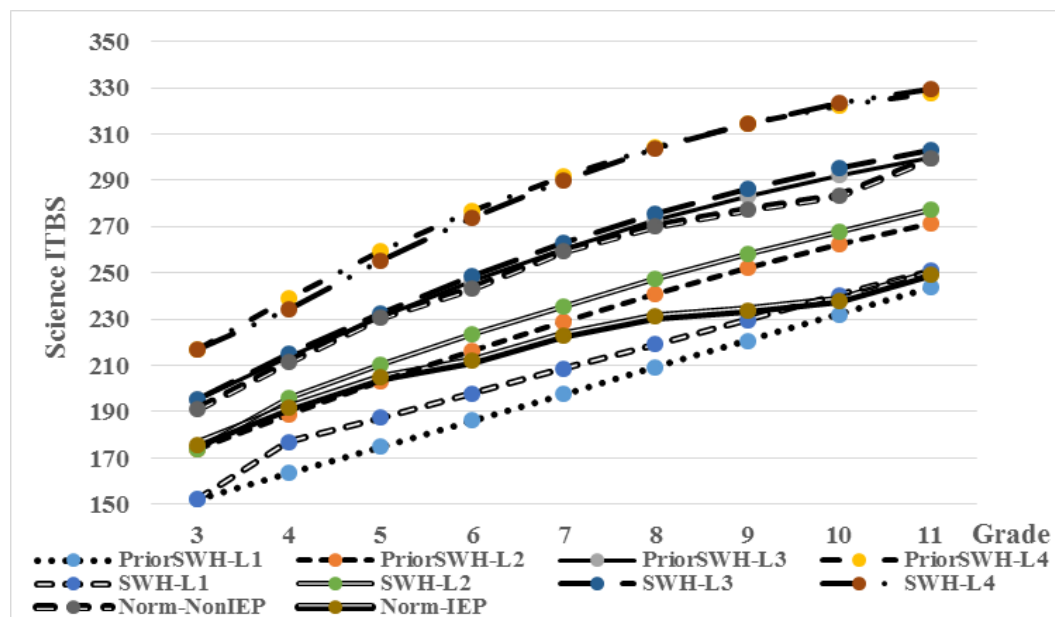


Figure 4.9 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 447.99, p < .001$). The level-1 variance component showing differences between each student's observed and predicted science achievement scores over time slightly changed, $\sigma^2 = 344.53$. By adding student's initial reading achievement level to level-2 model, the intercept variance decreased approximately $([1130.78 - 447.99]/1130.78)$ 60% indicating that student's initial reading achievement level helped explain 60% of the variation of science achievement scores across students.



Grade	3	4	5	6	7	8	9	10	11
Non-SWH									
Level1	152.03	163.43	174.85	186.28	197.72	209.18	220.65	232.13	243.63
Level2	173.56	188.66	202.94	216.41	229.06	240.91	251.95	262.17	271.59
Level3	195.1	213.88	231.02	246.53	260.4	272.64	283.25	292.21	299.55
Level4	216.63	239.1	259.11	276.66	291.75	304.38	314.55	322.26	327.51
L4-L1	64.6	75.67	84.26	90.38	94.03	95.2	93.9	90.13	83.88
SWH									
Level1	152.03	176.89	187.43	197.99	208.56	219.15	229.75	240.36	250.99
Level2	173.56	196.04	210.06	223.26	235.65	247.24	258.01	267.97	277.12
Level3	195.1	215.2	232.68	248.53	262.74	275.32	286.26	295.57	303.24
Level4	216.63	234.35	255.31	273.8	289.83	303.41	314.52	323.18	329.37
L4-L1	64.6	57.46	67.88	75.81	81.27	84.26	84.77	82.82	78.38

Figure 4.10 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Compared to Iowa Normative Population

Student's initial reading achievement also helped explain $([840.46-603.28]/840.46)$ 28% of the variation of science achievement scores across students with different exposure of the SWH approach.

The difference in deviance statistics between Model M and Model F, $115917.7 - 113797.6 = 2120.05$ is distributed as a Chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(5) = 2120.05, p < .01$, indicating that the science achievement trajectories were better predicted by adding the level-2 predictor student's initial reading achievement level into Model F.

Model K: Investigating All Variables Simultaneously

Building from all previous models, this model explored the effect of all level-2 variables simultaneously. Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were centered at third grade to allow for interpretation of the intercept as the average science achievement score for students in grade three.

Level-1 Model:
$$\text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE}-3_{ti}) + \pi_{2i} * (\text{GRADE}-3_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

Level-2 Model:
$$\pi_{0i} = \beta_{00} + \beta_{01} * (\text{GENDER}_i) + \beta_{02} * (\text{IEP}_i) + \beta_{03} * (\text{SES}_i) + \beta_{04} * (\text{INREAD}_i) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (\text{GENDER}_i) + \beta_{12} * (\text{IEP}_i) + \beta_{13} * (\text{SES}_i) + \beta_{14} * (\text{INREAD}_i)$$

$$\pi_{2i} = \beta_{20} + \beta_{21} * (\text{GENDER}_i) + \beta_{22} * (\text{IEP}_i) + \beta_{23} * (\text{SES}_i) + \beta_{24} * (\text{INREAD}_i)$$

$$\pi_{3i} = \beta_{30} + \beta_{31} * (\text{GENDER}_i) + \beta_{32} * (\text{IEP}_i) + \beta_{33} * (\text{SES}_i) + \beta_{34} * (\text{INREAD}_i) + r_{3i}$$

$$\pi_{4i} = \beta_{40} + \beta_{41} * (\text{GENDER}_i) + \beta_{42} * (\text{IEP}_i) + \beta_{43} * (\text{SES}_i) + \beta_{44} * (\text{INREAD}_i) + r_{4i}$$

Mixed Model:
$$\text{SCIENCES}_{ti} = \beta_{00} + \beta_{01} * \text{GENDER}_i + \beta_{02} * \text{IEP}_i + \beta_{03} * \text{SES}_i + \beta_{04} * \text{INREAD}_i + \beta_{10} * \text{GRADE}-3_{ti} + \beta_{11} * \text{GENDER}_i * \text{GRADE}-3_{ti} + \beta_{12} * \text{IEP}_i * \text{GRADE}-3_{ti} + \beta_{13} * \text{SES}_i * \text{GRADE}-3_{ti} + \beta_{14} * \text{INREAD}_i * \text{GRADE}-3_{ti} + \beta_{20} * \text{GRADE}-3_{ti}^2 +$$

$$\begin{aligned}
& \beta_{21} * \text{GENDER}_i * \text{GRADE-3}_{ti}^2 + \beta_{22} * \text{IEP}_i * \text{GRADE-3}_{ti}^2 + \\
& \beta_{23} * \text{SES}_i * \text{GRADE-3}_{ti}^2 + \beta_{24} * \text{INREAD}_i * \text{GRADE-3}_{ti}^2 + \beta_{30} * \text{SWH}_{ti} \\
& + \beta_{31} * \text{GENDER}_i * \text{SWH}_{ti} + \beta_{32} * \text{IEP}_i * \text{SWH}_{ti} + \beta_{33} * \text{SES}_i * \text{SWH}_{ti} + \\
& \beta_{34} * \text{INREAD}_i * \text{SWH}_{ti} + \beta_{40} * \text{CONSWH}_{ti} + \\
& \beta_{41} * \text{GENDER}_i * \text{CONSWH}_{ti} + \beta_{42} * \text{IEP}_i * \text{CONSWH}_{ti} + \\
& \beta_{43} * \text{SES}_i * \text{CONSWH}_{ti} + \beta_{44} * \text{INREAD}_i * \text{CONSWH}_{ti} + r_{0i} + \\
& r_{3i} * \text{SWH}_{ti} + r_{4i} * \text{CONSWH}_{ti} + e_{ti}
\end{aligned} \tag{4.21}$$

The multilevel model was estimated with results presenting in Table 4.5. The intercept of science achievement score in grade 3 was significant ($\beta_{00}=134.72, p < .001$). The achievement gap of the grand mean of science achievement between genders ($\beta_{01} = -0.52, p = 0.822$) and IEP status ($\beta_{02} = -4.18, p = 0.361$) was not significant. However, the achievement gap of the grand mean of science achievement was significant between students from different SES levels ($\beta_{03} = -6.36, p = 0.047$) and students with different initial reading achievement levels ($\beta_{04} = 20.69, p < .001$). At grade three, the rate of change was significant ($\beta_{10} = 10.94, p < .001$) and varied between students with different genders ($\beta_{11} = -2.38, p < .001$), IEP statuses ($\beta_{12} = -3.91, p < .001$), and initial reading achievement levels ($\beta_{14} = 3.39, p < .001$). SES was unrelated ($\beta_{13} = -0.38, p = 0.626$). The curvature parameter was not significant ($\beta_{20} = 0.04, p = 0.839$) nor varied between students with different IEP statuses ($\beta_{22} = 0.28, p = 0.070$), SES statuses ($\beta_{23} = 0.11, p = 0.273$). However, the curvature parameter was significant varied between gender ($\beta_{21} = 0.23, p = 0.003$) and among students with different initial reading achievement levels ($\beta_{24} = -0.34, p < .001$).

The elevation of science achievement scores for students exposed to the SWH approach was significant ($\beta_{30} = 19.40, p < 0.01$) and significantly varied among levels of initial reading achievement ($\beta_{34} = -6.24, p < .001$) but not varied between genders ($\beta_{31} =$

0.20, $p = .926$), IEP statuses ($\beta_{32} = 4.07$, $p = .372$), and SES levels ($\beta_{33} = 2.96$, $p = .357$). The shift of the slope was not significant ($\beta_{40} = -1.42$, $p = .223$) nor significant varied between gender ($\beta_{41} = 0.79$, $p = .143$), IEP statuses ($\beta_{42} = 1.40$, $p = .154$), SES levels ($\beta_{43} = -1.30$, $p = .051$), and initial reading achievements ($\beta_{44} = 0.59$, $p = .102$).

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 420.90$, $p < 0.01$). The level-1 variance component showing differences between each student's observed and predicted science achievement scores over time slightly changed, $\sigma^2 = 345.74$. By adding all level-2 variables simultaneously, the intercept variance decreased approximately ($[1130.78 - 420.90] / 1130.78$) 63%. The deviance statistics of Model K was the smallest when compared to models that explored each level-2 variable separately.

Model L: Final Model

The variables of most interest was elevation and shift of slope associated with exposure to the SWH approach. According to Model K, the results indicated that only the shift of the elevation was significantly varied among students with different initial reading achievement levels, therefore only the effect of student's initial reading achievement level is focused.

Model L explores the effect of student's initial reading achievement level, InRead (Low=1, Medium=2, Proficient= 3, and High= 4) by controlling the effect of IEP, SES, and Gender (grand centering). Since the science achievement scores were collected from third grade and to be consistent with Model B, the GRADE and GRADE² variables were

centered at third grade to allow for interpretation of the intercept as the average science achievement score for students in grade three.

$$\text{Level-1 Model: } \text{SCIENCES}_{ti} = \pi_{0i} + \pi_{1i} * (\text{GRADE-3}_{ti}) + \pi_{2i} * (\text{GRADE-3}_{ti})^2 + \pi_{3i} * (\text{SWH}_{ti}) + \pi_{4i} * (\text{CONSWH}_{ti}) + e_{ti}$$

$$\text{Level-2 Model: } \pi_{0i} = \beta_{00} + \beta_{01} * (\text{CGENDER}_i) + \beta_{02} * (\text{CIEP}_i) + \beta_{03} * (\text{CSES}_i) + \beta_{04} * (\text{INREAD}_i) + r_{0i}$$

$$\pi_{1i} = \beta_{10} + \beta_{11} * (\text{CGENDER}_i) + \beta_{12} * (\text{CIEP}_i) + \beta_{13} * (\text{CSES}_i) + \beta_{14} * (\text{INREAD}_i)$$

$$\pi_{2i} = \beta_{20} + \beta_{21} * (\text{CGENDER}_i) + \beta_{22} * (\text{CIEP}_i) + \beta_{23} * (\text{CSES}_i) + \beta_{24} * (\text{INREAD}_i)$$

$$\pi_{3i} = \beta_{30} + \beta_{31} * (\text{CGENDER}_i) + \beta_{32} * (\text{CIEP}_i) + \beta_{33} * (\text{CSES}_i) + \beta_{34} * (\text{INREAD}_i) + r_{3i}$$

$$\pi_{4i} = \beta_{40} + \beta_{41} * (\text{CGENDER}_i) + \beta_{42} * (\text{CIEP}_i) + \beta_{43} * (\text{CSES}_i) + \beta_{44} * (\text{INREAD}_i) + r_{4i}$$

$$\begin{aligned} \text{Mixed Model: } \text{SCIENCES}_{ti} = & \beta_{00} + \beta_{01} * \text{CGENDER}_i + \beta_{02} * \text{CIEP}_i + \beta_{03} * \text{CSES}_i + \\ & \beta_{04} * \text{INREAD}_i + \beta_{10} * \text{GRADE-3}_{ti} + \beta_{11} * \text{CGENDER}_i * \text{GRADE-3}_{ti} + \\ & \beta_{12} * \text{CIEP}_i * \text{GRADE-3}_{ti} + \beta_{13} * \text{CSES}_i * \text{GRADE-3}_{ti} + \beta_{14} * \text{INREAD}_i * \\ & \text{GRADE-3}_{ti} + \beta_{20} * \text{GRADE-3}_{ti}^2 + \beta_{21} * \text{CGENDER}_i * \text{GRADE-3}_{ti}^2 + \\ & \beta_{22} * \text{CIEP}_i * \text{GRADE-3}_{ti}^2 + \beta_{23} * \text{CSES}_i * \text{GRADE-3}_{ti}^2 + \\ & \beta_{24} * \text{INREAD}_i * \text{GRADE-3}_{ti}^2 + \beta_{30} * \text{SWH}_{ti} + \\ & \beta_{31} * \text{CGENDER}_i * \text{SWH}_{ti} + \beta_{32} * \text{CIEP}_i * \text{SWH}_{ti} + \beta_{33} * \text{CSES}_i * \text{SWH}_{ti} \\ & + \beta_{34} * \text{INREAD}_i * \text{SWH}_{ti} + \beta_{40} * \text{CONSWH}_{ti} + \\ & \beta_{41} * \text{CGENDER}_i * \text{CONSWH}_{ti} + \beta_{42} * \text{CIEP}_i * \text{CONSWH}_{ti} + \\ & \beta_{43} * \text{CSES}_i * \text{CONSWH}_{ti} + \beta_{44} * \text{INREAD}_i * \text{CONSWH}_{ti} + r_{0i} + \\ & r_{3i} * \text{SWH}_{ti} + r_{4i} * \text{CONSWH}_{ti} + e_{ti} \end{aligned} \quad (4.22)$$

The multilevel model was estimated with results presenting in Table 4.6. In order to allow easier interpretation of results, the value of InREAD (Low=1, Medium=2, Proficient= 3, and High= 4) was substituted into Equation 4.22.

Low initial reading achievement:

$$\text{SCIENCES}_{ii} = 152.93 + 12.56 * \text{GRADE-3}_{ii} - 0.13 * \text{GRADE-3}_{ii}^2 + 14.37 * \text{SWH}_{ii} - 0.61 * \text{CONSWH}_{ii} \quad (4.23)$$

Medium Low initial reading achievement:

$$SCIENCE_{ii} = 173.62 + 15.95 * GRADE-3_{ii} - 0.47 * GRADE-3_{ii}^2 + 8.12 * SWH_{ii} - 0.02 * CONSWH_{ii} \quad (4.24)$$

Proficient initial reading achievement:

$$SCIENCE_{ii} = 194.31 + 19.34 * GRADE-3_{ii} - 0.81 * GRADE-3_{ii}^2 + 1.88 * SWH_{ii} + 0.57 * CONSWH_{ii} \quad (4.25)$$

High initial reading achievement

$$SCIENCE_{ii} = 215.00 + 22.73 * GRADE-3_{ii} - 1.15 * GRADE-3_{ii}^2 - 4.36 * SWH_{ii} + 1.16 * CONSWH_{ii} \quad (4.26)$$

The results for Model L were similar to Model J. Model L showed a significant intercept of science achievement score in grade 3 ($\beta_{00} = 132.25$, $p < .001$) and significantly varied among students with different initial reading achievement levels ($\beta_{04} = 20.69$, $p < .001$). Student with extremely low initial reading achievement had grand-mean science achievement score at 152.93 and students with low initial reading achievement had grand-mean science achievement at 173.62. The rate of change was significant ($\beta_{10} = 9.17$, $p < .001$) and varied among levels of student's reading achievement ($\beta_{14} = 3.39$, $p < .001$). The curvature parameter was not significant ($\beta_{20} = 0.21$, $p = .192$); however, the curvature parameter did significantly varying among levels of initial reading achievement ($\beta_{24} = -0.34$, $p < .001$).

The elevation of science achievement scores for students after exposed to the SWH approach was significant ($\beta_{30} = 20.61$, $p < .001$) and significantly varied among levels of initial reading achievement ($\beta_{31} = -6.24$, $p < .001$). The elevation of students with extremely low initial reading achievement was highest at 14.37 points and decreased as initial reading achievement level got higher. Students with high initial reading achievement had negative elevation (-4.36) indicating that their science scores dropped

with the SWH exposure. The shift in slope was not significant ($\beta_{40} = -1.20, p = .256$) nor was it significantly varied among groups of students ($\beta_{44} = 0.59, p=0.102$). Figure 4.11 illustrates the comparison between predicted science achievement score trajectories before and after exposure to the SWH approach for students with different initial reading achievement levels.

The effect size analysis showed that when controlling for student gender, IEP, and SES, the standardized mean difference between before and after adopting the SWH approach was large (Cohen's $f = 1.04 = 19.40/\sqrt{345.74}$) and the change of slope was trivial (Cohen's $f = 0.08 = 1.42/\sqrt{345.74}$). The effect of student's initial reading achievement level on the change of elevation was medium in size (Cohen's $f = 0.25 = 6.24/\sqrt{606.41}$), while the effects of student IEP and SES status on the change of elevation were small (Cohen's $f = 0.16$ and 0.12 , respectively). The effect of gender on the change of elevation was trivial (Cohen's $f = 0.00 = 21/\sqrt{606.41}$). The effects of student's initial reading achievement level and gender on the change of slope small (Cohen's $f = 0.14$ and 0.19 , respectively). The effects of student's IEP and SES statuses on the change of slope were medium in size (Cohen's $f = 0.34$ and 0.31 , respectively).

Since Model L was developed based on the same data set and predicted variables as Model M, the variance component and deviance statistic remained unaffected. Comparing model L to Model F, by adding all level-2 variables simultaneously, the intercept variance decreased approximately ($[1130.78-420.90]/1130.78$) 63%. The variation associated with the SWH exposure decreased ($[840.46-606.41]/840.46$) 28%.

The final model where InRead was the main interest variable and GENDER, IEP, and SES were used as covariate showed similar trends with previous models where

InREAD was used as only level-2 variable. The predicted achievement gaps were smaller than those found in the InREAD-only model.

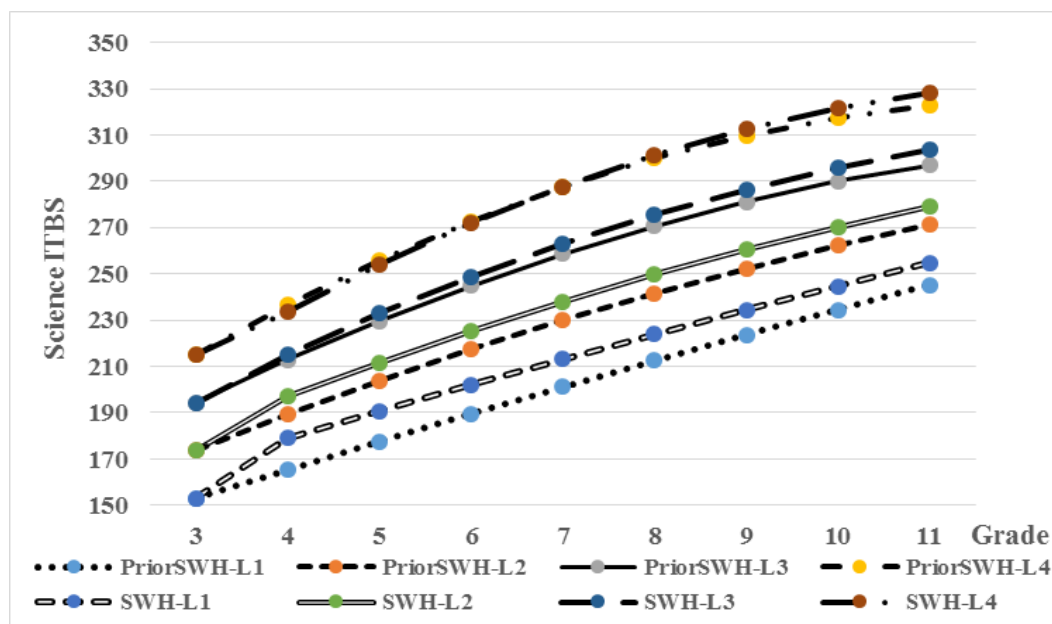
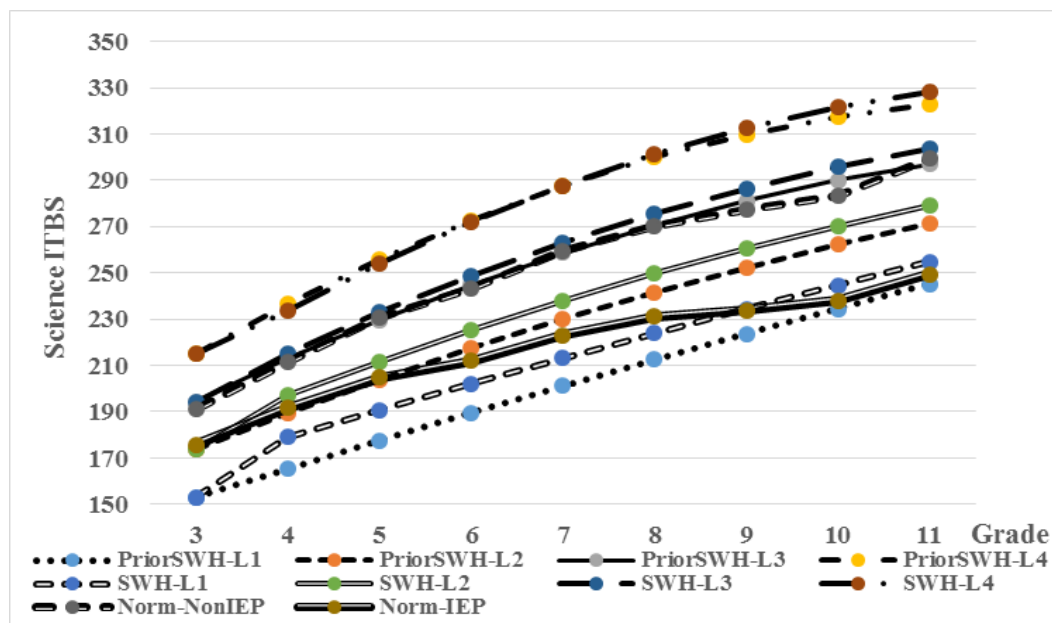


Figure 4.11 Science Achievement Trajectories before and after exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Controlling for Gender, IEP, and SES Status.



Grade	3	4	5	6	7	8	9	10	11
Non-SWH									
Level1	152.94	165.37	177.54	189.46	201.11	212.51	223.66	234.54	245.17
Level2	173.62	189.1	203.64	217.24	229.9	241.62	252.4	262.24	271.13
Level3	194.31	212.84	229.75	245.03	258.69	270.73	281.14	289.93	297.1
Level4	215	236.58	255.85	272.82	287.48	299.84	309.88	317.62	323.06
L3-L1	41.37	47.47	52.21	55.57	57.58	58.22	57.48	55.39	51.93
SWH									
Level1	152.94	179.13	190.69	201.99	213.04	223.83	234.36	244.64	254.65
Level2	173.62	197.21	211.73	225.31	237.95	249.65	260.41	270.23	279.11
Level3	194.31	215.3	232.78	248.64	262.87	275.48	286.47	295.83	303.57
Level4	215	233.39	253.83	271.96	287.79	301.31	312.52	321.42	328.02
L3-L1	41.37	36.17	42.09	46.65	49.83	51.65	52.11	51.19	48.92

Figure 4.12 Science Achievement Trajectories before and after Exposure to the SWH Approach for Students with Different Initial Reading Achievement Levels Compared to Iowa Normative Population Controlling for Gender, IEP, and SES Status

Table 4.6 Results of Final Model Centering at Grade 3 to 11

		Model	Model	Model	Model	Model	Model	Model	Model	Model	
		L (G3)	M (G4)	N (G5)	O (G6)	P (G7)	Q(G8)	R (G9)	S (G10)	T (G11)	
composite model	Intercept	B₀₀	**132.25	**141.63	**151.44	**161.67	**172.32	**183.41	**194.92	**206.85	**219.21
		SE	4.45	3.67	3.19	2.91	2.75	2.72	2.91	3.45	4.41
	Gender	B₀₁	-0.52	-2.67	*-4.37	** -5.62	** -6.40	** -6.72	** -6.59	** -6.00	* -4.96
		SE	2.29	1.97	1.78	1.68	1.64	1.66	1.73	2.08	2.55
	IEP	B₀₂	-4.18	-7.81	** -10.89	** -13.41	** -15.38	** -16.79	** -17.64	** -17.94	** -17.68
		SE	4.57	4.02	3.70	3.53	3.46	3.50	3.25	4.24	5.15
	SES	B₀₃	* -6.36	* -6.63	* -6.69	* -6.52	* -6.15	* -5.55	-4.72	-3.70	-2.45
		SE	3.20	2.87	2.69	2.59	2.57	2.62	2.67	3.19	3.80
	INRT	B₀₄	**20.69	**23.74	**26.10	**27.79	**28.79	**29.11	**28.74	**27.69	**25.96
		SE	1.58	1.32	1.16	1.06	1.01	1.00	1.05	1.23	1.55
Grade	Intercept	B₁₀	**9.17	**9.59	**10.02	**10.44	**10.87	**11.29	**11.72	**12.14	**12.57
		SE	1.38	1.11	0.87	0.71	0.68	0.79	0.75	1.27	1.56
	Gender	B₁₁	** -2.38	** -1.93	** -1.47	** -1.01	-0.56	-0.10	0.36	0.82	1.28
		SE	0.65	0.53	0.42	0.36	0.36	0.42	0.41	0.65	0.78
	IEP	B₁₂	** -3.91	** -3.36	** -2.80	** -2.24	* -1.69	-1.13	-0.57	-0.02	0.54
		SE	1.23	0.98	0.77	0.64	0.66	0.80	0.73	1.27	1.54
	SES	B₁₃	-0.38	-0.16	0.05	0.27	0.49	0.71	0.92	1.14	1.36
		SE	0.79	0.63	0.51	0.45	0.47	0.57	0.54	0.87	1.05
	INRT	B₁₄	**3.39	**2.71	**2.02	**1.34	**0.66	-0.02	** -0.71	** -1.39	** -2.07
		SE	0.47	0.38	0.30	0.25	0.24	0.27	0.26	0.43	0.53

Note. * significant at 0.05 level ** significant at 0.01 level

The Impact of Starting Level

This section details the results of the second set of the research questions exploring whether the grade level that students are first to exposed to the SWH approach affected their science achievement.

Model U

Building on Model C in section 2, StartLev was added to the level-2 model. Since the science achievement scores were collected from third grade and to be consistent with section 2, the GRADE and GRADE² were centered at third grade to allow for interpretation of the intercept as the average science achievement score for students in grade three.

Level-1 Model: $SCIENCE_{it} = \pi_{0i} + \pi_{1i}*(GRD_3_{it}) + \pi_{2i}*(GRADE-3_{it})^2 + e_{it}$

Level-2 Model: $\pi_{0i} = \beta_{00} + \beta_{01}*(STARTLEV_i) + r_{0i}$
 $\pi_{1i} = \beta_{10} + \beta_{11}*(STARTLEV_i)$
 $\pi_{2i} = \beta_{20} + \beta_{21}*(STARTLEV_i)$

Mixed Model: $SCIENCE_{it} = \beta_{00} + \beta_{01}*STARTLEV_i + \beta_{10}*(GRADE-3_{it}) + \beta_{11}*STARTLEV_i*(GRADE-3_{it}) + \beta_{20}*(GRADE-3_{it})^2 + \beta_{21}*STARTLEV_i*(GRADE-3_{it})^2 + r_{0i} + e_{it}$ (4.27)

The multilevel model was estimated with results presenting Table 4.7. For easier interpretation of results, the value of StartLev (Never had SWH experience= 0, Started the SWH approach at high school= 1, Started the SWH approach at middle school= 2, Started the SWH approach at elementary school= 3, and Started the SWH approach at kindergarten= 4) was substituted into Equation 4.27.

Never had experience:

$$SCIENCE_{ti} = 201.14 + 13.26 * GRADE-3_{ti} - 0.20 * (GRADE-3_{ti})^2 \quad (4.28)$$

High school:

$$SCIENCE_{ti} = 199.04 + 15.04 * GRADE-3_{ti} - 0.37 * (GRADE-3_{ti})^2 \quad (4.29)$$

Middle school:

$$SCIENCE_{ti} = 196.94 + 16.82 * GRADE-3_{ti} - 0.54 * (GRADE-3_{ti})^2 \quad (4.30)$$

Elementary school:

$$SCIENCE_{ti} = 194.84 + 18.6 * GRADE-3_{ti} - 0.71 * (GRADE-3_{ti})^2 \quad (4.31)$$

Kindergarten:

$$SCIENCE_{ti} = 192.74 + 20.38 * GRADE-3_{ti} - 0.88 * (GRADE-3_{ti})^2 \quad (4.32)$$

The results showed that the intercept of science achievement scores in grade 3 was significant ($\beta_{00} = 201.14$, $p < .001$) but not significantly varied among the starting grade levels ($\beta_{01} = -2.10$, $p = .089$). The rate of change ($\beta_{11} = 1.78$, $p < .001$) and the curvature parameters ($\beta_{12} = -0.17$, $p < .001$) were significantly varied among starting grade levels.

Further, comparing to Model C, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 858.67$, $p < .001$). The level-1 variance component showing significant differences between each student's observed and predicted science achievement scores over time was slightly changed, $\sigma^2 = 422.66$. By adding grade levels when students were first exposed to the SWH approach to the level-2 model, the intercept variance decreased approximately ($[861.77 - 858.67] / 861.77$) 0.4% suggesting that student's starting level helped explain only 0.4% of the variation of science achievement scores across students.

The difference in deviance statistics between Model C and Model U, 116532-116508 = 24 is distributed as a Chi-square with five degrees of freedom. The likelihood ratio test was significant, $\chi^2(3) = 24, p < .01$, indicating that the science achievement trajectories were better predicted by adding the level-2 predictor student's starting level of SWH exposure.

Model V

Model V explored the impacts of student's starting time on student's science achievement while controlling for (grand-centering) the effects of gender, IEP status, SES, and individual's initial reading achievement level.

Level-1 Model: $SCIENCE_{ti} = \pi_{0i} + \pi_{1i}*(GRADE-3_{ti}) + \pi_{2i}*(GRADE-3_{ti})^2 + e_{ti}$

Level-2 Model: $\pi_{0i} = \beta_{00} + \beta_{01}*(CGENDER_i) + \beta_{02}*(CIEP_i) + \beta_{03}*(CSES_i) + \beta_{04}*(STARTLEV_i) + \beta_{05}*(CINREAD) + r_{0i}$

$$\pi_{1i} = \beta_{10} + \beta_{11}*(CGENDER_i) + \beta_{12}*(CIEP_i) + \beta_{13}*(CSES_i) + \beta_{14}*(STARTLEV_i) + \beta_{15}*(CINREAD_i)$$

$$\pi_{2i} = \beta_{20} + \beta_{21}*(CGENDER_i) + \beta_{22}*(CIEP_i) + \beta_{23}*(CSES_i) + \beta_{24}*(STARTLEV_i) + \beta_{25}*(CINREAD_i)$$

Mixed Model :

$$SCIENCE_{ti} = \beta_{00} + \beta_{01}*CGENDER_i + \beta_{02}*CIEP_i + \beta_{03}*CSES_i + \beta_{04}*STARTLEV_i + \beta_{05}*CINREAD_i + \beta_{10}***(GRADE-3_{ti}) + \beta_{11}*CGENDER_i*(GRADE-3_{ti}) + \beta_{12}*CIEP_i*(GRADE-3_{ti}) + \beta_{13}*CSES_i*(GRADE-3_{ti}) + \beta_{14}*STARTLEV_i*(GRADE-3_{ti}) + \beta_{15}*CINREAD_i*(GRADE-3_{ti}) + \beta_{20}*(GRADE-3_{ti})^2 + \beta_{21}*CGENDER_i*(GRADE-3_{ti})^2 + \beta_{22}*CIEP_i*(GRADE-3_{ti})^2 + \beta_{23}*CSES_i*(GRADE-3_{ti})^2 + \beta_{24}*STARTLEV_i*(GRADE-3_{ti})^2 + \beta_{25}*CINREAD_i*(GRADE-3_{ti})^2 + r_{0i} + e_{ti} \quad (4.33)$$

The multilevel model was estimated with results presenting in Table 4.7. For easier interpretation of results, the value of StartLev (Never had SWH experience=0, Started the SWH approach at high school=1, Started the SWH approach at middle school=2, Started the SWH approach at elementary school= 3, and Started the SWH approach at kindergarten=4) was substituted into Equation 4.33.

Never had experience:

$$SCIENCE_{ti} = 203.80 + 13.16 * GRADE-3_{ti} - 0.23 * GRADE-3_{ti} \quad (4.34)$$

High school:

$$SCIENCE_{ti} = 200.67 + 14.84 * GRADE-3_{ti} - 0.37 * GRADE-3_{ti} \quad (4.35)$$

Middle school:

$$SCIENCE_{ti} = 197.54 + 16.53 * GRADE-3_{ti} - 0.51 * GRADE-3_{ti} \quad (4.36)$$

Elementary school:

$$SCIENCE_{ti} = 194.41 + 18.21 * GRADE-3_{ti} - 0.65 * GRADE-3_{ti} \quad (4.37)$$

Kindergarten:

$$SCIENCE_{ti} = 191.27 + 19.90 * GRADE-3_{ti} - 0.79 * GRADE-3_{ti} \quad (4.38)$$

The results for the Model V were similar to Model U. Model V had a significant intercept for science achievement score in grade 3 ($\beta_{00}=203.80, p < .001$) which significantly varied among timing of exposure to the SWH approach ($\beta_{04}=-3.13, p = .004$). The negative β_{04} indicates that students who started the SWH approach at an earlier age had significantly lower grand mean at grade three. For example, the average student who never had SWH experience had predicted grand-mean science achievement score of 203.80 compared to 200.67 for those whose exposure occurred at high school. The rate of change was significant ($\beta_{10}=13.16, p < .001$) and varied among timing of exposure to the

SWH approach ($\beta_{14}=1.68, p < .001$). The positive β_{14} indicates that students who started the SWH approach at an earlier age had significant higher rate of change. For example, at grade three, students who never had SWH experience had the rate of change of science achievement scores of approximately 13.16 compared to the student who started the SWH approach in elementary school whose rate of change was 18.21. The curvature parameter at grade three was not significant ($\beta_{20} = -0.23, p = .102$) significantly varied among timing of exposure to the SWH approach ($\beta_{24} = -0.14, p = .007$).

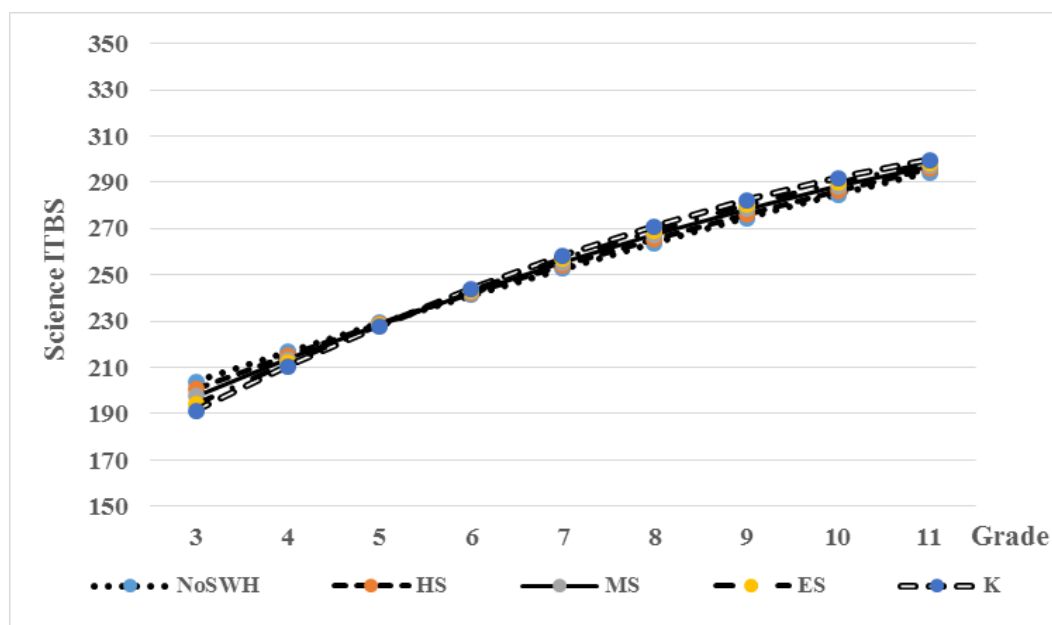
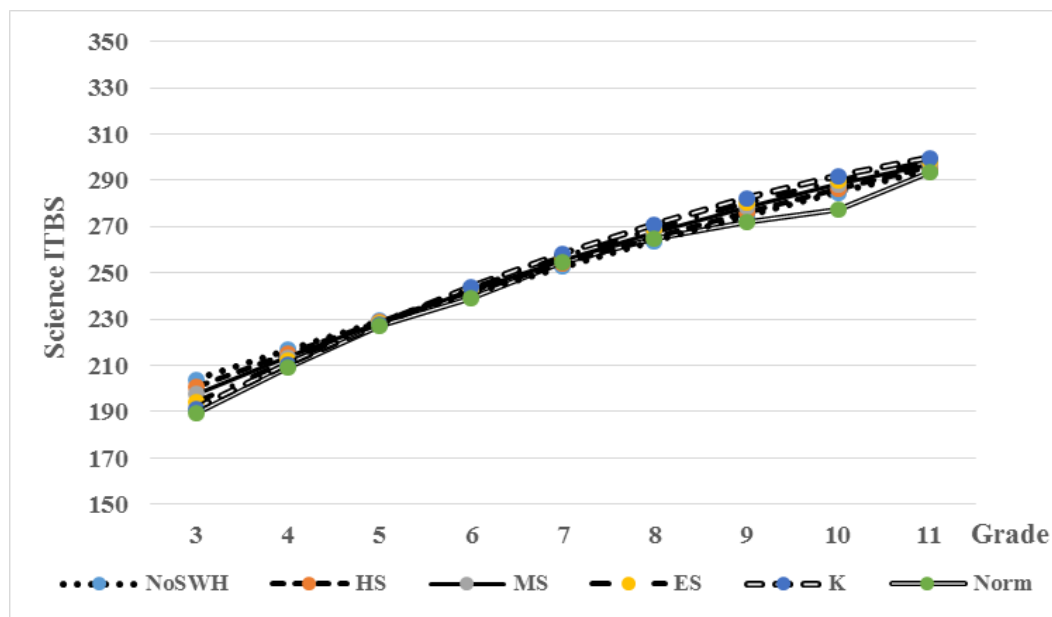


Figure 4.13 Science Achievement Trajectories for Different Timing of the First Exposure of the SWH Approach

Table 4.7 Results of Models Exploring First Timing of Exposure

			Model U	Model V
Fixed effect				
Composite model	Intercept	B_{00}	**201.14 (4.39)	**203.80 (4.57)
	Gender	B_{01}		-0.07 (1.65)
	IEP	B_{02}		1.43 (2.68)
	SES	B_{03}		-3.06 (1.77)
	InREAD	B_{04}		**15.51 (1.15)
	StartLev	B_{05}	-2.1 (1.23)	*-3.13 (1.30)
GRADE	Intercept	B_{10}	**13.26 (1.60)	**13.16 (1.55)
	Gender	B_{11}		** -2.33 (0.68)
	IEP	B_{12}		** -3.73 (1.14)
	SES	B_{13}		* -1.84 (0.78)
	InREAD	B_{14}		**4.21 (0.47)
	StartLev	B_{15}	**1.78 (0.50)	**1.68 (0.46)
GRADE ²	Intercept	B_{20}	-0.20 (0.15)	-0.23 (0.13)
	Gender	B_{21}		**0.27 (0.07)
	IEP	B_{22}		*0.31 (0.12)
	SES	B_{23}		*0.19 (0.09)
	InREAD	B_{24}		** -0.42 (0.06)
	StartLev	B_{25}	** -0.17 (0.05)	** -0.14 (0.05)
Random Effect				
Intercept		τ_0	858.67	371.12
level-1		σ^2	422.66	417.31
Deviance			116508	114201
Estimate parameters			7	7

Note. * significant at 0.05 level ** significant at 0.01 level



	3	4	5	6	7	8	9	10	11
NoSWH	203.8	216.73	229.2	241.21	252.77	263.86	274.5	284.68	294.4
Gain		12.93	12.47	12.01	11.56	11.09	10.64	10.18	9.72
HS	200.67	215.14	228.88	241.87	254.13	265.64	276.42	286.45	295.74
Gain		14.47	13.74	12.99	12.26	11.51	10.78	10.03	9.29
MS	197.54	213.56	228.56	242.53	255.49	267.42	278.33	288.22	297.09
Gain		16.02	15	13.97	12.96	11.93	10.91	9.89	8.87
ES	194.41	211.97	228.23	243.19	256.85	269.2	280.25	289.99	298.44
Gain		17.56	16.26	14.96	13.66	12.35	11.05	9.74	8.45
K	191.28	210.39	227.91	243.85	258.21	270.98	282.16	291.77	299.78
Gain		19.11	17.52	15.94	14.36	12.77	11.18	9.61	8.01
Norm	189.35	209.31	227.31	239.1	254.52	265.08	271.82	277.29	293.5
Gain		19.96	18	11.79	15.42	10.56	6.74	5.47	16.21

Figure 4.14 Science Achievement Trajectories for Different Timing of the First Exposure of the SWH Approach Compared to Iowa Normative Population

Further, the variance component estimate showed significant variation in science achievement scores at grade 3 ($\tau_{00} = 371.12$, $p < 0.01$). The level-1 variance component showed differences between each student's observed and predicted science achievement

scores over time, $\sigma^2 = 417.31$. By adding StartLev into the level-2 model controlling for gender, IEP, SES, and initial reading achievement, the intercept variance decreased approximately $([861.77-371.12]/861.77)$ 57% from Model C and $([1385.87-371.12]/1385.87)$ 73% from Model A suggesting that Model V helped explained 73% of the variation of science achievement scores across students.

The difference in deviance statistics between Model C and Model V, $116532-114201 = 2331$ is distributed as a Chi-square with twelve degrees of freedom. The likelihood ratio test was significant, $\chi^2(12) = 2331, p < .01$, indicating that the science achievement trajectories were better predicted by adding all level-2 predictors.

Starting Level Summary. All predicted achievement trajectories showed non-linear growth, consistent with the Iowa normative data. The predicted achievement trajectory of students whose timing of exposure to the SWH approach was as early as kindergarten were found to be closely aligned with the Iowa norm trajectory. It also showed highest year-to-year gain scores compared to the predicted trajectories of other groups. However, the predicted achievement of students who never had experience with the SWH approach was much higher than the Iowa normative data at third grade. Moreover, its year-to-year gain scores were also stable and did not drop as fast as found in Iowa norm and other predicted achievement trajectories

Summary

The multilevel model for changes with the shift of elevation and slope technique was utilized to examine the growth of science ITBS scores after students exposed to the SWH approach. The results showed the shift of elevation and slope of achievement

trajectories after the school adopting the approach indicating the effectiveness of the approach on students. However, not all students benefited equally. Disadvantaged students tended to have bigger improvement in their science ITBS scores. The results shed some light on long term effects of the SWH approach. The next chapter attempts to answer the research questions. Discussion of the results and implication will be elaborated too.

CHAPTER V

DISCUSSION AND IMPLICATIONS

This study aimed to examine the longitudinal impacts of the Science Writing Heuristic (SWH) approach as an argument-based inquiry approach on student's science achievement. Results from this study not only confirmed the positive impacts of the SWH approach on student reported by other related studies, but also highlighted the unique impacts of the long-term exposure of the approach.

This chapter begins with a discussion of the main findings addressing the three main research questions. Before addressing the research questions guiding this study, the reliability of the predicted models are discussed. Since the data structure in this study was complex, dealing with more than ten years of information, statistical models were used as a way to describe and simplify the complicated data, it was crucial for the researcher and readers to feel confident in the developed statistical models and their interpretations. Then results responding to each research question are discussed including their implications. Finally, the implication for future research and the limitations of this study are provided.

Research Questions

Research Question I

Research Question I-A: *Do students have better growth in their science test scores when they experience the SWH approach in science classroom compare to traditional science classroom before exposure the SWH?*

Overall, after students exposed to the SWH approach, student's science achievement significantly increased compare to prior to the SWH exposure. Moreover, the science achievement also continuously grew at faster rate when students exposed to the approach over time.

Research Question I-A: *Do students have better growth in their science test scores when they experience the SWH approach in science classroom compare to the Iowa normative population?*

The predicted achievement growth trajectories were non-linear, consistent with the Iowa normative achievement growth trajectory. At third grade, the SWH students had approximately the same average science score with the Iowa normative population; however, the SWH students' science scores increased at faster rate. As the results, at grade 11, the SWH students' science ITBS score was 5.2 point higher than the average score of the Iowa normative population.

Research Question I-B: *Do students have better growth in their science test scores when they experience the SWH approach in science classroom compare to the before adopting the approach?*

To further address the first research question whether the SWH approach impacts student science ITBS achievement compare to the period prior to adopting the SWH approach and also comparing to average scores of Iowa students in more details, the reliability of results and their interpretation are discussed below.

Research Question I: Reliability of the results

The data structure of this study was complicated since it related to the longitudinal professional development and ten years of students' ITBS scores in a school district. There were various factors that contributed to the complexity of the database, such as individual teachers' participation, individual students' experiences with the SWH approach, etc. However, interpretation of results was simplified by focusing on the predicted standardized test scores of students who started being exposed to the SWH approach at fourth grade and continuously exposed to the approach throughout their schooling experience. According to the multilevel model analysis (Singer & Willett, 2003, Peugh, 2010), there are numerous parameters used to demonstrate the reliability of the predicted models such as reliability estimate, significant reduced deviance statistic parameters, and percentage of variance explained. For this study, all parameters collectively indicated that the estimate models were statistically reliable. More importantly, the high consistency between predicted achievement growth trajectories from the analysis and achievement trajectories found in the Iowa normative achievement demonstrated the practical aspects of the predicted results. Combining statistical and practical aspects found within the estimated results, the researcher was confident that the predicted models suitably represented the phenomena.

Research Question I: The effectiveness of the SWH approach

In relation to the first research question, the researcher argues that regardless of student individual variables, the SWH approach significantly helped students improve their science ITBS scores by initially raising the score up and then continue to raise the

scores over the years after that. According to the predicted model (Model F), the results showed significance in discontinuity of both slope and elevation indicating that the SWH approach significantly altered the original growth trajectories of students prior to the SWH phase. The predicted science achievement trajectories found in this study were consistent with theoretical growth models purposed by Singer and Willett (2003) where discontinuity in elevation and slope were the key impacts of the intervention. Students who had experience with the SWH approach had their science ITBS scores initially boosted up 2.56 points which is considered a small effect (Cohen's $f = 0.14$). Moreover, students' science ITBS scores continuously grew at a better rate than the non-SWH group when they had longer exposures to the SWH approach over years; however, the effect was trivial (Cohen's $f = 0.05$). The small and trivial effect sizes showing the impact of the SWH approach on student science ITBS scores found in this study were smaller than effect sizes reported from other previous SWH approach-related studies. For example, Gunel, Hand, and Prain (2007) reported small to large effect sizes in the end-of-unit tests in studies in which teachers implemented the SWH approach in their science classrooms. However, it is important to note that the most critical differences leading to different results between this study and other SWH approach-related studies were the type of measurement and time frame of the study. Most of the previous research within the SWH approach context used teacher and/or researcher-generated measurements and usually were conducted within shorter time frames (3-6 months). Some studies were conducted in longer time frames (1-3 years). However, those studies could not track individual changes over time, instead they relied on changes in averaged scores. While, this study was

situated in a longer time frame (10 years) and was able to track individual changes assessing by the standardized tests.

The trivial and small effect sizes found in studies using standardized tests to evaluate the impacts of an inquiry-based intervention were generally expected. The small effect size found in this study was consistent with various published meta-analyses (Bredderman, 1983; Schroeder et al, 2007). Previous studies (Bredderman, 1983, Tretter and Jones, 2003) generally showed null to small impacts of an inquiry-based intervention on students standardized achievement test scores compared to the impacts found on tests that were developed by teachers and/or researchers. These small effects could be explained by the idea of multiplelevel-multifaceted assessment purposed by Ruiz-Primo, Shavelson, Halmilton, and Klein (2002). In general, the biggest magnitude impact of an inquiry-based intervention was expected where the measurements were closely aligned with actual implemented classroom practices and the magnitudes were decreasing as the distance of the assessment from the curriculum implemented increased. This study was involved with the standardized ITBS tests, which, according to the notion of multiplelevel-multifaceted assessment, fell into the furthest distance on the continuum. In this case, the ITBS tests administrated annually were not closely aligned with activities and concepts implemented in the actual classrooms. The science ITBS tests were expected to be the least sensitive measurement to measure the impacts of the SWH approach on students resulting in observable but small or even trivial impacts. Considering the ITBS is less sensible to classroom instruction, I, as the researcher considered the statistical significant impacts of discontinuity of both elevation and slope, even with small effect size, as highly significant in this context. To develop better

understanding of the impacts of the SWH approach on science achievement, two types of impact: initial and additive impact are discussed in more details.

Research Question I: Initial and Additive Impact

The significant discontinuity of both elevation and slope showed two distinctive functions; initial and additive impact, of the effects of the SWH approach on student science achievement. First, the initial impact demonstrated that student science achievement was raised up after students had had experience with the SWH approach. Second, the additive impact demonstrated that student science achievement scores continuously grew at a better rate over time compared to growth trajectory of students in the traditional classroom.

The results suggested that regardless of student individual variables, the SWH approach was effective for improving student scientific conceptual understanding as assessed by the ITBS tests. The impact of the SWH approach was not only found in helping students to raise their science ITBS scores initially, the impact was also found to be additive helping student achievement scores grow at a faster rate than normal when the students were immersed within the SWH approach for longer periods of time. According to the predicted model (Model F), the combination of both initial and additive impact predicted that at eleventh grade the average science ITBS score was approximately ten points higher than average science ITBS scores of students from the traditional science classroom. This meant that after being exposed to the SWH approach at third grade, student science achievement scores were gradually increasing. By the time students reached eleventh grade, the average science achievement score could potentially reach

approximately ten points or approximately one grade level above average science ITBS scores of students from the traditional science classroom.

While the initial impact of an inquiry-based intervention was usually reported by other researchers, the additive impact was not critically explored, which could be explained by the lack of longitudinal studies among the science education community. With the longitudinal data enabling the ability to track individual changes over a long period of time, this study allowed the researcher to observe the potential additive impact of the SWH approach on science achievement. Students and teachers needed time to adapt themselves to the new environment provided by a new approach. Results from related studies (Gunel, 2006; Cavagnetto, 2006) within the same school district had shown that teachers as much as, their students needed time to practice and develop understanding of the new approach in order to successfully shift their classroom orientation. Students who were shifted from traditional science classrooms to more active-learning classroom environments, also needed time and opportunities to adapt to a new set of skills requiring for new tasks and learning processes.

The initial impact of the SHW approach has been well documented by previous short-term studies (McDermott & Hand, 2010; Hand, Hohenshell & Prain, 2004; Nam, Choi & Hand, 2010). By engaging in the active construction and negation processes of questions, claims, and evidence, students were initially helped to develop better understanding of scientific concepts. The ideas of additive impact within the SWH approach studies arose when Hand, Hohenshell, and Prain (2004) found that the impact of non-traditional writing activities on student achievement increased over time when students had multiple experiences. By continuously being immersed within the new

approach from year to year, students started to develop understanding and required skills for the new learning processes. In results, the initial impacts were not only maintained, but also could increasingly grow over time.

However, the idea of additive impact of an inquiry-based intervention is contradictory to the results found in Breddermann's (1983) study. Breddermann (1983) suggested that the initial advantage from the activity-based science instruction could be maintained but not additive over time. Moreover, the initial impact might be diminished when the intervention was removed. In contrast, the results found in this study clearly demonstrated an additive impact of the SWH approach intervention when students were continuously exposed to the approach throughout their school experience. When students had opportunities to participate in an argument-based inquiry science classroom over the years, the initial impacts were not only maintained but also added up helping student achievement grow faster than normal. The additive impact found within the SWH approach studies would appear to show the strength of the approach and a critical function of continuously providing inquiry-based instruction experience for students throughout the years in order to promote better gain.

Research Question II

Research Question II: *Does the SWH approach have equal impact on students from different backgrounds?*

The SWH approach did not impact all students equally. The disadvantaged students' science achievement significantly increased and grew at faster rate than non-disadvantaged students.

To further address the second research question whether the SWH approach equally impacted science ITBS achievement among students from different backgrounds; gender, Individual Educational Program (IEP), Socioeconomic Status (SES), and initial reading achievement levels, the reliability of results and their interpretation are discussed below.

Research Question II: Reliability of the results

To answer the second research question, models were developed from a fundamental model used in answering the first research questions. The considered high reliability found in the fundamental model was carried on to these models. Moreover, the highly reduced deviance statistic parameters and high percentage of variance explained, associated with student individual variables; gender, IEP, SES, and reading achievement levels, were found when each variable was explored. More importantly, the high consistency between predicted achievement growth trajectories from the analysis and achievement trajectories found in Iowa normative population demonstrated the practical aspects of the predicted results. Combining statistical and practical aspects found within the estimated results, the researcher was confident that the predicted models suitably represented the phenomena.

Research Question II: Closing Achievement Gap

In relation to the second research question, the researcher argues that the SWH approach did not impact the participating students equally. In general, disadvantage students who usually had low science achievement had greater benefits from participating

in the SWH classroom. The SWH approach also positively impacted students who were usually proficient in science, however, the impacts were much smaller than the impacts found among disadvantaged students. Thus, the SWH approach can help reduce the persistent achievement gaps among disadvantaged and other students.

IEP and SES Achievement Gaps

The results showed that the expected science achievement trajectories of IEP students who were exposed to the SWH approach showed significantly large initial impact (Cohen's $f = 0.44$) by initially raising the score 14.76 points. The predicted achievement of IEP students also showed significant medium additive impact (Cohen's $f = 0.38$) by continuously growing at an additional rate of 2.44 points per year. That is, the IEP-students experienced 7.6 times bigger initial impact and 2.8 times bigger additive impact compared to non-IEP students who also had experience with the SWH approach.

For low SES-students who had experience with the SWH approach, the expected science achievement trajectories showed significantly medium initial impact (Cohen's $f = 0.39$) by initially raising the score 12.28 points. Then, the predicted achievement of low SES students became only trivial additive impact (Cohen's $f = 0.02$) by continuously growing at a better rate at 1.34 points per year. When compared to non-low SES students who also had experience with the SWH approach, the low-SES students experienced 7.9 times bigger initial impact and approximately the same additive impact.

Over all, implementing the SWH approach in school-wide science classrooms had significantly helped improve student science ITBS achievement scores although they were observable small impacts. However, when considering students' individual

variables, the bigger initial and additive impacts were found among disadvantaged students (IEP and low-SES). Though the bigger impacts were found among disadvantaged students, the non-disadvantaged students were not negatively impacted. The non-disadvantaged students still benefited from the approach however, with a smaller effect size. The higher initial and additive impacts found among low achieving students, over time, helped student achievement grow faster than traditional rate, resulting in the problematic existing achievement gaps being reduced.

Achievement Gaps by Gender

The results demonstrated that the existing achievement gap between males and females was eventually narrowed down as well. However, the impacts of the SWH approach on boys were different from girls but these impacts were not statistically different. Although the impacts were insignificant, the researcher found that the initial impact for male students was approximately 2.5 times larger but the additive impact for male students was 0.6 times smaller than those found in females. That is, boys had bigger initial increase; however, science achievement of girls grew at a faster rate. In results, the achievement gap was wider at elementary school and eventually closed at high school.

For this study, the insignificant impacts between boys and girls were reasonable since, the existing achievement gap between genders in Iowa normative achievement measured by the ITBS was usually small. According to average science ITBS scores of Iowa students, male students seemed to have marginally, less than 3.5 points (less than 2%), advantage over female students in elementary and middle school. However, the trend reversed in high school where female students had slightly better average science ITBS

achievement scores. Though other studies using other various measurements usually reported that female students were more at risk of low science achievement (Secker & Lissitz, 1999), results from this study showed that the girls did not appear to have the same problem. The results strengthened the idea the SWH approach as an argument-based inquiry usually impacted all students across the board, especially students who normally struggle in classroom.

Achievement Gaps by Reading Achievement Level

When student's reading achievement scores were considered, the results were similar to results found in other variables. The SWH approach had small impacts on students who were already proficient in reading measured by the ITBS. The expected science achievement trajectories of these students showed a marginal initial impact by initially raising the score only 0.98 points then, the predicted science achievement continuously growing at insignificant better rate at only 0.33 points per year. The results were consistent with previously found within non-IEP and non-low SES students. The results strengthen the idea that the SWH approach had small positive impacts on students who already had done well in science ITBS tests. However, when examining the impacts on students in 13% top and 13% bottom end of reading achievement continuum, the predicted achievements were significantly impacted. The results showed that the SWH approach significantly initially raised science achievement scores of students who had extremely low reading achievement level (percentile rank below 26th) by 14.32 points, approximately 1 grade level, comparing to non-SWH group. But the SWH approach negatively impacted highly proficient students by initially bringing down their science

achievement scores by 5.69 points comparing to non-SWH group, approximately one-fourth of a grade level.

The additive impact for these two extreme groups of students were also different but they were not statistically significant. While, students with extreme below proficient reading level had negative additive impact which mean that their initially boosted up scores insignificantly decreased by 0.87 points over the years after that, the additive impact of the highly proficient group was positive but insignificant. This mean that their initial decreased science scores were picked up over time by continuously increasing 0.94 point per year for the years after engaged with the SWH approach.

Combing the initial and additive impacts, the results demonstrated that the existing achievement gap found between extremely high reading achieving students and extremely low reading achieving students were narrowed down by approximately five points (approximately 10%). Since the additive impacts were insignificant, the narrowed gap was not significantly reduced over time but rather was maintained. Comparing to the achievement gap of Iowa students, the existing achievement gaps between disadvantaged and non-disadvantaged students were usually wider as students moved to higher grades. Thus, extremely low reading achieving students showed significant science achievement improvement. The science achievement gap could be narrowed down and more importantly it was maintained.

Research Question III

Research Question III: *Does grade levels when students start experiencing with the SWH approach have impact on student science ITBS achievement?*

To address the third research question whether the time that students started experiencing with the SWH approach impacted on science ITBS achievement. The reliability of results and their interpretation were discussed below.

Research Question III: Reliability of the results

To answer the third research question, the StartLev variable was created. StartLev theoretically was a level-three variable where a group of students (cohort) first started to be exposed to the SWH approach. However, since there were limited numbers of cohort to effectively conduct a three-level analysis, the analysis was carried out by using StartLev as level-two variable. The analysis of three-level and two-level multilevel model showed equivalent results. One advantage of two-level model analysis over three-level model was that two-level model could allow simpler interpretation. While the reduced deviance statistic demonstrated that StartLev statistically significantly predicted student science achievement score, StartLev helped explained only 0.4% of the variation of science achievement scores across students. The results also showed wide range, approximately 12.5 points, of predicted average science scores at third grade. It was unusual that third grade students from the same school district could have an average score 12.5 points apart from each other. Though, the results showed statistical significant models, this wide range of predicted science achievement scores at third grade required cautions in attempting to interpret these results.

Research Question III: The Impacts of Starting Point

After controlling for gender, IEP, SES, and reading achievement level, the results indicated that the grade levels that students first started experiencing the SWH approach as an argument-based inquiry had significant impacts on student science ITBS achievement trajectories. Students who first start to experience the SWH approach as early as kindergarten had the highest rate of achievement growth. The predicted science achievement scores of kindergarten students grew at the highest rate at 19.90 points per year at third grade which was 1.5 times faster than students who never had experience with the SWH approach. It was also about 1.3 times faster than students who started experiencing the approach at the high school level and about 1.1 times faster than students who started experiencing the approach at the elementary school level. With higher growth rate, the model predicted that at eleventh grade, students who had experience with the SWH approach throughout their schooling would have predicted a science ITBS score approximately four points or about a half grade level higher than predicted scores of students who had experience with the SWH approach only in high school.

Discussion

Given new emerging movements of the new framework of activities for scientists and engineers (NRC, 2012) and the Next Generation Science Standards (NGSS) (NRC, 2013), the results found in this study highlighted the importance of the SWH approach, an immersion-oriented intervention of argument-based inquiry (Cavagnetto, 2010), as a critical vehicle for students to develop scientific understanding within the next generation of science in the schools. The immersion approach of argument-based inquiry, as well as

the new framework, did not see scientific argument as separated skills, instead, scientific argumentation is the way to do science. Argumentation has to be embedded throughout scientific activities as students develop or refine questions for investigation, designed experiments, collect and interpret data, and construct and defend claims based on their evidence (Cavagnetto, 2010; NRC, 2012). Moreover, both the SWH approach and the NGSS critically emphasize on the interdisciplinary learning and the developing of big ideas, instead of broader scientific facts.

By shifting student's role from passive learning in the traditional classroom to an active learners in the argument-based inquiry, students' learning is enhanced. There are five key characteristics of the SWH approach as argument-based inquiry possibly contributing to the enhancement in student's learning. First, the SWH approach as well as the NGSS encourage teachers to shift the goals of science classroom from delivering factual contents to actively constructing understanding of scientific concepts and big ideas. Students are given more times and opportunities to explore and develop understanding of core ideas in depth, rather than remember all the factual knowledge. Second, the SWH classroom creates non-threatening learning environments where individual ideas were equally valued. By moving away from a *right* or *wrong* answer, students no longer feel threaten or judged by teachers or peers. Every idea is encouraged and can contribute to learning community. Third, students are given opportunities and encouraged to take control of their own learning by generating their own questions, investigations, evidence, and claims. By taking control of their learning, students view classroom activities as learning tools to develop their own understanding contributing to their own interests and learning intention (Wheatley, 1991), rather than just the tasks that

teachers ask them to do. In addition, the SWH approach highlights the importance of prior knowledge that students bring into classroom. Teachers elicit students' prior knowledge through concept mapping, questions, or discussion in group or individual. This helps teachers design classroom intervention that build from where students are and connect existing knowledge to new understanding in a meaningful way.

Lastly, negotiation and discussion are critical elements of the SWH approach where students are expected to engage in a continuous cycle of negotiating and clarifying meaning and explanation with their peers and teachers (Cavagnetto & Hand, 2012). Students are explicitly asked to make connections between their questions, claims and evidence and be able to defend their claims based on evidence. These activities are highly consistent with the argumentative model conceptualized by Ford (2008) highlighting the interplay between *construction* and *critique* of knowledge claims in both scientific reasoning and practice. Students in the SWH classroom are encouraged to move between *construction* and *critique* elements where students as a learning community *construct* their own claims based on gathering evidence. When presenting a new knowledge claim, students *critique* the knowledge claim by comparing the presenting claim with their own claim and seeking for errors in the claim or links between claim and evidence. By engaging in *construction* of their own knowledge claims, making connection between claims and evidence, and more importantly, *critiquing* peers claims, as a result, students could potentially develop deeper understanding of scientific concepts.

By creating learning processes which focus on prior knowledge, non-threatening environments, letting students control their learning, and continuous negotiation, the SWH approach as an immersion-oriented intervention of argument-based inquiry helps

enhance student science learning. As discussed in the results section, the results confirmed the benefits of inquiry-based intervention on students found in previous studies. Moreover, the results also revealed two new critical elements of benefits of the SWH approach emerging from this study; initial and long-term impact. After exposing to the SWH approach, all students benefited from the immersion-oriented instruction of argument-based inquiry, however, the impacts could be different across individual students.

First, disadvantaged students who usually were at risk of having low achieving in science had higher initial and long-term impacts than non-disadvantaged students. By engaging in scientific argumentation in classroom, disadvantaged students saw high initial increasing in their science achievement and higher growth rate when they continuously engaged in the argument-based instruction. These results were not new. Secker (2002) found that the instructional choices that teachers implemented in classroom do not impact all students equally, even with same classroom. Science achievement could be influenced by the demographic profiles of individual students. The bigger impacts found among disadvantaged students when exposed to an inquiry-based instruction in science in classroom were well published. Previous studies including other short-term the SWH-related studies (Kingir, Geban, & Gunel, 2012, Villanueva, Taylor, Therrien, & Hand, 2012) has showed similar results. These studies showed that the SWH approach encouraged teachers to shift their attention from heavily relying on transferring content to actively constructing understanding of scientific concepts and big ideas. For disadvantaged students, big ideas can be used to connect scientific concepts between formal and informal learning settings in meaningful ways. Moreover, by creating non-

threatening learning environments where individual ideas were valued and encouraged, disadvantaged students can engage in an active construction of knowledge through constant negotiation of ideas with peers and selves. With the SWH approach, the disadvantaged students are able to better develop big ideas, and scientific concepts through negotiation and argumentation using the SWH approach. As a result, the achievement gaps among disadvantaged and non-disadvantage students are closing.

Furthermore, students who were normally proficient in science saw positive impacts too; however, the scale of impacts were smaller than those found among disadvantaged students. This could be explained by the fact that the SWH approach may have little impacts on students who are already proficient in science or there is smaller room for science achievement to grow comparing to disadvantaged groups. In addition, high proficient students tended to see initial a dip of their science achievement. The dipping scores could be explained by the regression to the mean as a natural feature of the test or the approach was negatively impacts these group of students. However, the impacts of the immersion approach were more evident when these group of students had more opportunities and continuously engaged with the approach over a longer time. Traditional science classes tend to require students to remember factual knowledge, rather than develop understanding of a concept (Villanueva, Taylor, Therrien, & Hand, 2012). When shifting from traditional classrooms to the SWH classroom, high achieving students might have troubles coping with new classroom environments and struggle when the focuses of classroom were initially shifted from remembering the facts to building understand big ideas. However, when more opportunities were provided long term, these students were able to pick up new required skills and reach the same level as they had.

Future research are needed to confirm this phenomenon. Some adjustment might be needed to help teachers implementing the SWH approach that equally enhance student learning for all, rather than certain groups.

Finally, the results started to highlight the differential impacts of the SWH approach on students in different ages. The impacts on student's science achievement seemed to be bigger when students started engaging with argument-based inquiry at younger ages. Even though, the results of the developed model must be interpreted with caution, the results started to differentiate the impacts of the same approach on students with different ages. Students who started as young as kindergarten seem to have greater benefits from the argument-based inquiry. One possible explanation could be that first, younger students had engaged more in their classroom, resulting in better achieving growth. The SWH classrooms encourage students to engage in their own learning, requiring students to constantly negotiate with themselves and peers. Marks (2000) found that student academic engagement was higher in younger students. Engagement in the classroom could lead to higher achievement and contribute to students' social and cognitive development (Finn, 1993; Newmann, 1992). Second, the additive impact might play important roles in explaining the results. When the starting point was earlier, younger students had more time and opportunities to develop required skills and to be familiar with the approach. While, older students had less time and opportunities and had to go through a radical shift from a traditional classroom to an inquiry-based classroom. The additive impact over the years could contribute to higher growth rate found in students who started exposure as early as kindergarten level. There are other possible explanations e.g., high school teachers may have more pressure to delivery more contents

than elementary school teachers or some skills corresponding to the SWH approach could be developed in only early age. More research is needed to explore these issues.

Implications

Activities and learning environment provided in the SWH classroom could contribute to student learning as Wise (1996) suggested that if students were placed in a classroom where instruction was connected to students' interests and present understanding and they had opportunities to experience collaborative scientific inquiry, achievement will be accelerated. The data suggested that after the participating school district had adopted the SWH approach as the main approach in the science classroom, the overall student science achievement assessed by the ITBS tests had steadily increased compared to prior to adopting the SWH approach. The impacts of the SWH approach were also found to be bigger among disadvantaged students. The results suggested that students who usually had low science achievement could see better growth in their science achievement after exposure to the approach. The SWH approach offers better and more opportunities for students to engage in authentic scientific investigations and negotiation. Students are encouraged to generate claims based on evidence gathering from investigation. Negotiation and argumentation are crucial aspects of the approach. Students are required to self-negotiate and publicly-negotiate when generating questions, designing investigation, collecting data, interpreting results, generating claims, and reflecting on ideas. In the SWH classroom, student ideas are constantly challenged by teachers, peers, and the students themselves through discussion.

Moving on to the Next Generation of Science Standard

The results found in this study were critically encouraging at this period when the science education community is facing new movements in the Next Generation Science Standards proposed by NRC (2013). The framework of the SWH approach is highly aligned with the framework of the three spheres of activity for scientists and engineers put-forward by NRC (2012). The framework highlights argumentation as central activity, constantly interacting with other spheres. The framework proposed by NRC (2012) and the SWH approach highlighted the importance of scientific argumentation by requiring students to use critical thinking skills to argue, critique, and analyze in order to identify strengths and weaknesses and to develop or refine ideas, design, modeling, and explanation. Moreover, the new standards also focus on a smaller but deeper set of scientific core ideas and their application of content. The SWH approach pushed teachers to shift the classroom focus from delivering factual contents to constructing scientific understanding around big ideas as well. As a results, students are able to develop deeper understanding about scientific ideas by holding on to big ideas and their real-life applications. The results from this study encourage the researcher to feel confident that the SWH approach is well fitted in the new framework and the next generation of science standards. The SWH approach can help teachers and students moving forward in the next generation of science standards by steadily increasing student science achievement and narrowing down the existing problematic achievement gaps.

The Sustainability of the School-Wide Adoption of the SWH Approach

The context of this study was very unique, dealing with longitudinal data prior, during, and after the SWH professional development program of one school district. The results found in this study start to shed some light on the sustainability impact of the SWH professional development program. The professional development projects are not generally expanded into a successful school reform. The impacts of an intervention usually decrease and are not sustainable after support from professional developers or researchers are removed (Johnson, Fargo, & Kahle, 2010). Previous studies by Omar (2004), Gunel (2006), and Cavagnetto (2007) had demonstrated the impacts of the SWH approach on teachers and students during the professional development program taking place. With supports from the staffs and researchers and time for practicing, some participating teachers had successfully shifted from using the traditional approach to the SWH approach. However, the teachers' changes were not linear. In fact, the reported changes were complicated varying by individuals. Each teacher used different paths and pace to shift. Moreover, the SWH approach had been reported as having positive impacts on student's assessment by discussion participation, critical thinking test, teacher-generated tests, ITBS tests, and ITED tests. However, there was no evidence of how teachers and students progress after the program had ended where supports from staffs and researchers were removed.

For this study, the ITBS data was collected from 2000 to 2011 covering all phases of the SWH professional development program including pre, during, and post phase. Regardless of level of implementation of the individual teacher, in general, all teachers in the school district have been collectively adopting the SWH approach. The collective participation of groups of teachers from the same school is one of the important focuses

of professional developments leading to success (Desimone, Porter, Garet, Yoon, & Birman, 2002). In general, the positive impacts of the SWH approach have been maintained even after all supports had been removed. The school district reported regular changes of teachers and staffs. Teachers moved in and out from the school district, however, since the SWH approach was promoted as the main approach in teaching science, all new teachers had to have some kind of training from teachers in the school district who had more experience with the approach. As a result, the impacts on the student science achievement could be sustained and observed even if the program had been finished.

The Impact of the SWH Approach on Standardized Test

One critical implication of this study was the significant impacts of the SWH approach as an argument-based inquiry on student standardized test scores. Even though standardized tests could ensure highly valid and reliable results, their negative impact on science reform had been criticized. When the standardized test scores were used to make high-stake decisions, it has been reported that teachers were pushed to explicitly teach to the tests as the main method to raise the scores. This phenomena is contradictory to the reform efforts that encourage science teachers to use assessments that are closely aligned with the instructional goals (Aydeniz, 2007; Madden, 2008). The results demonstrated that the SWH approach had significantly helped students improve their science achievement, especially for at-risk students. However, the ITBS tests used in this study was only one measurement assessing student science content learning. To comprehensively evaluate the impacts of an intervention, teachers and researcher must

gather evidence and information from different measurement tools, such as classroom notes, teacher or student interview, classroom grade, etc.

Implications for Future Research

The purpose of this study was to explore how the SWH approach affected student science ITBS achievement in the long-term using Hierarchical Linear Model (HLM). The results have demonstrated the positive impacts of the approach on students by initially raising the achievement scores up and continuously helping the scores grow faster. This study also showed the SWH approach did not affect all students equally. Disadvantaged students seemed to benefit more from participating in the SWH classrooms. The SWH approach also helped narrow the achievement gaps among at-risk students; females, IES status, social economic backgrounds, and reading achievement levels. However, due to the homogenous context of student's ethnicity, this study could not effectively explore the impacts of the approach on students with different racial backgrounds. One implication for future research is to explore whether the SWH approach can help narrow down achievement gaps among racial backgrounds as well. Moreover, the contexts of each school are unique. The participating school district had less than 25% of low-SES and IEP students. And its science achievement was usually higher than Iowa normative population. The researcher is keen to explore the longitudinal impacts of the SWH approach on other schools with broader contexts, such as low-achieving schools, low-poverty schools, schools for gifted students, or school with different cultural backgrounds.

Although this study addressed the general impacts of the SWH classrooms on student achievement, it failed to establish longitudinal relationship between teacher's implementation of the approach and student achievement. Future research with teacher's factors are needed to be conducted to understand how teacher's implementation impacts student achievement in the long-term. For example, what are the differences of initial and additive impacts between low and high implementation? How do teacher's changes affect student's growth over year?

In addition, this study started to explore the function of the age of students when they started having experience with the SWH approach. With constraints of the interpretation, the results started to show significant different impacts of students at different ages. Future research with better design can be conducted to unpack this issue by increasing the number of cohorts to allow more appropriate three-level analysis.

Another issue that should be addressed in future research is the other impacts of the SWH approach on students. Though the SWH approach focuses in helping teachers and students teach and learn science, the argumentation skills could be transferred to other subject areas. The parallel-design study could be conducted in the future to explore the impacts of the SWH approach on student achievement in other subject areas such as reading or mathematics. Moreover, the results showed that students, especially at-risk students, benefited from the SWH approach. Future research could explore the changes of student attitude toward science and career choices as a result of having better science achievement.

Finally, this study demonstrated the sustainability of the SWH approach within the participating school district. The impacts of the SWH approach have been showed even after the SWH professional development had ended for years. Internal professional development has been employed in the school to help train new teachers. In order to understand the functions and impacts that the internal professional development has on new teachers and their students, future research needs to be conducted.

Limitations

The researcher recognizes the limitations of design and methodology surrounding this study. First, there was the lack of uniformity of the ITBS and student's demographic data. Since, this study dealt with over ten years of exiting data from a school district, missing data due to the typical movement, in and out, of students during this period was very common. Moreover, the ITBS tests have only been fully administrated in all batteries since 2003, thus some of ITBS test batteries might not be administrated at some level in some years. For example, there was neither data of science scores of elementary students prior to 2003, nor reading ITBS scores in 2001 and 2002. In addition, as mentioned in the context of this study, prior to 2010, all students who finished 5th grade in this school district had to join 6th grade in another school district. Thus, 6th grade data of students was not accessible. However, one of the strengths of the multilevel analysis is that there are statistic procedures that could be employed to ensure the validity and reliability of results.

Second, due to the complexity of data, some assumptions had to be made to simplify data structure. For example, the researcher assumed that all students in the same

cohort had the same amount of experience with the SWH approach though, some students might move in the school district during the school year. Thus, the model could not recognize all variability of individual students. Third, because the majority of students in this school district were Caucasians, there was not enough sampling of students to explore the impact of the SWH approach on achievement gap by race.

More importantly, there was the lack of information about teacher's implementation. Scholars agree that what actually happened in each classroom played critical role in how students learn. Moreover, as found in previous researches (Omar, 2002; Gunel, 2006; Cavagnetto, 2006), teachers had gradually changed over time after adopting the SWH approach. However, the models in this study could not recognize variations of individual teacher implementation factors across classes, and times. Analysis without these factors limited the researcher to draw a general conclusion about relationships between teacher implementation or teacher professional growth and student achievement growth.

Finally, this study explored the impact of the SWH approach on students in one voluntary school district. Though, the location and context of this school district could be considered typical for Iowa, contexts might not be able to represent contexts of other schools in Iowa, states, or countries. With this constraint, it is crucial for the researcher and readers to be careful not to overgeneralize the results beyond the limited context.

REFERENCES

- Abd-El-Khalick, F. (2004). Introduction. In F. Abd-El-Khalick, S. Bougaoude, N. Lederman, R. Mamok-Naaman, A. Hopstein, M. Nioz, D. Treagrest, & H. Tusan (Eds.), *Inquiry in science education: International perspective* (pp. 397–400). *Science Education*, 88, pp. 397–419. doi:10.1002/sce.10118
- Asay, L. D., & Orgill, M. (2010). Analysis of essential features of inquiry found in articles published in *The Science Teacher*, 1998–2007. *Journal of Science Teacher Education*, 21(1), 57-79. doi:10.1007/s10972-009-9152-9
- Aydeniz, M. (2007). Understanding challenges to the implementation of assessment reform in science classrooms: A case study of science teachers' conceptions and practices of assessment. Unpublished Dissertation. Florida State University, Tallahassee, FL.
- Aydeniz, M., & Southerland, S. A. (2012). A national survey of middle and high school science teachers' responses to standardized testing: is science being devalued in schools?. *Journal of Science Teacher Education*, 23(3), 233-257. doi:10.1007/s10972-012-9266-3
- Berland, L. K., and Reiser, B. J. (2009). Making sense of argumentation and explanation. *Science Education*, 93(1), 26–55. doi:10.1002/sce.20286
- Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research*, 53(4), 499-518. doi: 10.3102/00346543053004499
- Brickhouse, N. W. (2006). Celebrating 90 years of science education: Reflections on the gold standard and ways of promoting good research. *Science Education*, 90(1), 1–7. doi: 10.1002/sce.20128
- Bybee, R. W., Taylor, J. A., Gardner, A., Van Scotter, P., Carlson Powell, J., Westbrook, A., & Landes, N. (2006). The BSCS 5E instructional model: Origins and effectiveness. *Colorado Springs, CO: BSCS*.
- Cavagnetto, A. R. (2006). *Setting the Question for Inquiry: The Effects of Whole Class Vs Small Group on Student Achievement in Elementary Science* (Doctoral dissertation). Retrieved from ProQuest.
- Cavagnetto, A. R. (2010). Argument to Foster Scientific Literacy A Review of Argument Interventions in K–12 Science Contexts. *Review of Educational Research*, 80(3), 336-371. doi: 10.3102/0034654310376953
- Cavagnetto, A., & Hand, B. (2012). The importance of embedding argument within science classrooms. In Khine, M. S. (Ed.), *Perspectives on Scientific Argumentation* (pp. 39-53). Springer Netherlands.

- Chanlen, N., & Hand, B. (2011). The impacts of the science writing heuristic (SWH) approach on students critical thinking skills. Paper presented at the International Meeting of the Association for Science Teacher Education, Minneapolis, MN
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational evaluation and policy analysis*, 24(2), 81-112. doi: 10.3102/01623737024002081
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, 84(3), 287-312. doi: 10.1002/(SICI)1098-237X(200005)84:3<287::AID-SCE1>3.0.CO;2-A
- Finn, J. D. (1993). *School engagement and students at risk*. Washington, DC: National Center for Education Statistics.
- Ford, M. J. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, 92(3), 404-423. doi:10.1002/sce.20263
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis. *Review of Educational Research*, 82(3), 300-329. doi:10.3102/0034654312457206
- Geier, R., Blumenfeld, P. C., Marx, R. W., Krajcik, J. S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, 45(8), 922-939. doi:10.1002/tea.20248
- Gilbert, J. K., Boulter, C., & Rutherford, M. (1998). Models in explanations, Part 1: Horses for courses?. *International Journal of Science Education*, 20(1), 83-97. doi:10.1080/0950069980200106
- Greene, J. P., & Winters, M. A. (2003). Testing high stakes tests: Can we believe the results of accountability tests? (Report 33). New York: Manhattan Institute, Center for Civic Innovation.
- Grimberg, B. I., & Hand, B. (2009). Cognitive Pathways: Analysis of students' written texts for science understanding. *International Journal of Science Education*, 31(4), 503-521. doi:10.1080/09500690701704805
- Gunel, M. (2006). Investigating the impact of teachers' implementation practices on academic achievement in science during a long-term professional development program on the Science Writing Heuristic. Unpublished PhD, Iowa State University. Ames.

- Gunel, M., Hand, B., & Prain, V. (2007). Writing for learning in science: A secondary analysis of six studies. *International Journal of Science and Mathematics Education*, 5(4), 615-637. doi:10.1007/s10763-007-9082-y
- Hand, B. (2008). Introducing the science writing heuristic approach. In B. Hand (Ed.), *Science inquiry, argument and language: A case for the science writing heuristic*. Rotterdam, The Netherlands: Sense.
- Hand, B. M., Prain, V., & Yore, L. (2001). Sequential writing tasks' influence on science learning. In Tynjala, P., *Writing as a Learning Tool* (pp. 105-129). Springer Netherlands.
- Hand, B., & Keys, C. (1999). Inquiry investigation: A new approach to laboratory reports. *The Science Teacher*, 66, 27-29.
- Hand, B., Hohenshell, L., & Prain, V. (2004). Exploring students' responses to conceptual questions when engaged with planned writing experiences: A study with year 10 science students. *Journal of Research in Science Teaching*, 41(2), 186-210. doi:10.1002/tea.10128
- Harrison, A. G., & Treagust, D. F. (2000). A typology of school science models. *International Journal of Science Education*, 22(9), 1011-1026. doi:10.1080/095006900416884
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.
- Jimenez-Aleixandre, M. P., and Erduran, S. (2007). *Argumentation in science education: An overview*. In S. Erduran & M. P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–28). Dordrecht, Netherlands: Springer.
- Johnson, C., Fargo, J., & Kahle, J. B. (2010). The cumulative and residual impact of a systemic reform program on teacher change and student learning of science. *School science and mathematics*, 110(3), 144-159. doi:10.1111/j.1949-8594.2010.00017.x
- Karplus, R. and Their, H. D. (1967). *A new look at elementary school science*. Chicago, IL: Rand McNally.
- Kingir, S., Geban, O., & Gunel, M. (2012). How does the science writing heuristic approach affect students' performances of different academic achievement levels? A case for high school chemistry. *Chemistry Education Research and Practice*, 13(4), 428-436. doi: 10.1039/C2RP20013A
- Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. *Journal of Research in Science Teaching*, 43(7), 607-636. doi:10.1002/tea.20141

- Lemke, J. (2004). The literacies of science. *Crossing borders in literacy and science instruction: Perspectives on theory and practice*, 33-47. Retrieved from <http://jaylemke.squarespace.com/storage/Literacies-of-science-2004.pdf>
- Loyd, B. H., Forsyth, R. A., & Hoover, H. D. (1980). Relationship of elementary and secondary school achievement test scores to later academic success. *Educational and Psychological Measurement*, 40: 1117–1124.
- Luft, J.A. (2001). Changing inquiry practice and beliefs: The impact of an inquiry-based professional development programme on beginning and experienced secondary science teachers. *International Journal of Science Education*, 23(5), 517-534. doi:10.1080/09500690121307
- Madden, D. B. (2008). Middle school educators' perspectives of changes in teaching practices caused by implementation of the FCAT Science. Unpublished Masters Thesis, Florida State University, Tallahassee, FL.
- Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American educational research journal*, 37(1), 153-184. doi:10.3102/00028312037001153
- Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39(1), 17-38. doi:10.1007/s11165-007-9072-7
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47(4), 474-496. doi:10.1002/tea.20347
- Moje, E. B. (2007). Developing socially just subject-matter instruction: A review of the literature on disciplinary literacy teaching. *Review of research in education*, 31(1), 1-44. doi:10.3102/0091732X07300046001
- Muthen, B. O. (1991). Analysis of longitudinal data using latent variable models with varying parameters. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions*. Washington, DC: American Psychological Association.
- Muthen, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research*, 22(3), 376-398. doi:10.1177/0049124194022003006
- National Research Council (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: National Academy Press.
- National Research Council (2013). *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

- National Research Council (Ed.). (1996). *National science education standards*. National Academy Press.
- National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. National Academies Press.
- Newmann, F. M. (Ed.). (1992). *Student engagement and achievement in American secondary schools*. New York, NY: Teachers College Press.
- Omar, S. (2004). In-service teachers' implementation of the Science Writing Heuristic as a tool for professional growth. Unpublished PhD, Iowa State University. Ames.
- Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation*, 9(6), 1-12. Retrieved, January 19, 2013, from <http://pareonline.net/>
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48(1), 85-112. doi:10.1016/j.jsp.2009.09.002
- Prain, V., & Hand, B. (1996). Writing for learning in secondary science: Rethinking practices. *Teaching and Teacher Education*, 12(6), 609-626.
- Raudenbush, S. W., & Xiao-Feng, L. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological methods*, 6(4), 387. doi:10.1037/1082-989X.6.4.387
- Riverside Publishing (2003). *The Iowa Tests: Guide to research and development*. Itasca, IL: Riverside Publishing.
- Rosemier, R. A. (1962). An investigation of discrepancies in percentile ranks between a grade eight administration of *ITBS* and a grade nine administration of *ITED*. Unpublished study, Iowa Testing Programs, Iowa City, IA.
- Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39(5), 369-393. doi:10.1002/tea.10027
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436-1460. doi:10.1002/tea.20212
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific

- modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632-654. doi:10.1002/tea.20311
- Schwarz, C., & White, B. (2005). Metamodeling knowledge: Developing students' understanding of scientific modeling. *Cognition and Instruction*, 23(2), 165 – 205. doi:10.1207/s1532690xci2302_1
- Secker, C. V. (2002). Effects of inquiry-based teacher practices on science excellence and equity. *The Journal of Educational Research*, 95(3), 151-160. doi:10.1080/00220670209596585
- Shelley, M., Gonwa-Reeves, C., Baenziger, J., Seefeld, A., Hand, B., Therrien, W., & Taylor, J. (2012). Multilevel Models for Estimating the Effect of Implementing Argumentation-Based Elementary Science Instruction. *Society for Research on Educational Effectiveness*.
- Shymansky, J. A., Kyle, W. C., & Alport, J. M. (1983). The effects of new science curricula on student performance. *Journal of Research in Science Teaching*, 20(5), 387-404.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford university press.
- Snow, R.E. (1974). Representative and quasi-representative designs for research in teaching. *Review of Educational Research*, 44, 265–291.
- Spybrook, J., Raudenbush, S. W., Liu, X. F., Congdon, R., & Martínez, A. (2006). Optimal design for longitudinal and multilevel research: Documentation for the “Optimal Design” software. *Survey Research Center of the Institute of Social Research at University of Michigan*.
- Therrien, W. J., Hughes, C., & Hand, B. (2011). Introduction to special issue on science education and students with learning disabilities. *Learning Disabilities Research & Practice*, 26(4), 186-187. doi:10.1111/j.1540-5826.2011.00339.x
- Tretter, T. R., & Jones, M. G. (2003). Relationships Between Inquiry-Based Teaching and Physical Science Standardized Test Scores. *School Science and Mathematics*, 103(7), 345-350. doi:10.1111/j.1949-8594.2003.tb18211.x
- Tymms, P. (2004). Effect sizes in multilevel models. Schagen, I. & Elliot, K. (Eds.), *But what does it mean: The use of effect sizes in educational research* (pp. 55-66), Londres, UK: National Foundation of Educational Research..
- Villanueva, M. G., & Hand, B. (2011). Science for all: engaging students with special needs in and about science. *Learning Disabilities Research & Practice*, 26(4), 233-240. doi:10.1111/j.1540-5826.2011.00344.x

- Villanueva, M. G., Taylor, J., Therrien, W., & Hand, B. (2012). Science education for students with special needs. *Studies in Science Education*, 48(2), 187-215. doi:10.1080/14703297.2012.737117
- Von Secker, C. E. and Lissitz, R. W. (1999), Estimating the Impact of Instructional Practices on Student Achievement in Science. *Journal of Research in Science Teaching*, 36(10), 1110–1126. doi:10.1002/(SICI)1098-2736(199912)36:10<1110::AID-TEA4>3.0.CO;2-T
- Vygotsky, L. L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard university press.
- Wallace, C. S., & Kang, N. H. (2004). An investigation of experienced secondary science teachers' beliefs about inquiry: An examination of competing belief sets. *Journal of research in science teaching*, 41(9), 936-960. doi:10.1002/tea.20032
- Wallace, C. S., Hand, B. B., & Prain, V. (2004). *Writing and learning in the science classroom* (Vol. 23). Springer.
- Wheatley, G. H. (1991). Constructivist perspectives on science and mathematics learning. *Science Education*, 75(1), 9-21. doi:10.1002/sce.3730750103
- Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, 47(3), 276-301. doi:10.1002/tea.20329
- Windschitl, M., Thompson, J., & Braaten, M. (2008). Beyond the scientific method: Model-based inquiry as a new paradigm of preference for school science investigations. *Science education*, 92(5), 941-967. doi:10.1002/sce.20259
- Wise, K. C. (1996). Strategies for teaching science: What works?. *The Clearing House*, 69(6), 337-338.