

THESIS
3
2005

This is to certify that the
dissertation entitled


ELEMENTARY SOCIAL STUDIES TEACHERS'
IMPLEMENTATION OF CURRICULUM-EMBEDDED
PERFORMANCE ASSESSMENT IN SOUTH KOREA:
A MIXED METHOD STUDY

presented by

JINYOUNG CHOI

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in CURRICULUM, TEACHING
AND EDUCATIONAL POLICY


Major Professor's Signature

12/13/2005

Date

MSU is an Affirmative Action/Equal Opportunity Institution

LIBRARY
Michigan State
University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE
MAR 23 2008		
07 22 08		

**ELEMENTARY SOCIAL STUDIES TEACHERS' IMPLEMENTATION OF
CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT IN SOUTH KOREA:
A MIXED METHOD STUDY**

By

Jinyoung Choi

A DISSERTATION

**Submitted to
Michigan State University
In partial fulfillment of the requirements
for the degree of**

DOCTOR OF PHILOSOPHY

Department of Teacher Education

2005

ABSTRACT

ELEMENTARY SOCIAL STUDIES TEACHERS' IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT IN SOUTH KOREA: A MIXED METHOD STUDY

By

Jinyoung Choi

The study investigated elementary social studies teachers' implementation of curriculum-embedded performance assessment in South Korea. It examined teachers' implementation of performance assessment in terms of the formats, cognitive demands, and purposes of the assessment. It then examined what factors influenced their implementation.

This study used a mixed method design combining quantitative and qualitative methodologies. Data sources included questionnaires, interviews and documents. 700 teachers completed the questionnaire. Analyses of these data included regression analysis and structural equation modeling. Eight case study participants were selected from the respondents of the pilot study. The case studies involved analyses of interview data and assessment documents.

Two noticeable findings about policy and practice emerged. First, the case studies show that teachers believed they were responding to the reform, yet their practices each differed considerably from those of the others. The study identified four patterns in implementing the performance assessment reform — *reluctant*, *superficial*, *transitional*,

and *profound* implementation — on a spectrum from *symbolic* to *authentic* implementation. Second, teachers easily could implement the surface aspect of assessment (performance assessment formats), but struggled with implementing the deep aspects of assessment (higher-level cognitive demands and assessment *for* learning).

The following significant findings about the factors influencing teachers' implementation emerged from the quantitative analyses. First, all three groups of factors (school contexts, learning opportunities, and teacher capacity & will) were important for the success of the reform when we consider both the direct and indirect effects of these factors on teacher implementation. Second, *authentic teacher learning*, or active teacher learning aligned with performance-based student learning, was the most important factor that influenced teacher implementation. Third, teacher capacity and will factors had the second strongest direct influences on teacher implementation and mediated the influences of the other factors on teacher implementation. Fourth, most school contexts factors indirectly influenced teacher implementation through either the learning opportunities or teacher capacity and will factors.

The author suggests that teachers need *scaffolding* to change their implementation from reluctant compliance to profound change in practice, and discusses several ways of scaffolding to help teachers progress in their implementation of the reform, based on the findings.

Copyright by
Jinyoung Choi
2005

This dissertation is dedicated to my parents,
Daesoon Choi and Jeeyeon Lee,
for their unconditional love and support,
which has given me the confidence to face the world.

ACKNOWLEDGEMENTS

My journey to completing my doctoral program has been long and arduous. Many have eased my way. Without their considerable help, this dissertation would not have been possible. In this space, I express my sincere appreciation.

I am extremely grateful for the tremendous support that I have received from my dissertation director and committee chair, Dr. Mary Kennedy. She has profoundly shaped the research and writing of this study. She read numerous drafts, found time to meet with me whenever I requested it, and provided insightful feedback. She offered endless encouragement throughout my thinking and writing process, providing the right balance of challenge and support. She is the best teacher, mentor, and role model for me. I am forever grateful to her.

I am also most appreciative of my dissertation committee, Dr. Jere Brophy, Dr. Mark Reckase, and Dr. Cheryl Rosaen, for their encouragement, guidance, expertise, and time. Each offered insight and advice relating to his or her area of specialization during my doctoral journey. Jere Brophy enhanced my knowledge of social studies teaching and learning. Mark Reckase's statistical expertise served me well in both my understanding and completion of the structural equation modeling. Without Cheryl Rosaen's support and guidance, I could not have completed my preliminary examination and would not have advanced to the dissertation. I am also grateful to Dr. Gary Sykes for providing the support needed for producing this work. I developed my dissertation topic in his class, and my dissertation makes reference to many of his assigned readings.

Thank you also to Dr. Betsy Becker and Dr. Mary Kennedy for my research assistantship in the project, *Teacher Qualifications and the Quality of Teaching* (TQQT), which gave me the opportunity to understand American education better and made the pursuit of this degree financially possible. My experiences in this research project have been invaluable in providing me with a solid foundation for my professional career. For their intellectual support and friendship, I also express my appreciation to the TQQT staff: Soyeon Ann, Scott Metzger, Yanxuan Qu, and Meng-Jia Wu, who have worked with me for over four years.

Many colleagues and friends at Michigan State University have also contributed to this dissertation, both directly and indirectly. I especially thank JaeMin Cha, Sejung Choi, KaiLonnie Dunsmore, Shinho Jang, Chang Huh, Seung Hyun Kim, Mi Ran Kim, Suzanne Knezek, Ann Lawrence, Heejun Lim, Karen Lowenstein, Michael Morris, Hyun Soon Park, Ji-Won Son, Kyung Oh Song, and Yonghee Suh for their collegiality and compassionate friendship during my doctoral journey.

I cannot articulate sufficiently how much I thank Jeonghee Noh for being my strongest, most effective supporter on every step of this journey. She has offered encouragement and support through many years of studying in the United States, in both happy and difficult times. I am also thankful to my siblings, Sunkyung, Hyunjoo, and Seongjoon. I especially thank Hyunjoo, who cooked for me daily.

I express my deepest love and acknowledgements to my parents, Daesoon Choi and Jeeyeon Lee. Words could never adequately express how grateful I am for their unconditional love and support from Day One. This dissertation is dedicated to them.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER 1	
INTRODUCTION	1
Purpose of the Study	7
Significance of the Study	8
Overview of the Dissertation	10
CHAPTER 2	
LITERATURE REVIEW	12
What Does the South Korean Performance Reform Look Like?	12
Defining the South Korean Performance Assessment Reform	12
Key Features of the South Korean Performance Assessment Reform	14
Variation in Teachers' Responses to Reform	17
Formats of Assessment	20
Cognitive Demands of Assessment	20
Purposes of Assessment	22
Factors Influencing Teachers' Varied Responses to Reform	24
Teachers' Capacity and Will to Implement a Reform	26
Teacher Learning for Building Teachers' Capacity and Will	30
School Contexts in which Teachers Work as Contextual Factors	33
CHAPTER 3	
RESEARCH DESIGN AND METHODS	37
Research Design for the Present Study	37
Data Collection Instruments	39

Pilot Study	39
Revised Questionnaire	41
Interview Protocol	44
Data Collection Procedures	44
Questionnaire Data	45
Interview Data	45
Documents	45
Participants	46
Survey Participants	46
Case Study Participants	48
Data Analysis	50
Unit of Analysis	50
Quantitative Analyses	50
Analyses of Cases	53
Methodological Caveats	56

CHAPTER 4

TEACHERS' IMPLEMENTATION OF CURRICULUM-EMBEDDED

PERFORMANCE ASSESSMENT	58
Four Cases of Teachers' Responses to the Reform.....	58
Pattern 1: Profound Implementation.....	63
Pattern 2: Transitional Implementation	70
Pattern 3: Superficial Implementation.....	75
Pattern 4: Reluctant Implementation	81
What does it mean to Implement Performance Assessment?	85
Teachers' Reports on Their Implementation of the Reform	88
Formats of Assessment	88
Cognitive Demands of Assessment	91
Purposes of Assessment.....	92
Summary of the Second Section.....	94

CHAPTER 5

FACTORS INFLUENCING IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT	96
Preliminary Analyses.....	96
School Contexts and Teacher Implementation	99
Teacher Learning Opportunities and Teacher Implementation.....	102
Teacher Capacity & Will and Teacher Implementation.....	103
Correlations among Factors.....	104
Hierarchical Regression Analysis.....	106
Effects of School Contexts Factors	110
Effects of Teacher Learning Opportunities Factors.....	110
Effects of Teacher Capacity & Will Factors	112
Structural Equation Modeling (SEM).....	114
Measurement Model	118
Structural Model	125
Hypotheses Testing.....	127

CHAPTER 6

TEACHERS' LEARNING OPPORTUNITIES AND THEIR IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT	138
Teachers' Learning Opportunities for Implementing the Reform	138
Professional Development.....	140
Collaboration	148
Involvement in School Activities	151
Summary of the First Section.	153
Structural Equation Modeling (SEM) Analysis.....	154
Measurement Model	157
Structural Model	159
Hypotheses Testing.....	160
Summary of the Second Section.....	167

CHAPTER 7

DISCUSSION, IMPLICATIONS, AND CONCLUSION.....	168
Discussion.....	168
Teachers' Different Responses to the Performance Assessment Reform	169
Factors Influencing Teachers' Implementation of Performance Assessment	172
Implications	179
Concluding Remarks	189
APPENDICES	190
Appendix A. Questionnaire	191
Appendix B. Descriptions of Questionnaire Items.....	199
Appendix C. Interview Protocol.....	201
Appendix D. Categories and Examples of Interview Language	205
Appendix E. Correlation Tables	211
REFERENCES	214

LIST OF TABLES

Table 1	Framework for Investigating Teachers' Assessment Practices	23
Table 2	Summary of Data Collected.....	46
Table 3	Characteristics of Survey Participants	47
Table 4	Selecting Case Study Participants.....	49
Table 5	Characteristics of Interview Participants	49
Table 6	Summary of Data Sources and Data Analysis in Relation to Research Questions	54
Table 7	Descriptive Statistics for the Independent Variables	97
Table 8	Descriptive Statistics for the Outcome Variables.....	98
Table 9	Frequency for Class Size	100
Table 10	Correlations among Constructs/Factors.....	104
Table 11	Results of Reliability Analysis.....	107
Table 12	Results of the Hierarchical Regression Analysis	109
Table 13	Comparison of the Initial and Modified Second Order CFA Models	122
Table 14	Parameter Estimates of the Measurement Model	123
Table 15	Correlation Analysis among Latent Variables	124
Table 16	Results of the SEM Analysis	128
Table 17	Standardized Estimates for Direct, Indirect, and Total Effects.....	133
Table 18	Total PD Hours on Content Areas	141
Table 19	Frequencies of Learning Activities.....	146
Table 20	Frequencies of Collaboration with Colleagues.....	149
Table 21	Involvement in School Activities.....	151
Table 22	Results of the Proposed CFA Model.....	157
Table 23	Correlation Analysis of Latent Variables	158
Table 24	Results of Fit Indices in the Measurement Model and Structural Model	160
Table 25	Results of the SEM Analysis	161
Table 26	Standardized Estimates for Direct, Indirect, and Total Effects.....	165

Table 27	Descriptions of Questionnaire Items	199
Table 28	Categories for Sorting Data about Teachers' Assessment Practices	205
Table 29	Categories for Sorting Data about Teachers' Views	209
Table 30	Correlations between School Contexts and Teacher Implementation	211
Table 31	Correlations between Teacher Learning Opportunities and Teacher Implementation.....	212
Table 32	Correlations between Teacher Capacity & Will Variables and Teacher Implementation.....	213

LIST OF FIGURES

Figure 1.	Four Patterns in Implementing the Performance Assessment Reform	61
Figure 2.	Frequency of Use of Different Assessment Formats	88
Figure 3.	Formats of Assessment	90
Figure 4.	Frequency Ratings for Different Cognitive Demands of Assessment	91
Figure 5.	Cognitive Demands of Assessment	92
Figure 6.	Emphasis Placed on Different Assessment Purposes	93
Figure 7.	Purposes of Assessment	94
Figure 8.	Regression Model	108
Figure 9	Hypothesized Model	114
Figure 10.	Specified SEM Model	117
Figure 11.	Hypothesized Second Order CFA for Teacher Implementation	119
Figure 12.	Modified Second Order CFA model for Teacher Implementation	121
Figure 13.	Modified Hybrid Model	126
Figure 14.	Results of Hypothesis Testing	130
Figure 15.	The Relationship Between PD Hours on Content Areas and Teacher Implementation	142
Figure 16.	Box Plots of Implementation by PD Hours Spent on Specific Areas	144
Figure 17.	Teacher Learning Activities and Teacher Implementation	147
Figure 18.	Box Plots of Influence of Frequency of Collaboration on Implementation	148
Figure 19.	The Relationship between Collaboration and Teacher Implementation	150
Figure 20.	Involvement in School Activities and Implementation	152
Figure 21.	Box Plot of Implementation by Involvement in School Activities	153
Figure 22.	Hypothesized Teacher Learning Model	155
Figure 23.	Specified Hybrid Model	156
Figure 24.	Results of Hypothesis Testing	162
Figure 25.	Progress of Implementing the Reform	179
Figure 26.	A Model for Combining Mandates with Scaffolding Instruments	182

CHAPTER 1

INTRODUCTION

Over the past decade or so, there have been criticisms that the most widely used traditional multiple-choice tests fail to measure important aspects of learning such as higher order thinking skills and abilities to perform real world tasks as well as to provide enough (or detailed) information for teachers' instructional and diagnostic purposes (Darling-Hammond, Aness, & Falk, 1995; Resnick & Resnick, 1992; Wiggins, 1993). This growing concern over the heavy use of traditional multiple-choice tests has spurred interest in developing alternative assessments, such as authentic assessment, performance based assessment, and portfolio assessment. Proponents of these alternative assessments argue that new assessment methods will benefit both students' learning and teachers' practices. On the student side, these new assessments measure the kind of competency that matters in society, the workplace, and schools (Wiggins, 1993). On the teacher side, these new assessments provide useful information for teachers to improve student learning by evaluating what students understand and whether they apply knowledge to their real-world situations (Darling Hammond & Aness, 1996).

This American educational research inspired South Korean educators' need for alternative assessments and their increasing concerns with traditional multiple-choice tests. South Korean educators pointed out that even though the national curriculum¹ had emphasized teaching and assessing higher-order thinking, teacher instruction and

¹ In South Korea, the government controls curriculum, including the curriculum framework textbook and teacher manual publication. Elementary schools have only a single kind of textbook for each subject of each grade. Teachers' instruction and assessment practices are based mainly on the national curriculum.

assessment focused on lower-level cognitive demands (e.g., recall of factual knowledge) (Beck, 1998; Choi, 1998; Jung, 2001; The Korea Institute of Curriculum and Evaluation, 1999). In addition, the traditional high-stakes multiple-choice tests resulted in a race to raise test scores through teaching to the tests. Parents paid attention only to the test scores their children received, and in order to get higher test scores, students worked hard to memorize facts. Over-reliance upon traditional multiple-choice tests had detrimental effects on teaching and learning, deepening the gap between the intended curriculum and the enacted curriculum. Introspections on these problems led South Korean educators to consider alternative assessments that would align well with the goals of the national curriculum and would provide a different kind of information than traditional multiple choice tests provided.

In this context, the South Korean Ministry of Education began to focus its attention on creating a new assessment system to better evaluate the kinds of higher-order thinking emphasized in the national curriculum. In 1998, the Ministry of Education via its document, *Visions for Korean education 2002: Creation of new school culture*, mandated that all elementary, middle, and high schools use performance assessment. While the performance assessment reform was mandated nationally and urged all teachers in South Korea to use performance assessment, the mandate was for individualized, classroom-based performance assessment rather than standardized large-scale performance assessments. Although the standards and content for performance assessment had to be based on the national curriculum, specific performance assessment tasks and rubrics were to be developed by classroom teachers. Similarly, although the assessment reform has emphasized performance assessment, traditional paper-and-pencil tests can be used in

classrooms. The reform does not say that teachers should use only performance assessment. Instead, since teachers have relied heavily on traditional multiple-choice tests, the assessment reform suggests that teachers change their assessment practices from such extensive reliance on traditional tests to incorporate a greater use of performance assessment in their assessment portfolios.

The new assessment policy in South Korea aimed to align the national curriculum with local assessment. New learning theories, such as constructivism and socio-cultural theory had inspired a reconceptualization of curriculum and assessment (Shepard, 2000). Reflecting these new learning theories, the national social studies curriculum in South Korea emphasized higher-order thinking, inquiry, and real-world problem-solving (The South Korean Ministry of Education, 1997). However, the reformed national curriculum conflicted with the traditional multiple-choice tests most commonly used by South Korean elementary teachers. These exams have been used mainly to assess mainly recall and low-level comprehension of factual knowledge by asking students to select “right answers”. They have had difficulty providing a clear picture of what students can do with their knowledge, an assessment goal targeted by the national curriculum. This recognition of mismatch between curriculum and assessment, requested teachers to include alternative assessments to obtain a more comprehensive picture of students’ understanding as emphasized in the national curriculum. Performance assessments based on students’ investigations of open-ended and complex problems were suggested as a better way of assessing students’ higher order thinking such as reasoning and analysis.

The assessment reform also hoped to improve instructional practices. The emphasis on recall of factual knowledge in the traditional multiple-choice tests made

teachers inattentive to components of social studies learning such as higher-order thinking, decision making, and communication. Because questions on traditional multiple-choice tests were based mainly on the single national social studies textbook for the grade level, teachers had to cover all information presented in this textbook and students had to read and memorize this information. Although the reformed national curriculum recommends that teachers use the textbook as merely one of many instructional resources, teachers have felt pressurized to cover the entire content of the textbook in order to improve students' scores on traditional multiple-choice tests. As a result, students regard social studies as uninteresting. Darling-Hammond et al (1995) note that if the assessment tasks require students' active learning, students' attitude toward learning can be changed. Thus, because the type of assessments plays a powerful role in teachers' decision-making about what and how to teach, the reform proponents believe that changes on assessment practices can engender changes in instructional practices as well as student learning.

Those who advocate them, assert that performance assessments foster quality instruction (Darling-Hammond et al., 1995; Khattri, Kane, & Reeve, 1995; Khattri & Sweet, 1996; Liu, 2000; Shepard et al., 1995). Since the mandated performance assessment reform in South Korea requires teachers to assess higher order thinking displayed in the performance assessment tasks (e.g., essays, reports, and presentation), the Ministry of Education expects that teachers' instructional practices will also change, to focus more on students' thinking (Jung, 2001). This seems a likely scenario when considering that teachers' traditional assessments influenced teachers' instructional practices in the past. For example, teachers' traditional instructional practices focused on teaching factual knowledge represented in the social studies textbook because students

could obtain higher scores in the traditional multiple-choice tests if they memorized textbook content.

However, changes expected in the mandated assessment reform are unlikely to occur easily and quickly. Evidence shows that changing educational practices are not easy and become more difficult when the change involves transformation of the existing structure of schooling (Cuban, 1993; Tyack & Cuban, 1995). The performance assessment reform called for a “deviation from traditional instruction and assessment practices and thus challenged the established organizational structure of schooling” (Khattri & Sweet, 1996). Cuban (1993) divides the reforms of the past century into “incremental” and “fundamental” changes. Incremental reforms aim to change teachers’ practices within existing structure of schooling by, for example, decreasing class size and adding new courses, under the assumption that the basic structure needs only a few repairs. Fundamental reforms assume that the basic structure is flawed and needs new structure. Thus, fundamental reforms are more difficult to implement.

In Cuban’s terms, South Korea’s new forms of assessment are fundamental reforms because they require a new paradigm. Shepard (2000) notes that curriculum, learning theory, and assessment in the new paradigm are the “direct antithesis of principles in the old paradigm” (p.18). New assessments need to be compatible with new views of curriculum, teaching, and learning, and assessment reforms that seek fundamental changes are difficult to implement and sustain. Teachers who long have lived in the old paradigm find it difficult to implement new assessments. Imagine a teacher who learned in traditional ways during his or her own schooling. What if this teacher learned to teach in traditional ways during pre-service and in-service teacher education programs?

What if this teacher has taught and assessed in traditional ways for many years? What if this teacher feels no dissatisfaction with his or her traditional instructional practices?

Despite the problems and challenges faced by teachers who have lived in the old paradigm, the current policy simply pushes them to implement the assessment reform and many have struggled with implementing it. Although the South Korean government has mandated that teachers implement the performance assessment reform, teachers' practices may not fulfill the government's hopes. This problem is analogous to the problem teachers face. Although teachers try to teach the content they want their students to learn, based on the curriculum framework, it may not be easy for all students to learn what the teachers expect them to learn, especially if learning requires complex thinking. As teachers should think about the problems and difficulties their students have when they teach, policy changing or improving teachers' practices should start by understanding problems and challenges teachers faces when they implement the reform.

Fullan (1992) notes that in order to achieve substantial education reform, educators must understand problems faced at the level at which change actually occurs. Thus, this present study starts with examining how teachers respond to the assessment reform to understand what it means to implement education reform at the local level. Then, it identifies what factors influence teachers' implementation of this reform in order to develop strategies for helping teachers to bring about more fundamental changes in their assessment practices.

The Purpose of this Study

The purpose of this study is to investigate how South Korean teachers respond to the mandated performance assessment and to identify the factors that influence teachers' implementation of this reform. Specifically, this study examines the features of teachers' learning opportunities, and their influences on implementation of the assessment reform. For the purpose of this study, I examine South Korean teachers' assessment practices in upper elementary social studies classrooms using both case studies and questionnaire data. The specific research questions are as follows:

1. How do teachers implement the assessment reform?
 - (1) What kinds of *formats of assessment* do teachers use?
 - (2) What kinds of *cognitive demands* do teachers assess?
 - (3) What are teachers' *purposes of assessment*?

2. What are the factors influencing teachers' implementation of the assessment reform?
 - *School contexts in which teachers work:*
 - (1) Does the school community's attitude toward the performance assessment reform influence implementation?
 - (2) Does the school climate influence implementation?
 - (3) Does the availability of school resources influence implementation?

- *Teachers' learning opportunities:*

- (4) Does the number of hours spent by teachers to learn contents specific for the reform in professional development (PD) programs influence implementation?
- (5) Do the authentic teacher learning activities facilitated in PD sessions influence implementation?
- (6) Does teachers' collaboration influence implementation?
- (7) Does teachers' involvement in school activities influence implementation?

- *Teachers' capacity and will to implement the reform*

- (8) Do teachers' beliefs about instruction and assessment influence implementation?
- (9) Does teachers' knowledge about performance assessment influence implementation?
- (10) Does teachers' willingness influence implementation?

Significance of the Study

This study attempts to provide policymakers, school administrators, and teacher educators with information about what is needed to help teachers to successfully implement an assessment reform by examining both what actually occurs when teachers implement the assessment reform and what factors influence their implementation. Investigating teachers' enacted practices, as translated from the policy's intended practices, is important for understanding the progress of the reform. Based on a diagnosis

of the current progress of the reform, policymakers and teacher educators can design future plans to help teachers to successfully implement the assessment reform.

This study attempts to contribute to policymakers', school administrators', and teacher educators' understanding of the importance quality of learning opportunities in helping teachers to bring about fundamental changes in their assessment practices. In recent years, many educational researchers have claimed that teacher learning is key to connecting policy and practice (Cohen & Hill, 2000; Little, 1993; Supovitz & Turner, 2000). Since the assessment reform requires teachers to do something new, unfamiliar, and possibly uncomfortable, teachers need to learn how they can implement the reform with comfort. This study cautions South Korean policymakers, school administrators, and teacher educators that teachers may not implement the reform successfully without sufficient and quality learning opportunities.

This study attempts to expand upon previous studies that examined the effects of professional development. Garet, Porter, Desimone, Birman, & Yoon (2001) note that although there is a large body of literature that describes quality professional development, relatively little research has been conducted to examine the effects of professional development programs on teachers' practices. By examining the effect of different aspects of professional development on teachers' assessment practices, this study attempts to provide such evidence.

Overview of the Dissertation

In this chapter, I have established the central questions of this study: How do teachers respond to the mandated assessment reform and what factors influence their implementation of the performance reform? Investigating this question will contribute to our understanding of how we can help teachers to bring about fundamental changes in their assessment practices.

In the next chapter, I survey research on the policy implementation that explored teachers' different responses to the reform and identified factors influencing teachers' practices. I then review both South Korean and American literature on performance assessments to develop a framework for investigating teachers' assessment practices.

In the third chapter, I detail the methodology employed in this study. First, I provide a brief description of an earlier pilot study. Then, I describe why I chose a mixed research design combining survey and case study methodologies to investigate my research questions, and explain how I collected my data, including details on participants, data sources, instruments, and data collection procedures. I also provide the methods of data analysis for each methodology.

The fourth chapter presents analyses of both the case study and questionnaire data to describe how South Korean teachers implemented the mandated assessment reform. The chapter opens with four case studies that unpack teachers' practices to illustrate teachers' different responses to the reform. Then the chapter uses the questionnaire data on teachers' implementation of the reform to verify findings from the case studies.

Chapters five and six report findings on the factors influencing teachers' implementation of the performance assessment reform. The fifth chapter presents findings from regression analysis and structural equation modeling, employed to examine the influences of three groups of factors (school contexts in which teachers work, teachers' opportunities to learn about performance assessment, and teachers' capacity and will to implement the reform) on teachers' implementation of performance assessment. While the fifth chapter presents a comprehensive model that includes all three groups of factors to show their influence on teachers' implementation of the reform, the sixth chapter focuses on the relationship between one of three groups of factors, teachers' learning opportunities and teachers' implementation of the reform.

The first section of the chapter six describes the relationships between the surveyed teachers' implementation of performance assessment and their experiences with different types of learning opportunities by including more learning opportunities variables that could not be tested in chapter five. The second section of chapter six reports findings from structural equation modeling employed to examine both the direct and indirect influences of learning opportunities on teachers' implementation of performance assessment.

In the final chapter, I discuss the findings from the case study and questionnaire data. Based on what I have learned from my study, I suggest how policymakers and teacher educators can help teachers to change their assessment practices. I conclude my study with my final thoughts about effective approaches to educational reform.

CHAPTER 2

LITERATURE REVIEW

This chapter consists of three sections. First, I review South Korean educational research to describe the South Korean performance assessment reform. Second, I review both South Korean and American educational research that examines factors influencing teachers' reform practices. Little research focuses specifically on teachers' classroom-based assessment practices, so this literature review includes studies that investigate teachers' instructional practices because instruction and assessment are related closely. Third, I develop a conceptual framework for investigating teachers' responses to the South Korean performance assessment reform.

What Does the South Korean Performance Reform Look Like?

I briefly describe the South Korean assessment reform to help readers to understand the South Korean assessment policy in terms of definitions, and key aspects of the assessment reform. I also review South Korean literature on performance assessment.

Defining the South Korean Performance Assessment Reform

To evaluate the South Korean performance assessment, it is important to understand what qualities pertain to such assessment. I first examine three commonly used terms for the new assessment in American educational research, and illustrate the features of this reform by comparing South Korean performance assessment to the three different assessments that inspired it. A variety of names describe assessment reforms that move

teachers away from traditional tests to more innovative and meaningful forms of assessment, such as alternative assessment, authentic assessment, and performance assessment. These different terms for new assessments are:

- **Alternative assessment:** Assessment intended to distinguish the form of assessment from traditional, fact-based, multiple-choice testing.
- **Authentic assessment:** Assessment intended to highlight the “real world” nature of tasks and assessment contexts that make up the assessment.
- **Performance assessment:** Assessment that requires students to actually perform, demonstrate, construct, and develop a product or a solution under defined conditions and standards (Department of Education, 1997).

The South Korean performance assessment system appears to entail all the qualities of these three new assessments. The national curriculum framework developed by the South Korean Ministry of Education (1997) defines performance assessment as assessment that requires teachers to observe students’ performance tasks directly and judge the learning outcomes professionally. From this definition, tasks should be connected with curricular aims and the tasks should be important and meaningful for students by making connections with students’ lives. Performance is defined as constructing answers or products rather than selecting a right answer (The South Korean Ministry of Education, 1997).

Thus, the South Korean performance assessment can be defined as including all three new forms of assessment and a fourth in that (1) it offers an *alternative* approach to traditional multiple-choices tests, (2) it uses *authentic*, real-world tasks, (3) it assesses

student *performances*, and (4) it connects to *curricular aims*. Although a literal translation of the South Korean reform policy is “performance assessment”, the best English translation of the policy is ‘*curriculum-embedded performance assessment*’² because the South Korean performance assessment reform emphasizes that the performance assessment occurs in a particular classroom and is embedded in the curriculum, not large scale on-demand performance assessments.

Key Features of the South Korean Performance Assessment Reform

Six interrelated features of the South Korean performance assessment reform emerge from the literature review. The key aspects of performance assessment are compared with aspects of traditional tests. Although the reform moves teachers away from over-reliance on traditional multiple-choice tests, it does not forbid the use of these tests. The reform mandate emphasizes performance assessment that elicits students’ higher order thinking and problem solving abilities emphasized in the national curriculum reform, but allows teachers to decide which assessment methods are best for their students.

Performing tasks (vs. Selecting a answer). The assessment reform suggests that teachers employ performance assessment tasks requiring students to demonstrate their understanding by constructing responses or by performing tasks, rather than by selecting “right” answers. This feature of the South Korean assessment reform is well connected to the reform’s emphasis on learners as active knowledge constructors, rather than on their passive reception of targeted content (The South Korean Ministry of Education, 1997).

² Since the name of the reform is long, I will use the terms, “the performance assessment reform”, “the assessment reform” or “the reform” interchangeably to refer to “the curriculum-embedded performance assessment reform”.

Performance assessment suggests that while it is important to assess what students know, it also is important to assess what students can do with their knowledge. By examining both processes and products of students' learning, teachers can understand what students know as well as how they think and solve the problems (Jung, 2001). The gained information is useful for teachers to provide feedback to students for improving students' future learning.

Curriculum-embedded (vs. On-demand). The assessment reform emphasizes that assessment should be embedded in the curriculum. That is, assessment occurs in a classroom setting and it is not separated from the learning process. This concept can be understood better when compared with traditional assessment practices. In South Korea, traditional assessment practices were closer to on-demand assessments that take place as scheduled events, outside of normal classroom settings. Mandating the assessment reform, the government removed the employment of large-scale multiple-choice tests that were on-demand assessments. Instead, the reform recommends the increased use of performance assessments embedded in teachers' daily curriculum.

Emphasis on higher order thinking (vs. Recall of knowledge). The assessment reform requires teachers to design assessments, based on a new understanding of what kinds of learning are important for students. The performance assessment reform emphasizes assessing students' higher order thinking as emphasized in the national curriculum. While traditional tests effectively can determine whether or not students acquire a body of knowledge, performance assessments are useful for assessing complex cognitive skills such as analysis, synthesis, and application (Cho, 1999; Jung, 2001; Keon, 2000). Traditional multiple-choice tests asked students to match what the national

curriculum expected students to learn, whereas assessment reform expects that performance assessments eliciting students' higher thinking allows teachers to change the focus of their instructional practices from delivering knowledge to students to helping students to engage actively in their learning process.

Engaging authentic tasks (vs. Decontextualized). The assessment reform emphasizes that assessment tasks be the actual performance relevant to students' real life contexts. While traditional tests have been criticized for the use of decontextualized questions, the performance assessment requires students to do tasks which resemble the context in which adults do their work (Jung, 2001). Students, thus, would not be able to complete these tasks by memorizing facts. Students need to apply what they learn, to solve real-world problems. Allowing students to perform, demonstrate, and construct something provides them with opportunities to apply what they know to problems they might face in the future (Wiggins, 1998).

Assessment for learning (vs. Assessment of learning). There are different purposes of using assessments (Darling Hammond & Aneess, 1996). One is that assessment is an efficient means to sort students, and another is that assessment is a means for identifying and supporting students' strengths and weaknesses. Whereas traditional tests in South Korea are important for the former goal, the new assessment policy emphasizes the latter. Because South Korean traditional tests are used to assign grades at the end of the semester and to distinguish students who are successful from those who are not, the tests are not diagnostic (Cho, 1999; Keon, 2000). Chappuis, Stiggins, Arter, & Chappuis (2004) labeled this purpose of assessment as "assessment *of* learning". The purpose of assessments emphasized in the reform is not to rank order and sort students

according to test scores, but to provide feedback to students and teachers to improve teaching and learning (The Korea Institute of Curriculum and Evaluation, 1999). By using performance tasks, teachers can know how students think and diagnose where students have learning difficulties. The information obtained from performance assessment helps teachers to develop future lessons for improving students' learning. Chappuis et al. (2004) labeled this purpose of assessment as "assessment *for* learning".

Teachers as expert assessors. Students' abilities are assessed by their own teachers. In traditional tests, assessors, in most cases, are machines that, or outsider who, do not know individual students and they rely only on test scores to assess students' abilities. However, the performance assessment in South Korea emphasizes teachers' roles as expert assessors (The Korea Institute of Curriculum and Evaluation, 1999). Because they see their students every day, teachers use various resources to evaluate students' strengths and weaknesses. This idea of teachers as expert assessors indicates how much teachers' capacities need to be built to assess students' abilities.

Variation in Teachers' Responses to Reform

Although teachers had the same reform policy, it was anticipated that there would be variations in teachers' implementation of the assessment reform. Evidence shows that teachers respond to reforms in different ways (Fuhrman, Clune, & Elmore, 1988; Grant, 1998; Spillane & Jennings, 1997; Spillane & Zeuli, 1999). While some teachers could implement only surface elements of the reform, such as materials used and grouping assignments, some teachers could implement both surface and deep elements of the reform.

Spillane & Zeuli (1999) found variations in teachers' responses to a mathematics reform. From observations of 25 classrooms, they identified three patterns of teacher practice in response to the reform. Pattern 1 reflected teachers' practices using conceptually grounded tasks and conceptually centered discourse to help students to understand principled mathematics knowledge by doing mathematics. When teachers taught in ways that fit this pattern, they fundamentally changed their instructional practices. Pattern 2 reflected teachers' practices using conceptually oriented tasks and procedure-bounded discourses. This pattern suggests that teachers' responses to the reform can be a mixture of old practices and new practices. Pattern 3 reflected teachers' practices that were totally unaligned with the intent of the reform. The results of the study indicated that teachers could enact some dimensions of teachers' practices successfully, but had difficulties enacting other dimensions.

Saxe, Franke, Gearhart, Howard, & Crockett (1997) also documented several patterns of changing assessment practices using the cases. One pattern was teachers implemented a new form of assessment in a way that served 'old' functions (what to assess). In this study, a teacher used open-ended problems, but not to evaluate students' mathematics understanding and skills. While the teacher used a new format of assessment, what the teacher assessed (function), employing the new format of assessment, was about whether students' answers were "right". Another pattern was that teachers re-purposed old assessment forms to evaluate new functions. In this study, a teacher used an old assessment format to encompass a reform function. The teacher used exercises to evaluate the percentage of "right" answers as well as their written explanations.

The findings from these two studies are unsurprising, since implementation is a developmental process and changes take time. While some teachers may implement reform at the surface level, bringing out changes that they can do easily, some teachers may implement the reform at the deeper level. Coburn (2003) defines deep change as “change that goes beyond surface structures or procedures (such as changes in materials, classroom organization, or the addition of specific activities) to alter teachers’ beliefs, norms of social interaction, and pedagogical principles as enacted in the curriculum”(p.4). When teachers implement a reform, they are likely to focus on implementing surface dimensions of practices, rather than substantial dimensions (Spillane & Jennings, 1997). Based on the literature review, I hypothesize that there are variations in teachers’ responses to assessment reform, in terms of different aspects of the reform (surface vs. deep).

To investigate how teachers respond to South Korea’s assessment reform, I developed three aspects of assessment practices that reflect this review of both South Korean and American Literature. The first two aspects are based on a framework from Saxe, Franke, Gearhart, Howard, & Crockett (1997), which focuses on how teachers implement the *format* and the *cognitive demands* of assessment. I then add a third aspect, *purpose* of assessment, to their framework. While formats of assessment would be easier to implement, representing the surface element of the reform, cognitive demands and purposes of assessment would be more difficult to implement, representing the deeper or substantial elements of the reform. In the following section, I describe these three aspects of assessment.

Formats of Assessment

Formats of assessment refers to the particular types of tools teachers use to assess their students. The forms of assessment is typed into three: (1) teachers' own paper and pencil objective tests including multiple-choice, true/false, matching, and short answer fill-in tests, (2) published tests including standardized objective achievement tests and objective tests provided in published text materials, (3) performance assessment defined as the observation and rating of students' product that demonstrate their proficiency (Stiggins & Conklin, 1992). While this study focus on teachers' use of performance assessment, teachers' use of other assessment formats is also examined for the comparison purposes.

Performance assessment formats are structured variously. Among these formats are: (1) open-ended exercises requiring students to construct response to prompts or problems within a short time, (2) extended tasks, such as projects, that require more time, (3) demonstrations that take the form of student presentations of their work, (4) portfolios that collect student work and show development, and (5) teachers' observations that gauge student classroom performance (Khatti & Sweet, 1996). The South Korean assessment reform suggests that teachers use a wide range of performance assessment formats.

Cognitive Demands of Assessment

Cognitive demands of assessment refers to what kind of intellectual work is assessed. For example, a teacher uses an assessment format to evaluate students' factual knowledge, but a teacher wants to assess students' reasoning by using another format. Wiggins & McTighe (1998) indicate that different types of assessments can evaluate

different things. Since traditional tests such as traditional paper-and-pencil tests and quizzes often have been used to assess lower-level cognitive demands, such as factual information, concepts, and discrete skills, the performance assessment reform suggests that teachers also should use performance assessments to assess higher-level cognitive demands, such as what students can do with their knowledge.

Researchers at the Center on Organization and Restructuring of Schools (CORS) established standards for authentic assessment tasks in social studies (i.e., higher-level cognitive demands (Newmann. & Associates, 1996; Scheurman & Newmann, 1998). Because the South Korean assessment reform also emphasizes similar standards for student learning, as I described in the key characteristics of the assessment reform, I will use the CORS standards as criteria for evaluating whether teachers are assessing higher-level cognitive demands emphasized by the mandated assessment reform. The criteria developed from these standards for authentic student tasks and performance are outlined below.

First, performance assessment tasks require students to construct their knowledge. Rather than merely acquiring knowledge produced by others, students construct their knowledge through thinking processes such as interpreting, synthesizing, and evaluating complex information. For example, in a history lesson, students construct their historical interpretation through collecting, analyzing, and reasoning historical evidence. In addition, assessment tasks provide opportunities for students to consider alternative perspectives. Second, performance assessment tasks ask students to demonstrate understanding of big ideas, rather than knowing bits of facts and information. Teachers provide opportunities to think about relations among concepts, by asking “why” and “how” questions, not only by

“what” questions. Performance assessment tasks are designed to ask students to explain their thinking via various forms of oral, written, and symbolic language. Third, performance assessment tasks ask students to apply to real-world problems what they learn. Rather than merely checking what students know, teachers assess whether they can use their knowledge to solve problems they likely will encounter outside school.

Purposes of Assessment

Purposes of assessment refers to the use of assessment results. Assessment and its results can be used for two purposes. The first is to use assessment and its results for assigning grades and providing report cards to inform parents about student learning (*assessment of learning*). The second is to use assessment and its results for helping teachers teach better and students learn more (*assessment for learning*) (Stiggins, 2005). Since in the traditional assessment often used by South Korean teachers, assessment results are used mainly for assigning grades to students at the end of semester for the purpose of *assessment of learning*, the performance assessment reform emphasizes the importance of another assessment purpose (*assessment for learning*). The assessment reform suggests that teachers use the results of performance assessment to give feedback to students as well as to plan for future instruction that will improve student learning. Teachers use information obtained from performance assessment to determine what their students know as well as can do, through directly observing students’ demonstrations or products.

Brief description of the framework, as it applies to changes mandated by the South Korean government, is shown in Table 1.

Table 1

 Framework for Investigating Teachers' Assessment Practices

	Performance assessment	Traditional Assessment
Formats	<i>Performance assessment formats:</i> Short, open-ended tasks, extended open-ended tasks such as project and reports, portfolio, and demonstration, and observation	<i>Traditional formats:</i> multiple choice tests or selected-response questions
Cognitive Demands	<i>Higher-level of cognitive demands:</i> Explaining how students solve a problem, applying what students know to real-world problems, making arguments with evidence, considering alternative interpretations.	<i>Lower-level of cognitive demands:</i> Recall of factual knowledge
Purposes	<i>Assessment of learning:</i> Using assessment to guide instruction and give feedback to students	<i>Assessment for learning:</i> Using assessment for grading and student report card

Based on these three aspects of assessment, this study investigates variations in teachers' implementation of the performance assessment reform. In the next section, I will identify factors which may explain teachers' varied responses to the reform, based on a review of both American and South Korean educational literature. (The South Korean performance assessment reform is in part a response to American educational communities' increased interest in new assessments, and the South Korean policy makers modeled their reforms after American performance assessment). While the performance

assessment reform was based on American literature, the use of performance assessment was mandated by the government. It is similar to the state large-scale performance assessment or portfolio assessment that has been used in Maryland, Vermont, and Kentucky. However, even if the performance assessment reform was mandated nationally, it was local-based assessment rather than large-scale performance assessments. Thus, I will review both U.S. and South Korean literature that examined factors influencing individual teachers' instructional and assessment practices.

Factors Influencing Teachers' Varied Responses to Reform

Implementation of reforms depends on how implementers *interpret* policy and *respond* to it (McLaughlin, 1991; Spillane, 1998; Spillane & Jennings, 1997; Spillane & Zeuli, 1999). Implementers ignore, adapt, or adopt to a new policy depending on their beliefs, knowledge, and the contexts in which they work. Thus, the success of policy implementation depends on an understanding of problems that implementers face (Fullan, 1991).

Literatures show that there are interlocking reasons why reform policies have not succeeded in the U.S. Some educators argue that there are unchanging cultural beliefs about knowledge, teaching, and learning (Ball, 1990; Cuban, 1993; Tyack & Cuban, 1995). Some educators emphasize that teachers do not have sufficiently deep knowledge to change their practices as the reform policy intended (Cohen, 1990). Some argue that teachers have not received sufficient and appropriate learning opportunities to learn a new policy (Cohen, 1990; Cohen & Barnes, 1993). Simply giving teachers mandates, without

changing their beliefs about teaching and learning, understanding of the policy, and providing support, is unlikely to change their practices

McLaughlin (1991) argues that implementation studies need to be linked both to macro level analysis at the system level, and to micro level analysis at the individual level. Similarly, reviewing two studies examining local responses to the same policy from different perspectives (from the inside-out vs. from the outside-in), Knapp (2002) suggests studying connections between individual actors and their workplace environments. He noted that this kind of research is difficult to do, and that many studies focus on one or the other. Thus, a better approach would be to look at teachers' implementation of a reform, considering both individual factors and contextual factors because individual actors reside in the contexts in which they work.

While both individual and contextual factors are important for implementing a reform, influences of contextual factors are mediated by influences of individual factors. Examining college science faculty's instructional practices, Gess-Newsome, Southerland, Johnston and Woodbury (2003) found that removing structural barriers was necessary, but was an insufficient condition for changing instructional practices. Thus, to achieve changes in teachers' instructional practices, reformers should consider the importance of individual factors such as teachers' capacity and willingness to implement reforms, along with creating better work contexts for teachers.

Based on both U.S. and S. Korean literature, in this section, factors influencing teachers' practices are organized into three groups. The first includes individual factors comprising teachers' capacity and will to implement a reform. The second includes contextual factors that are related to the contexts in which teachers work. The third

includes factors related to teachers' learning opportunities. Learning opportunities are the factors that can act as both individual and contextual factors in South Korea. While teachers can choose their learning opportunities, teachers are required to earn a number of credits for their professional development (PD) programs. They usually take their professional development programs during summer and winter vacation with stipends. If teachers take the minimum number of PD hours they are ordered to take, PD can act as an external factor. However, if teachers decide to have more PD programs than the required minimum credits, PD can act as an individual factor. Also this factor can act as an influence that builds teachers' capacity and will to implement reform. While in the U.S. literature are found many factors affecting teachers' practice, I include only the relevant factors that can apply to the South Korean situation.

Teachers' Capacity and Will to Implement a Reform

The success of teachers' implementation of reform policy depends on teachers' *capacity* and *will* to change. McLaughlin (1991) argues that policy implementation depends on local capacity and the will of individual implementers. *Capacity* is defined as the ability to perform or produce (Pickett, 2001). This means that teachers need to learn knowledge and skills in order to act as a policy-makers intend. *Will* includes "attitudes, motivations, and beliefs that underlie an implementer's response to a policy's goals or strategies (McLaughlin, 1987; Spillane, 1999).

Teachers' beliefs about the nature of knowledge, teaching, and learning, influence their implementation of a reform (Ball, 1990; Cimbricz, 2002; Cohen & Hill, 2000; Cohen & Hill, 2001; Cronin Jones, 1991; Cuban, 1993; Prawat, 1992; Spillane, 2000; Tyack &

Cuban, 1995). New reform policies are interpreted through the lenses of teachers' beliefs. Ball (1990) found that a mathematics teacher had difficulties implementing reform policy because of her traditional beliefs about teaching and learning mathematics. The math teacher believed that there was a "right answer" to a mathematics problem; teachers deliver essential knowledge, and students receive it. These traditional beliefs shaped her instructional practices in ways inconsistent with the policy. Although the math teacher wanted students to engage in problem solving, as the policy initiative emphasizes, because she believed that there is a correct understanding, students' participation was structured and limited by the "right" solution path.

In contrast, when teachers' beliefs are consistent with those of the new policy, teachers implement it well. Observing and interviewing a fifth grade teacher, Spillane (2000) found that there were differences in instructional practices depending on the teacher's different views of different subjects (Language arts vs. Mathematics). The teacher was more successful in changing her practices in language arts than in mathematics because of her different views about the two subjects. Whereas in language arts the teacher believed that knowing was not simply memorizing facts and rules but applying them, in mathematics she viewed memorizing rules and procedures as important in learning mathematics. These different views caused the teacher to handle the two subjects differently.

Flexer and Gerstner (1993) also found that teachers' beliefs about teaching, learning, and assessment was an influential factor in implementing performance assessment. According to this study, because teachers believed it was important to teach and assess higher-order thinking such as problem-solving and critical thinking, and

performance assessment was useful to evaluate such student thinking, they could implement performance assessment successfully.

Teachers' willingness to implement a reform is another important individual factor in changing their practices (Grant, 1998; Mitman & Lambert, 1993; Spillane, 1999, 2004). Willingness refers to teachers' motivation to implement a proposed reform (Spillane, 1999). While teachers' beliefs mean what they think about reform ideas, teachers' willingness refers to their intentions for implementing reform ideas. Mitman & Lambert (1993) examined the reform implementation process in 17 schools and found that success of reform efforts relied heavily on teachers' willingness to change their instructional practices.

As well as teachers' beliefs about teaching and learning, and teachers' willingness to implement a reform, teachers' knowledge relating to the reform is a critical individual factor that affects how teachers implement a reform (Cohen, 1990; Gess-Newsome et al., 2003; Putnam, Heaton, Prawat, & Remillard, 1992; Wilson, 1990; Woodbury & Gess-Newsome, 2002). Cohen's (1990) case study supports this hypothesis. His case study shows that students will not learn new mathematics unless teachers know it and teach it. For example, Mrs. Oublier, a mathematics teacher, taught new topics, such as estimation, included in the new mathematics framework, that were not covered in traditional mathematics. Even though Mrs. Oublier thought that estimation was important for students to make sense of numbers, her instructional practice did not support students' learning of estimation in a way that agreed with the new mathematics framework. An interview with Mrs. Oublier shows that she lacked mathematical knowledge. Although she understood the purpose of teaching and learning of estimation, she did not have a deep

knowledge of either mathematics or teaching mathematics and thus struggled to teach the new mathematics framework.

Wilson (1990) in her case study of a mathematic teacher who enacted the state mathematics reform also showed that the teacher's lack of knowledge about reform-oriented instructional strategies influences how to teach mathematics. For instance, because the teacher did not know how and why to use manipulatives emphasized by the reform curriculum framework, he skipped some lessons presented in the reform textbook that needed to use manipulatives. Even if the teacher had the new textbook and materials he could use to implement the reform, he could not implement the reform as proposed because of his lack of knowledge about alternative ways of teaching. This case study indicated that what the teacher did was the best he could do because he had no in-service training to learn the reform and, consequently, he had not the capacity to implement the reform.

Several studies on performance assessment indicate that teachers' knowledge of the subject matter they teach and understanding of performance assessment affect their implementation of the reform (Firestone, Mayrowetz, & Fairman, 1998; Flexer & Gerstner, 1993; Francis, 1994). Flexer and Gerstner (1993) found that teacher knowledge of how to use performance assessment influenced their implementation and instructional practices. Teachers' lack of knowledge about it caused them to assess only computation and computational problem solving by giving paper and pencil exercises to students.

Teacher Learning for Building Teachers' Capacity and Will

Among factors shown to influence teachers' implementation of a reform, teachers' opportunities to learn the new policies are a crucial influence on their instructional practices. This hypothesis is related closely to the first hypothesis, because teacher learning helps teachers to build capacity, motivates them to change their practices, and changes their beliefs about knowledge, teaching, and learning regarding policy implementation (Cohen & Hill, 2000; Supovitz & Turner, 2000).

Several U.S. scholars argue that the more professional development opportunities teachers have, the more familiar they are with new policies. Supovitz, Mayer, and Kahle (2000) examine whether professional development efforts in the context of systemic reform can promote teachers' use of inquiry-based instruction. Their findings indicate that professional development programs can improve teachers' attitudes toward inquiry, their preparation to use inquiry-based pedagogy, and their actual use of inquiry-based teaching practices.

While the quantity of learning opportunities influences policy implementation, the quality of learning opportunities also is important in changing teachers' practices. That is, simply offering more learning opportunities does not necessarily help teachers to implement a new policy well. Thus, hypotheses about teacher learning should include consideration of the quantity and quality of teachers' learning opportunities. The characteristics of quality learning opportunities are identified from review of the literature.

First, quality teacher-learning opportunities are more content-specific (*content of learning*). When teachers' learning opportunities are more content-specific, teachers implement reforms better (Cohen & Hill, 2000; Cohen & Hill, 2001; Desimone, Porter,

Birman, Garet, & Yoon, 2002; Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet et al., 2001; Kennedy, 1998). Cohen and Hill (2000), for example, examined the influence of professional development on teachers' instructional practices. The findings indicated that teachers' learning opportunities significantly influenced their policy implementation when "those opportunities are situated in curriculum that is designed to be consistent with the reforms" (Cohen & Hill, 2000, p.329). Garet et al. (2001) also found that professional development programs that focused on academic subject matter (mathematics, in this study) were more likely to enhance teachers' knowledge and skills as well as to change their instructional practices.

Second, quality teacher-learning opportunities provide active learning activities to teachers (Garet et al., 2001; Ingvarson, Meiers, & Beavis, 2005; Knap, 2003). The reason why active learning opportunities (*pedagogy of learning*) are important is analogous to the reason why active learning is important in student learning. When teachers experience the same learning opportunities emphasized for student learning in the reform, teachers may be able to provide the same learning opportunities to their students. Grant, Peterson, and Shojgree-Downer's (1996) study also found that the pedagogy of learning for teachers is important as well as the content of learning for their practices. This study showed that when teachers actively participated in their learning through discussing the goals and content of a reform, researching topics related to the reform, and observing other colleagues' instruction, they extended their knowledge as well as changed their practices.

Third, teachers' collaboration with fellow teachers can provide quality learning opportunities to teachers (Borko, Elliott, & Uchiyama, 1999; Coburn, 2001; Davis, 2002; Gallucci, 2003; Ingvarson et al., 2005; Porter, 1999; Smith, 2000; Spillane, 1999).

Darling-Hammond, Aness, & Falk (1995) emphasized the importance of teachers' dialogue and professional collaboration to support teachers' authentic assessment. In the study, teachers explored what the central themes are for their teaching, discussed difficulties and dilemmas they faced, and worked together to find ways for solving problems through teachers' diverse experiences. This collaboration was beneficial because "it is not packaged or preconceived, but process-oriented, evolving from teacher dialogue and reflection", as noted by Darling-Hammond (1995, p.244). Smith (2000) also found that interacting with colleagues regarding new ideas helped teachers to reform their teaching practices by resolving conflicts between their established views on teaching and learning, and new ideas.

Fourth, teachers' involvement in designing and implementing the assessment system will influence changes in their assessment practices by promoting their appropriation of the assessment and helping teachers to gain knowledge about assessment (Khatti & Sweet, 1996). From the cross-case analysis, Khatti & Sweet (1996) show that a low level of involvement in the development and implementation process inhibits teachers' appropriation of performance assessments that will influence their changes in assessment practices. When teachers are involved in the development or implementation process such as establishing standards of performance, designing assessment tasks and rubrics, and participation in scoring assessments, teachers are more likely to appropriate the assessment because of these opportunities to learn about the assessment as well as to work through problems and issues teachers faced when assessing student learning.

School Contexts in which Teachers Work as Contextual Factors

While individual factors such as teachers' beliefs and knowledge are crucial to the implementation of a reform, contextual factors, including teachers' working conditions, need to be considered for implementation of a reform. Since the examination of school context is so extensive and complex, I limited my discussion to three aspects of school context that were most relevant to performance assessment reform in South Korea.

One important school context factor that influences teachers' implementation of a reform is available time to experiment with reform ideas in their classrooms (Borko, Mayfield, Marion, Flexer, & Cumbe, 1997; Flexer & Gerstner, 1993; Khattri et al., 1995; Sung, 2000; Wilson, 1990; Wolfe & Miller, 1997). Bringing new ideas into classrooms needs time for planning lessons, preparing materials, teaching new lessons, and learning new ideas. If teachers have insufficient time for implementing new ideas, they may not be able to implement them despite wishing to. For example, the case study of a mathematics teacher who enacted the state mathematics reform showed that even if the teacher recognized that teaching for understanding, as envisioned by the reform, is more important than rules and procedures (traditional ways of teaching mathematics), he reported that he could only work on the rules and procedures with limited time.

Implementing performance assessment would require extensive time because it takes more time to design and score performance assessment tasks than to use traditional multiple-choice tests. Khattri et al.(1995) found that a lack of time for planning and developing these materials, and for scoring tasks, hindered teachers' efforts to adopt performance assessment. In Flexer and Gerstner's study (1993) on the dilemmas faced by teachers developing performance assessment in mathematics, teachers did not have

enough time to select and prepare performance assessments, and the time they did have seemed fragmented. Teachers wanted to try to use the new assessments if there was time. A study that surveyed barriers for the implementation of performance assessment in South Korea confirmed that preparation time is important for teachers' implementation of performance assessment (Sung, 2000).

How the school community, including students, parents, fellow teachers, and principals, responds to a reform is a second important school context factor in changing teachers' practices (Eisner, 1999; Flexer, Cumbo, Borko, Mayfield, & Marion, 1995; Khattri & Sweet, 1996; Tyack & Cuban, 1995; Wilson, 1990). Teachers may resist changing their practices when their school communities, accustomed to traditional practices, do not support the change. Khattri and Sweet (1996) demonstrated that without the support of the school community, teachers had difficulty in implementing performance assessment effectively. For example, whereas teachers in a district did not succeed in implementing assessment reforms due to community opposition to Vermont's assessment reforms, teacher' implementation of performance assessment in another district went well because most stakeholders supported the reforms. Thus, school community support for reform ideas is critical to the reform process.

School climate is also an important school context factor that influences teachers' practices (Olsen & Kirtman, 2002). Good school climate is important for initiating teachers' implementation of a reform. Supovitz et al.(2000) found that reform-oriented school climate significantly influenced teachers' initial practices. However, the impact of school climate on teaching practice seemed to disappear as they participated in inquiry-based professional development programs. They explained that quality of professional

development would help teachers to overcome their school climate. Examining 13 schools that implemented two reform efforts in Kentucky, Bulach & Malone (1994) also found that school climate was a significant factor in successfully implementing school reform.

Another school context factor is class size. The Korea Institute of Curriculum & Evaluation (1999) surveyed barriers for implementing performance assessment from elementary teachers. Among barriers reported by teachers, class size was the highest ranked barrier to implementing performance assessment. Since assessing performance assessment tasks is time-consuming, if teachers have a large class size, they may not want to use performance assessment or may not implement performance assessment as proposed.

In summary, while these contextual factors should be considered as important influences on teachers' implementation of reform, removing contextual barriers inhibiting teachers' implementation would not be sufficient. Grant (1998) observed that two teachers, in the same organizational context, read the same reform policy but interpreted it in different ways because of their different personal knowledge and beliefs. Thus, I hypothesize that teacher capacity and will, as individual factors, would have the most immediate (direct) relationship with teachers' responses to the reform, while contextual factors would influence teachers' implementation indirectly through these individual factors. Although teachers work in the same contexts, if they do not want to change or cannot make changes, real change would not happen in their classrooms. Final decisions for changing practices are in teachers' hands. Examining three elementary schools' implementation of a school reform to change teaching, Martinez (2005) found that teachers' knowledge and beliefs about teaching mathematics and expectations about their

students were the direct sources that influenced teaching practices. If the success of the performance assessment reform is directly influenced by teachers' capacity and will as individual factors, then teachers' learning opportunities will be crucial because teacher learning will play a key role in promoting teacher capacity and will.

CHAPTER 3

RESEARCH DESIGN AND METHODS

This chapter presents research design and methods. First, this chapter presents the research design for investigating research questions including a rationale for choosing a mixed research design. Then it describes methods of data collection including data collection instrument, data collection procedures, and details of participants. Finally, it describes methods of data analysis.

Research Design for the Present Study

This study used a mixed method design that combines quantitative and qualitative approaches. Specifically, this study used a concurrent nested mixed design model. The concurrent nested model has one data collection phase. That is, both quantitative and qualitative data are collected simultaneously, but this model has a predominant method that guides the study. The method that has less priority (qualitative or quantitative) is nested within the predominant method (qualitative or quantitative) (Creswell, 2003). In this study, quantitative data (survey) and qualitative data (interviews and documents) were gathered simultaneously, but the quantitative method has priority. The qualitative case study was incorporated in order to enhance and to inform the quantitative methodology.

The main reason for combining both quantitative and qualitative methods was that the qualitative case study complemented the quantitative method. While the quantitative methodology gave a broad, generalizable set of findings by obtaining responses from many people, the qualitative methodology permitted me to collect more

depthual and detailed information that the quantitative methodology could miss by focusing on a small number of people (Patton, 1990). Thus, this study combined surveys with the qualitative case study because the qualitative case study can add depth and details to the study as well as examine aspects of a phenomenon that the survey cannot capture.

For example, in this study, the main purpose of the survey was to assess teachers' implementation of the performance assessment reform and to identify what factors influenced their implementation. While the survey was good at examining how teachers generally implement each aspect of assessment practices (i.e., what kinds of performance formats of assessment teachers use, what kind of cognitive demands teachers assess), it was hard to use a survey to understand teachers' complex assessment practices in an integrated way. While the survey can investigate broad patterns of implementation looking at parts (how teachers implement each aspect of assessment practice), it was hard to design the survey to look at the sum (aspects intertwined and integrated). The case study examined what kinds of formats teachers use to assess the kinds of cognitive demands and how teachers use this assessment information. Thus, the case study was combined with the survey because the case study could examine different facets of teachers' implementation that would be hard for the survey to capture. Using both a survey and a case study allowed me to obtain a more complete understanding of the implementation.

Another reason for choosing a mixed method design was that information provided by teachers through the qualitative case study was used to triangulate the survey information. Patton (1990) noted that triangulation can solve the problem of relying too much on any single data source or method because using multiple sources or multiple methods enhances the validity and credibility of findings. Thus, combining the survey

with the qualitative case study permitted the study to have more convincing and accurate findings.

Although I integrate findings from both the survey and the qualitative case study, the purpose of using the mixed method design was not to address all questions in both method parts. Some findings are derived from conducting different methods, but some findings are derived because of the use of both methods.

Data Collection Instruments

Pilot Study

A pilot study was conducted from August 2002 through May 2004. The main purpose of the pilot study was to develop and test a questionnaire instrument for this study. Based on a review of literature, the survey was designed to tap teachers' assessment practices and factors influencing their practices. One hundred and eighty teachers were asked to complete the survey and 163 teachers in 19 schools returned a completed questionnaire each, a 91% response rate.

The pilot study was beneficial for the instrument development. First, the questionnaire instrument was revised based on feedback from field testing. The questionnaire was reformatted to facilitate easier reading, wordings of some questions were changed for clarity, and scales of items were changed. Also, some items were removed based on findings from reliability analyses.

Second, measures of implementation of the assessment reform as outcomes were revised. The survey included four measures of teachers' implementation of the reform: (1) formats of assessment, (2) cognitive demands of assessment, (3) purpose of assessment,

and (4) teachers' perceived implementation of the reform. The first three measures were developed based on literature on assessment practices. The fourth measure, perceived implementation, was added to assess teachers' overall implementation of the reform. The pilot study found that teachers' perceived implementation of the assessment reform was not correlated with other aspects of assessment practices such as formats of assessment and cognitive demands of assessment. This meant that teachers may think that they were implementing performance assessment well, although they did not use a variety of performance formats of assessment or assess higher-level of cognitive demands that have been emphasized by the performance assessment reform. That is, their perceived implementation of the performance assessment reform was unrelated to what they actually were doing. Thus, the measure, perceived implementation of the reform, was not included in the revised questionnaire.

Third, the pilot study helped me develop the interview protocol identifying what I could not know from the questionnaire data. For example, from the questionnaire data it was hard to see teachers' holistic implementation on format of assessment, cognitive demands of assessment, and purposes of assessment. While the questionnaire data were useful to look at each aspect of assessment practices generally, in practice these aspects were intertwined when teachers assess student learning. The questionnaire data could not show a more holistic picture of teachers' implementation of the performance assessment reform. Based on findings of the pilot study, the study combined qualitative methodology with the questionnaire data, and the pilot study helped develop the interview protocol to examine aspects of implementation of the reform that I could not find from the questionnaire data

Based on the pilot study, two instruments used for this study (questionnaire and interview protocol) were developed. Characteristics of instruments are described in the next section.

The Revised Questionnaire

The revised questionnaire comprised five parts: (1) background information, (2) teachers' assessment practices in social studies, (3) teachers' capacity & will including teacher beliefs about teaching and learning, teacher knowledge about performance assessment, and teacher willingness to implement the reform, (4) teachers' learning opportunities including professional development, collaboration, and involvement in school activities, and (5) school contexts, including school communities' attitude toward performance assessment, school climate, and available resources. The questionnaire is attached in Appendix A. The questionnaire included 14 main measures used in statistical analyses. These are four school contexts, four teachers' learning opportunities, three teachers' capacity and will, and three teachers' implementation measures.

Teachers' implementation of the reform. Three of the 14 measures have implementation of the mandated performance assessment reform as their outcome variables. These are: (1) formats of assessment, (2) cognitive demands of assessment, and (3) purposes of assessment. Formats of assessment refers to the particular types of tools teachers use to assess their students. Formats of assessment were categorized into two types: (1) traditional assessment formats (e.g., multiple-choice tests) and (2) performance assessment formats (e.g., project and portfolio). Cognitive demands of assessment refer to what kind of intellectual work is assessed. Items measuring cognitive demands of

assessment were categorized into two: (1) lower-level cognitive demands (e.g., recall factual knowledge) and (2) higher-level of cognitive demands (e.g., apply social studies concepts to real world problems, and make arguments with evidence). Purposes of assessment refer to how the teacher uses assessment results. Purposes of assessment were categorized into two: (1) assessment *of* learning (e.g., grading and providing report cards) and assessment *for* learning (e.g., improving student learning and improving teaching). Although the questionnaire included items measuring teachers' use of traditional assessments (traditional formats, lower-level cognitive demands, and assessment *of* learning), these were not used in the statistical analyses (regression and structural equation model) because my main focus was to examine what factors influence teachers' implementation of performance assessment. However, data on traditional assessments were used in graphs and box plots to look at whether there have been changes in use of assessment (traditional vs. performance assessment). Appendix B describes the items included in each teacher implementation measure.

Teachers' capacity and will. Teachers' capacity and will measures included: (1) teacher knowledge about performance assessment, (2) teacher belief about teaching and learning, and (3) teachers' willingness to implement reform. Teacher knowledge refers to teachers' understanding of performance assessment. Teachers rated their knowledge about five areas (i.e., what to assess and how to develop performance assessment tasks). Teacher beliefs refer to teachers' views about teaching and learning social studies. Items for measuring teacher beliefs were developed based on the instrument Ravitz, Becker, & Wong (2000) developed. Teachers' willingness to implement the performance assessment was measured by one item. Appendix B describes items included to measure teacher

knowledge, teacher belief, and teacher willingness.

Teachers' learning opportunities. The measures of teachers' learning opportunities included: (1) content of professional development, (2) pedagogy of professional development, (3) collaborations with fellow teachers, and (4) involvement in school activities. Content of professional development refers to the total hours of professional development (PD) programs that teachers spent to learn six content areas. Pedagogy of professional development refers to the learning activities teachers engaged in during their professional development programs. Items for measuring pedagogy of professional development were categorized into two: (1) traditional teacher learning (i.e., lecturing) and authentic teacher learning (i.e., doing projects and presenting or leading discussion). Collaboration with fellow teachers was measured by asking to what extent teachers had opportunities for collaborating with fellow teachers. Involvement in school activities was measured by summing the number of school activities in which teachers were involved. Appendix B describes items included to measure each of the teacher learning opportunities measures.

School contexts in which teachers work. Measures of school context in which teachers work included: (a) school communities' attitudes toward reform, (b) available resources, and (c) school climate. School communities' attitude toward the performance assessment reform was measured by asking how much teachers agree or disagree with the statement that school communities (i.e., parents and students) support teachers implementing the performance assessment reform. Resources were measured by asking how much teachers agree or disagree with the statement that time and resources available are sufficient. School climate was measured by five items asking whether the school

encourages teachers' new ideas (i.e., the school encourages teachers to try new ideas and teachers in this school are continually learning and seeking new ideas⁰. Appendix B describes items included to measure school contexts.

Interview Protocol

The interview protocol consisted of four parts: (1) questions about teachers' implementation of the performance assessment reform in terms of three aspects of assessment (formats, cognitive demands, and purposes of assessment), (2) questions about influences of their implementation of the reform, including teachers' understanding about performance assessment, their beliefs about instruction and assessment, willingness to implement the reform, and other factors on their implementation of the reform, (3) questions about their learning opportunities, and (4) questions about teachers' background. The interview protocol is attached to Appendix C.

Data Collection Procedures

Prior to data collection, a description of this study and the data collection instruments were reviewed and approved by the University Committee on Research Involving Human Subjects (UCRIHS). Three data sources were collected. One of the data sources was written-survey responses from teachers that were analyzed quantitatively. The other two sources were transcripts of interviews with teachers and documents. These two data sources were collected for the qualitative case study. In the following section, data collection procedures for each data source are described.

Questionnaire data

Questionnaire data were collected in two phases. The first set of data was collected from July to August and the second set of data was collected from October to November. To combine two sets of data, I compared demographic characteristics of participants in these two data sets. The following characteristics were compared: (1) school location, (2) school type, (3) teachers' gender, (4) teacher's highest degree, (5) teaching experiences, and (6) grade taught. Because there were no significant differences between the two sets of data, the data sets were combined for my future analyses.

Interview data

Interviews were conducted in late July or early August. Each interview lasted approximately from 90-120 minutes and was conducted after school in the teachers' classrooms. The interview was designed to elicit data on teachers' assessment practices and factors influencing their practices. A semi-structured interview protocol was used to ensure consistent coverage of important themes (See Appendix C), but each interview was allowed to take its own direction once it started. I added questions based on what each teacher told me during the interviews. Prior to starting the interview, participants consented to being tape recorded. Each interview was numbered and tape-recorded.

Documents

Participants in the case study were asked to identify and provide documents relevant to their assessment practices. These documents include school curriculum framework, school and grade assessment plans, examples of assessment tasks, etc. Documents were reviewed, labeled, numbered, and filed. The assessment tasks actually

used in classrooms were the most important evidence to show how teachers implement performance assessment reform. Examining the assessment tasks was helpful in looking at what kinds of assessment function teachers use to assess an assessment task.

Table 2

Summary of Data Collected	
Timeline	Data Collected
December, 2002 – January, 2003	Collected the pilot survey data
February, 2003 – June, 2004	Revised the survey instrument Selected a case study sample
July-August, 2004	Collected the first set of survey data Conducted interviews Collected documents
October –November, 2004	Collected the second set of survey data

Participants

Survey Participants

The revised questionnaire was distributed to elementary teachers who taught social studies in three areas of South Korea, because these areas are located near the capital, Seoul, where teachers and students are generally similar. Teachers in third, fourth, fifth, and sixth grades were selected because teachers in first and second grades taught an integrated curriculum of social studies and science, thus it would be hard to separate their assessment practices for each subject. Seven hundred teachers completed the questionnaire, and overall response rate was 74 %.

Since teachers were not selected randomly, I compared some characteristics (gender, school type, and teaching experience) of those who participated in my study to the characteristics of all teachers who taught in these areas. I found no big differences between the sample and the population. Percentages for gender and school types are similar, but the sample had a larger number of teachers who had 0 to 4 teaching years and a smaller number of teachers who had teaching experiences over 20 years. In other categories of teaching experiences, the percentages were almost the same. Table 3 presents a summary of the characteristics of the survey participants.

Table 3

Characteristics of Survey Participants

	Category	Frequency (%)	Total
School Type	Private	33 (4.7%)	700
	Public	656 (95.3%)	
Grade taught	3 rd	165 (23.9%)	691
	4 th	169 (24.5%)	
	5 th	193 (27.9%)	
	6 th	164 (23.7%)	
Gender	Male	110 (15.9%)	694
	Female	584 (84.1%)	
Highest Degree	Baccalaureate degree	557 (81.9%)	694
	Graduate degree	123 (18.1%)	
Teaching	0 to 4 years	245 (35.0%)	700
Experience	5 to 9 years	130 (20.0%)	
	10 to 14 years	79 (11.3%)	
	15 to 19 years	82 (11.7%)	
	20 years above	154 (22.0%)	

The majority of survey participants were working in public elementary schools

(95.3 %) and females (84.1%). Since 97.2 % of all the schools in the three areas are public and 77.3 % of teachers in all the schools in these three areas are females, the distributions of teachers in the sample were similar to those in the population. The majority of teachers (81.9%) had an earned baccalaureate degree, and 18.1 % of the teachers had an earned graduate degree. The distributions for the grade-level are similar among four different grade levels. While 35.0 % of teachers had four years or less of teaching experience, 33.7 % of teachers had teaching experience of 15 years or longer.

Case Study Participants

Case study participants were selected purposively from the respondents of the pilot survey who volunteered to participate in my future research. Because the purpose of the case study is to understand that teachers respond to the reform in different ways, a stratified purposive sampling strategy was used to include teachers who had participated in the pilot study and whose questionnaires suggested different patterns of implementation of the assessment reform. This sampling method permitted me to examine different patterns of assessment practices. Based on two dimensions of assessment practices – *formats* and *cognitive demands of assessment*, I searched for teachers who were in different stages of implementation of the reform: (1) teachers who implement the reform successfully in both aspects of assessment (performance assessment formats and higher-level of cognitive demands), (2) teachers who implement the reform successfully in only one aspect of assessment, and (3) teachers who fail to implement the reform in both aspects of assessment. For example, if a teacher used performance assessment formats at least “once a month” and gave at least “moderate” emphasis to higher-level of cognitive

demands on average, the teacher was considered as “implementing successfully in both aspects”. If a teacher “never” or “once a semester” used performance assessment formats and gave “little” emphasis to higher-level of cognitive demands, the teacher was considered as having “failed to implement in both aspects”. Two volunteer teachers from each pattern were selected as shown in Table 4.

Table 4

Selecting Case Study Participants

Cognitive Demands / Formats	Performance assessment	Traditional formats
Lower-level cognitive demands	2	2
Higher-level cognitive demands	2	2

Table 5 presents a summary of the characteristics of interview participants.

Table 5

Characteristics of Interview Participants

Teacher ID	School Type	Grade Taught	Gender	Highest Degree	Teaching Experience
1*	Public	5	Female	Bachelor	11 years
2*	Public	6	Female	Bachelor	10 years
3*	Private	6	Male	Master	16 years
4*	Public	6	Male	Doctoral	12 years
5	Public	3	Female	Bachelor	6 years
6	Public	5	Female	Bachelor	27 years
7	Public	5	Female	Bachelor	14 years
8	Public	4	Female	Bachelor, in a master program	1 year

* Teachers selected to represent four different responses to the reform.

Teachers in the case studies had a wide range of teaching experiences (one year to 27 years). This sample includes two female teachers who had only an earned baccalaureate degree and two male teachers who had earned a master degree or both a master and doctoral degree. One teacher taught in a private elementary school.

Data Analysis

Unit of Analysis

While the performance assessment reform was mandated nationally and suggests all teachers in South Korea use performance assessment, the mandate was for individualized, classroom-based performance assessment. While content and standards for performance assessment were based on the national curriculum framework, specific performance tasks and rubrics were developed by classroom teachers. Although teachers were in the same grade and same schools under the same mandate, their practices would not be the same. Thus, individual classroom teachers should be the unit of analysis for this study.

Quantitative Analyses

Survey data were analyzed in three steps. First, preliminary statistics are obtained using the Statistical Package of the Social Science (SPSS). Descriptive statistics were obtained to determine the distributional characteristics of each variable including the mean, standard deviation, skewness, and kurtosis. Next, correlations among variables were examined. Then, reliability analyses including Cronbach alpha and inter-subscale correlations were conducted to test the internal reliability of each measure.

After these statistics were examined and the variables were organized into constructs, a hierarchical regression analysis was run for two purposes: (1) to examine relative direct influences³ of three groups of factors (school contexts, learning opportunities, and teacher capacity and will) on teachers' implementation of the reform and (2) to test whether there were potential mediators on teachers' implementation of the reform. The order of entering these groups of factors was decided based on the literature review. Since influences of the school contexts and learning opportunities would be mediated by teachers' capacity and will and teacher learning opportunities would be the kind of factors that can act as individual and external factors, I first entered the school context factors. Then, I entered teacher learning opportunities factors followed by teacher capacity and will factors.

The third step was to engage in structural equation modeling using the Analysis of Moment Structures (AMOS) to test the adequacy of the hypothesized model. The structural equation allowed me to test hypotheses that specify direct and indirect relationships among latent variables as well as observed variables (Kline, 1998). In general, structural equation models consist of two sub-models: (a) a measurement model and a structural model. The measurement model defines relations between the variables as indicators of underlying constructs (the observed variables) and the underlying constructs that the observed variables were designed to measure (unobserved latent variables). The structural model represents relations among the latent variables. Combining measurement and structural models, Kline (1998) styled as the hybrid model.

³ I acknowledge that the language of "direct", "influences", or "effects" is customary with regression and SEM analyses, but I caution readers that my statistical analyses cannot test causal relationships. I, however, can identify the strength and valence of the relationships between the factors and teachers' implementation.

I used two-step modeling to test the structural equation model (Kline, 1998). The first part of two-step modeling is to find an acceptable measurement model. A model is first specified as a confirmatory factor analysis measurement. The CFA model is then analyzed to determine whether it fits the data. For the purpose of evaluating the overall fit indices are reported. A chi-square statistics is commonly used to evaluate model fit, but this statistics is known to be too sensitive to sample size (Kline, 1998). Thus, a number of fit indices are considered to evaluate structural equation models. Fit indices included the chi-square statistics adjusted by the degrees of freedom (χ^2/df), the Tucker-Lewis index (TLI), the Comparative Fit Index (CFI), the Incremental Fit Index (IFI), and the Root Mean Square Error of Approximation (RMSEA). χ^2 / df less than 3 is considered as a good fit. The value of less than .05 for RMSEA indicates good fit. For TLI, IFI, and CFI, values, values over .9 are generally considered as an acceptable model fit (Kline, 1998). If the initial measurement model indicates an inadequate goodness of fit, the model needs to be re-specified, based both on theories and residual matrixes. After modification, differences in χ^2 statistics were obtained to examine significant improvement of the modified models.

If the overall fit of the CFA model is acceptable, the second stage of two-step modeling is to test a structural model. Testing a structural model includes testing the structural paths between the latent variables as well as testing the overall fit of the hypothesized model. If the initial model indicates an inadequate overall fit, the model needs to be modified, based both on theories and residual matrixes. After modification, differences in χ^2 statistics were obtained to examine significant improvement of the modified models.

Analyses of Cases

The case studies involved analyses of interview data and documents. The interview data were analyzed via the following steps: First, interviews were transcribed into printed text. I then read the transcripts carefully to make a coding list. While I obtained a general sense of the information through that reading, I made a list of themes derived from my research questions and emergent from the reading of the data. A third step was to code the content of transcripts in relationship to these themes. The N-vivo computer software allowed me to categorize the responses. All responses related to each theme were placed in the approach theme category. Next, analytic case study narratives were written for each teacher (within-case analysis). Finally, cases were compared to one another to understand the patterns of similarities and dissimilarities across them (across-case analysis). Documents provided by teachers were analyzed along with interview transcripts. More details on analyses of cases are described in Appendix D.

Table 6 illustrates how the data sources and analyses correspond with my research questions.

Table 6

Summary of Data Sources and Data Analysis In Relation to Research Questions

Research Questions	Data Source	Data Analysis
<i>1. How do teachers respond to the performance assessment reform?</i>		
a. What kinds of formats of the assessment do they use?	Survey, interviews, documents	Quantitative, Case study
b. What kinds of cognitive demands teachers assess?	Survey, interviews, documents	Quantitative, Case study
c. What are teachers' purposes of assessment?	Survey, interviews, documents	Quantitative, Case study
<i>2. What are the factors that influence teachers' implementation?</i>		
● Teacher capacity and will factors:		
a. Does teachers' belief influence their implementation?	Questionnaire	Regression, SEM
b. Does teachers' knowledge influence teachers' implementation?	Questionnaire	Regression, SEM
c. Does teachers' willingness influence teachers' implementation?	Questionnaire	Regression, SEM

Research Questions	Data Source	Data Analysis
<ul style="list-style-type: none"> Teacher learning opportunities: <ul style="list-style-type: none"> d. Does the content of professional development influence teachers' implementation? e. Does the pedagogy of professional development influence teachers' implementation? f. Does collaboration with fellow teachers influence teachers' implementation? g. Does involvement in school activities influence teachers' implementation? School contexts factors <ul style="list-style-type: none"> h. Does school community influence teachers' implementation? i. Do school resources influence teachers' implementation? j. Does school climate influence teachers' implementation? 	<ul style="list-style-type: none"> Questionnaire Questionnaire Questionnaire Questionnaire <ul style="list-style-type: none"> Questionnaire Questionnaire Questionnaire 	<ul style="list-style-type: none"> Regression, SEM, graphs SEM, graphs SEM, graphs SEM, graphs <ul style="list-style-type: none"> Regression, SEM Regression, SEM Regression, SEM

Methodological Caveats

Two important caveats about this study merit attention. The first caveat is that this study relied on self-report data. Thus, I caution that my conclusion is based on teachers' reports on their implementation of the reform and factors influencing implementation. There may have been some biases in teachers' responses, but I attempted to correct problems with self-report data. Problems with self-report data can be corrected in several ways. First, problems can be corrected in part by using more operational or behavioral categories (descriptive neutral terms) and avoiding judgmental language (Garet et al., 1999). For example, instead of asking direct opinions about quality of their implementation of performance assessment, measures about their implementation represent an accounting of behaviors. Second, problems with self-report data (interview) can be corrected in part by organizing interview questions around specific documents. For example, instead of asking teachers only about what and how to assess student learning, I asked them to bring the assessment tasks they used and then asked interview questions to understand their assessment practices. Third, problems with self-report data can be corrected in part by delimiting the period of the reference (Schwarz, 1999). For example, since I surveyed near the end of a semester, teachers could do a good job of recalling their assessment practices for that semester. Fourth, problems with self-report data can be corrected in part by using composite scores rather than single indicators (Mayer, 1999). For example, in my questionnaire, in order to examine assessment formats, I used a composite score combining four items: project, portfolio, demonstration, and observation. This procedure could obtain higher validity and reliability of self-report data

The second caveat is that my statistical analyses do not test causal relationships since the data is cross-sectional. All information was collected at the same time and there is no direct evidence that one set of variations in one variable, School Community (School Contexts) preceded and therefore caused the second set of variations in, for example, Collaboration (Learning Opportunities) (Miller, 1999). Thus, I acknowledge that the language of “direct”, “influences” or “effects” is customary for regression and SEM analysis, but I caution readers that I do not draw inferences about the direction of causation for the relationship. I cannot rule out the possibility that there is an alternative ordering of the relationship among variables, but I can identify the strength and valence of the relationships among variables, and the path model I hypothesize is supported by prior research as reviewed in Chapter 2.

CHAPTER 4

TEACHERS' IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT

This chapter examines how teachers implement the assessment reform. Both the questionnaire and case study data were analyzed to provide a comprehensive understanding about teachers' implementation of the reform. This chapter is organized into two sections. The first section starts by presenting cases showing that teachers responded to the reform in a variety of ways. Four different patterns in implementing the reform were presented in terms of three key aspects of the reform: (1) *formats* of assessment, (2) *cognitive demands* of assessment, and (3) *purposes* of assessment. The second section reports questionnaire data on teachers' implementation of the reform in terms of these three aspects. While the case study helped me describe teachers' different response to the reform with more in-depth and detailed information, the questionnaire data was useful to examine how teachers implement each aspect of assessment practices generally with a larger sample size.

Four Cases of Teachers' Responses to the Reform

In analyzing the transcripts of interviews and documents related to teachers' assessments, I attempted to identify different patterns in implementing the assessment reform in terms of three key aspects of the reform: (1) *formats* of assessment, (2) *cognitive demands* of assessment, and (3) *purposes* of assessment. From analysis of the interview sample of eight, four cases that show the different implementation patterns were

selected as representatives of each pattern. These four cases present similarities and contrasts in patterns of assessment practices. Teachers in the case study responded to the reform in a variety of ways, but seemed to work under the similar school assessment system in the context of the national curriculum and assessment reform.

All teachers' schools required teachers to use new formats of assessment. Following the mandate of the national curriculum and assessment reform, schools encouraged teachers not to rely heavily on traditional paper-and-pencil tests that required selected responses, but rather to increase use of new formats of assessment. Based on the national curriculum and assessment framework, each school revised its own curriculum and assessment framework.

Following the school assessment framework, teachers in the same grade designed the grade assessment framework containing details of assessment areas, assessment units, assessment method, and assessment period. In most schools, mainly a teacher in charge of planning the grade assessment designed this grade assessment framework. She or he usually provided samples of assessment tasks for teachers. Each assessment task mainly consisted of a specific unit or lesson for assessment, assessment questions, and assessment criteria.

Grade assessment plans collected from teachers interviewed appeared similar in three aspects of assessment practices. First, three broad areas of assessment were included: (a) knowledge/understanding, (b) inquiry/thinking skills, and (c) disposition. Second, teachers were encouraged to use a variety of assessment methods including both traditional formats and new formats of assessment. Although traditional paper-and-pencil tests were included in grade assessment plans, those plans emphasized use of new formats

of assessment. While paper-and-pencil tests were included for assessing the knowledge area, assessment tasks were designed for assessing two other assessment areas: inquiry and thinking skills and disposition. Third, the grade assessment plans appear to be designed for providing students with report cards. Based on the grade assessment plan, each teacher assessed students' learning.

Because of the centralized educational system in South Korea, teachers' assessment practices were expected to be similar, but in fact showed significant variations. Four of the eight teachers interviewed were selected for case studies because they typified trends in this variation. In the next section, the complexity of teachers' implementation of performance assessment reform will be explored via the four cases. Investigating the different responses to the performance assessment reform of Messrs. Choi and Kim, and Mmes. Lee and Park will help us understand four patterns of implementation.

The four patterns of performance assessment implementation were identified based on both teachers' views and intentions about teaching and learning, and teachers' practices in the three aspects of assessment. Figure 1 shows different patterns in implementation of the reform.

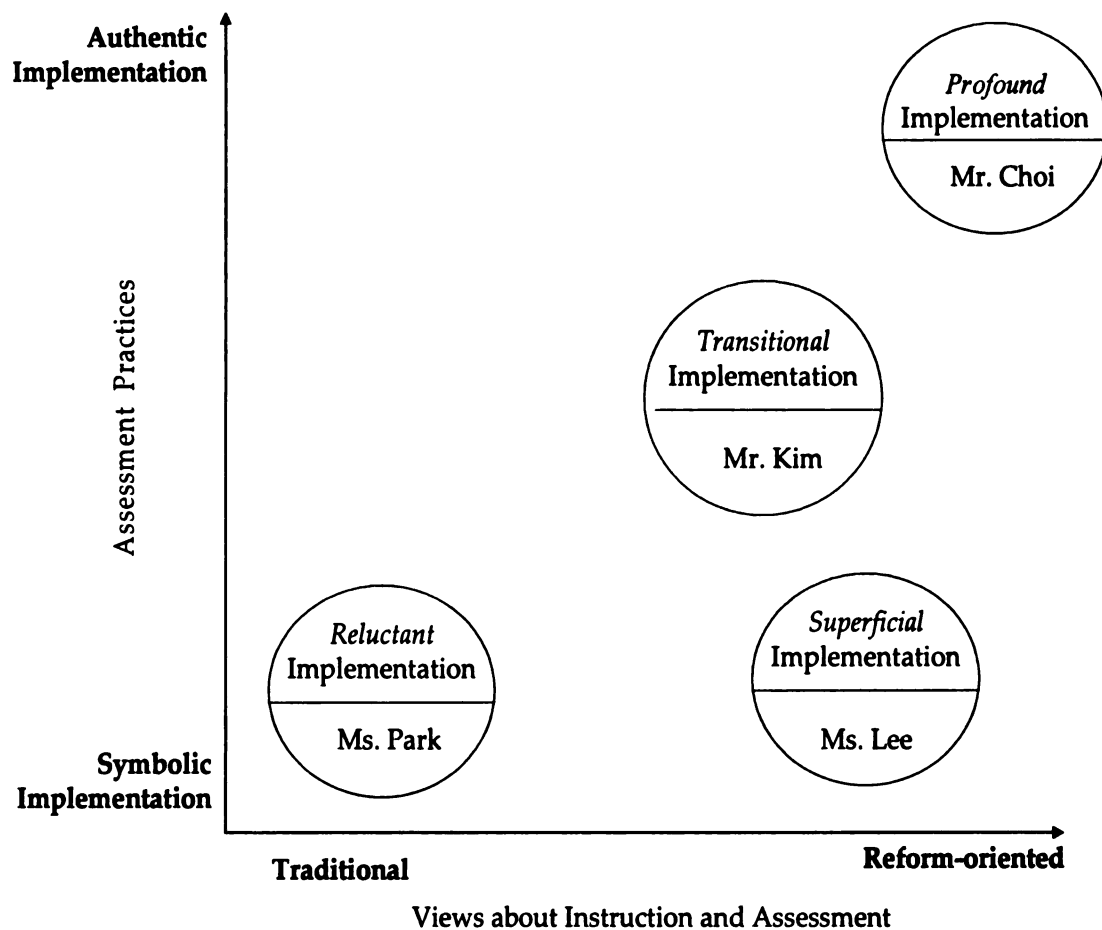


Figure 1. Four Patterns in Implementing the Performance Assessment Reform

The vertical axis in Figure 1 shows teachers' actual practices in response to the reform according to implementation of three aspects of assessment: formats, cognitive demands, and purposes of assessment. The high end is *Authentic Implementation* and the low end is *Symbolic Implementation*. Whereas Symbolic Implementation addressed only the letter of the law (formats of assessment), Authentic Implementation addressed both the letter (formats of assessment) and the spirit of the law (cognitive demands and purposes of

assessment). The horizontal axis in Figure 1 shows teachers' views about instruction and assessment. The right edge is more reform-oriented views, and the left edge is more traditional views. The four teachers were located in Figure 1 by examining both their views and their practices.

Mr. Choi is located at the right on the bottom axis and at the high on the right axis, exemplified *Profound Implementation* since his reform-oriented views about instruction and assessment were well aligned with his practices, which successfully implemented all three aspects of the reform: formats, cognitive demands, and purposes of assessment. Mr. Kim, located in the middle on the right axis, but leaning toward the right on the bottom axis, exemplified *Transitional Implementation*. Mr. Kim's views on instruction and assessment seemed close to reform-oriented, but he mixed traditional and reform-oriented assessment practices; however, since he was moving toward the profound implementation, his practices were labeled as Transitional Implementation.

The other two teachers, Ms. Lee and Ms. Park, showed symbolic implementation since they implemented only the surface element of the reform (formats of assessment). These two teachers were located at the low end on the right axis. However, they had different views and intentions about instruction and assessment. Ms. Lee was located at the very right on the bottom axis since she intended to implement the principles of the reform with more reform-oriented views. Ms. Lee's case showed an example of *Superficial Implementation* because she intended to implement the reform principles, but could only integrate formats of assessment into her classroom. In contrast, Ms. Park is located at the very left on the bottom axis since she did not want to implement the reform with very traditional views about instruction, but merely complied with implementation of

reform as she has been ordered. Ms. Park's case showed an example of *Reluctant Implementation* merely compliant with implementation of reform formats. Four case studies of profound, transitional, superficial, reluctant implementation were as follows.

Pattern 1: Profound Implementation

Mr. Choi exemplified profound implementation of the mandated assessment reform. He appeared to successfully integrate all three aspects of the performance assessment reform into his social studies classes: formats, cognitive demands, and purposes of assessment. Employing performance assessment formats, he assessed students' higher order thinking as well as used assessment results for improving student learning.

Mr. Choi is an experienced teacher who has been teaching for 12 years total in public elementary schools. He is currently teaching a sixth-grade class. He reported that most students in his classroom performed at an average level. He received his bachelor's degree in general education and master's degree in social studies education. He recently received his doctoral degree in social studies education. Throughout his interview, he showed strong interest in teaching social studies: "I am more confident in teaching social studies than other subjects. I majored in social studies education in the undergraduate program. I experimented with many new ideas in my social studies classes."

Formats of assessment. Mr. Choi chose to use performance assessment formats exclusively. He used a variety of performance assessment formats: essays, projects, portfolio, demonstrations, and observations. He most often used writing reports from inquiry projects as an assessment format. Mr. Choi provided students with topics for

students' inquiry projects. Students had to decide on their own research questions based on a topic. After researching their own questions, students wrote reports. Sometimes students went on field trips. After returning from field trips, students wrote reports about what they had learned from those trips. Before they went on field trips, the teacher helped students decide what to investigate while on their trips, then provided students with guidelines for writing reports and assessed their reports based on the guidelines.

Mr. Choi used the traditional multiple-choice paper-and-pencil test once a semester because sixth-grade teachers decided to take sixth-grade tests that included four subject tests: language arts, mathematics, social studies, and science. Four teachers made test questions for each subject and Mr. Choi was responsible for social studies. Although he employed the tests in his classrooms, he preferred not to use this type of assessment format because the tests were not aligned with what he taught to his students. Pointing out the mismatch between instruction and assessment, he said:

I don't like this type of tests. I think that assessment should be related to what students are taught by their teacher. Since these paper-and-pencil tests required students in the same grade to take the same tests, questions included in the tests were based on the students' textbooks. Since I use the textbook as a resource and cannot cover all of the content represented in it, I do not think that the results of this type of test can show what my students have learned in my classroom.

Mr. Choi considered the scores obtained from the traditional multiple-choice test as only a supplemental source of information when he wrote students' report cards. It seemed that he gave more weight to scores from performance assessment tasks when describing students' knowledge in their report cards. Mr. Choi said:

Teachers who teach the same grade usually give grades in Knowledge based on students' scores on traditional tests. Whereas I think that traditional tests can easily measure whether students acquire knowledge of facts, but these test scores alone are enough to say what students know. If a student is good at writing reports, but does not obtain a good score on a traditional test, I do not give a bad grade in Knowledge when writing her report card. I believe more in the scores from performance assessment tasks. Traditional test scores are used as references when I write report cards.

While Mr. Choi attempted to integrate various formats of performance assessment into his classroom, his students suffered from the new formats of assessments. The new formats of assessment asked them to show their thinking in their own words, but his students struggled with the new assessments. Although the performance assessment reform had been mandated several years earlier, the students still did not know how to express their own thinking. They were used to accepting historical facts represented by social studies textbooks. He stated the difficulties he initially faced when he used performance assessment formats: "Students had difficulty expressing their own thoughts in writing. When I initially asked my students to write inquiry reports, they just copied. I kept telling to my students that you should write in your own words."

Since the students were unfamiliar with new assessment formats, Mr. Choi taught them how to do well in the new assessments. The way he trained was to guide students by giving assessment criteria before the students did assessment tasks. The assessment criteria reflected his goal of student learning. For example, he used writing journals to train students to show their thinking in their own language. The journals comprised two sections. In the first, students wrote their thoughts about the news articles provided by the teacher. The second section required students to collect and organize other news articles. After that, students wrote their thoughts about the news articles. Before assigning writing

journals, Mr. Choi provided his students with the assessment criteria, such as supporting argument with evidence and suggesting alternative solutions to the problem. Based on the assessment criteria, he gave comments to the students on the journals, and the students revised their journals or tried to reflect Mr. Choi's comments when they wrote another journal entry. Articulating his expectations before assessing student work, Mr. Choi helped the students to build their confidence in communicating their own thinking about content.

Cognitive demands of assessment. Mr. Choi assessed both content knowledge and higher order thinking, employing the performance assessment formats. He had clear targets for assessing his students. Four areas were assessed: (1) Knowledge (Do students know facts, concepts, and generalizations?) (2) Inquiry (How do students collect, analyze, and synthesize information?), (3) Decision-making (Do students express their arguments about social issues and provide solutions?), and (4) Disposition/Attitude (Do students show interest in or positive attitudes toward learning social studies?)

Mr. Choi believed that knowledge and thinking skills such as Inquiry and Decision-Making should not be assessed separately, and that new formats of assessment would be more beneficial to see how students solve problems using what they know. He stated:

I think that the good thing about using performance assessment is to that it allows me to examine what my students understand by looking at their performance. I can assess what a student knows using new formats of assessment. Both Knowledge and Thinking skills can be assessed with these types of formats. Traditional paper-and- pencil tests have a limitation. When using this type of test, it is difficult to assess both of these assessment areas.

This view on assessment was reflected in the assessment tasks Mr. Choi used. Using this assessment task, he assessed both students' understanding about content and students' thinking. For example, in a history unit, students learned about the ancient Korean who established applied science in Korea. After finishing the unit, Mr. Choi asked students to write an essay on the question: How would you write a public statement for reforming people assuming that you are one of the applied scientists? He said that he wanted to assess students' abilities to develop their arguments based on their understanding of content. Three criteria were developed to assess this essay: (1) Did the student understand the arguments of the different applied scientists?, (2) Did the student's essay reflect the problems of the time?, and (3) Did the student write his or her arguments showing appropriate evidence? The first two criteria were to assess students' understanding about the content and the third was to assess students' thinking skills.

What Mr. Choi wanted to assess seemed closely related to his social studies instruction goal. He said, "A question about what to assess is a question about what to teach because assessment and instruction cannot be thought of separately." In his history classes, he emphasized understanding of multiple historical interpretations. He stated:

Students think that the historical knowledge represented in their social studies textbook is always true. I tried to get students to understand that historical knowledge in the social studies textbook is one of the interpretations created by historians. I taught that historical interpretation can be different based on different historical perspectives. Instead of memorizing the historical facts represented in the social studies textbook, students need to learn the importance of interpretation and to build their own perspectives.

Mr. Choi's goals of teaching history were connected to his assessment. One of the assessment tasks he used showed this connection. After students learned about a historical

unit on the Japanese colonization of Korea, each student was asked to compose an imaginary letter to a Japanese elementary student about how historical knowledge was distorted in documents written by the Japanese. Mr. Choi wanted to assess whether his students could apply content knowledge to a real world problem. Using this assessment task, he wanted his students to look at historical knowledge from multiple perspectives as well as to establish their arguments based on their analyses of historical data.

Analyses of these assessment tasks indicated that he wanted his students to learn beyond factual knowledge, and in order to assess more than only recall of factual knowledge, he used new formats of assessment rather than traditional paper-and-pencil tests.

Purposes of assessment. Mr. Choi used performance assessment results to improve student learning. He expressed the importance of feedback for improving student learning. While he was required to provide students and parents with report cards, his main purpose for assessment was not to provide report cards. He mentioned that report cards could not show what he assessed and did not help students and parents identify students' strengths and weaknesses. As he stated:

Report cards consist of four areas, Knowledge, Inquiry, Decision-making, and Disposition. These areas were graded with three levels — “excellent”, “good” and “needs effort” — in spring semester and were described in greater detail in fall semester. Neither grades nor written descriptions were enough to present the assessment results obtained from my assessments. If I use grades, students and parents just know the general status of the students' learning. If I use written descriptions, I cannot provide details on student learning because of limited space on the form. I can write just one or two sentences about what the students learned so I usually write what students did well, such as “He's good at collecting data” or “He's good at making arguments”.

He attempted to innovate in his assessment practices based on reform principles. Since the report cards required by the school did not fit his purposes for assessment, Mr. Choi used informal ways to communicate assessment results with students and parents. The most often used way was to provide direct comments on students' essays or reports as well as a one-page description about assessment results for parents. The one-page description included what the student did well, and what the student needed to improve. To allow parents to think about their child within the class, it also provided the student's grade and a paragraph on problems common to most students in the class. Based on the teacher's comments, students had to revise their essays or reports if the teacher asked them to.

Mr. Choi had a dilemma because he did not want to give grades — he preferred providing written descriptions, but parents and students wanted to know students' status. Instead of choosing either option, he decided to provide both numbers and descriptions about the assessment results. He said:

Students feel pressure to study only when they take paper-and-pencil tests. They think that only numbers on paper-and-pencil tests show what they know. I tried to change this belief. I tried to explain that even though a student obtained good scores in paper-and-pencil tests, the student did not have good thinking skills. However, changing this belief was not easy. I had to make a compromise to solve this problem. I decided to provide numbers as well as written descriptions when I assessed essays or reports. I gave scores out of a total of 100 instead of using A, B, C. Even though I disliked providing numbers, I decided to do that because this approach motivated students and helped them learn more.

While Mr. Choi used both numbers and written descriptions to communicate assessment results with parents and students, it is clear that his main purpose for assessment was to improve student learning by helping students revise their reports based

on his feedback, rather than by only assigning grades. He was more concerned about students' learning progress than about their mastery of content.

Pattern 2: Transitional Implementation

Mr. Kim exemplified transitional implementation. Mr. Kim had views about instruction and assessment that seemed close to reform-oriented, but he mixed traditional and reform-oriented assessments moving toward the reform practices. While he could integrate performance assessment formats into his classroom, his practices in other two aspects of assessment (cognitive demands and purposes of assessment) mixed traditional and reform oriented assessments.

Mr. Kim is also an experienced teacher who has been teaching for sixteen years total in a private elementary school. He is currently teaching a sixth-grade class. He reported that his students performed very well. He received his bachelor's degree in administration and master's degree in counseling. He was interested in trying new teaching methods in his classroom and most willing to implement the performance assessment reform.

Formats of assessment. Mr. Kim advocated a balanced approach between traditional and new formats of assessment. His typical social studies class showed how he balanced use of both traditional and new assessment formats. A description of Mr. Kim's typical social studies class follows.

Mr. Kim used project-based instruction to teach a history unit. Before starting the unit, he provided to students the topics or issues related to lessons in the unit. Students formed teams of four. After choosing a topic or an issue, members of the team scheduled

time to work on their project together and assigned a role for each member. Projects were conducted out of class. Students met to carry out their project after class or discussed via Instant Messenger. Every lesson started with a team's presentation. The team presented what they had learned doing their project. The team had to answer the teacher's or other students' questions. If the team missed some important points, the teacher supplemented the presentation. While each team was presenting their work, the teacher assessed their learning based on his observations. After the presentation was done, the teacher employed a timed, short-answer quiz to see whether students had understood the lesson well.

This description showed that Mr. Kim used various assessment formats including several new formats of assessment, such as projects, presentations, and observations, as well as traditional paper-and-pencil tests. One type of assessment format that was not included in this description was the mid-term and final paper-and-pencil tests. Most questions were multiple-choice and some required extended responses.

Cognitive demands of assessment. Mr. Kim showed interest in assessing both content knowledge and student thinking. In contrast to Mr. Choi, he believed that traditional assessment formats could assess both content knowledge and thinking skills. He felt confidence in his ability to redesign traditional paper-and-pencil tests to assess both content knowledge and student thinking. He stated:

The paper-and-pencil tests I designed have three types of questions: basic, applied, and advanced. The basic questions consist of multiple-choice items, and a correct answer from each item gets one point. Applied questions assessed inquiry skills, such as interpreting maps. Advanced questions assess students' thinking skills, such as comparing two arguments and writing your position. Because I included these three types of questions, my traditional paper-and-pencil tests assessed thinking skills.

Even if Mr. Kim could redesign the traditional paper-and-pencil tests to assess basic reasoning, such as explaining how and why, the questions he asked were unable to assess more complex thinking, such as applying what students know to a real problem or justifying their arguments through evidence, because the students had to finish the questions within a given time frame.

Mr. Kim's desire to assess both content knowledge and thinking skills was not reflected in the questions asked in the performance assessment formats. For example, students went on a field trip to a historical site (*Kyungbok Palace*) that was related to a unit of the study. Before going on the field trip, Mr. Kim provided students with a booklet of questions for them to answer on their return, based on their observations of the palace and the information they had collected. Most questions in the booklet assessed knowledge of facts, such as "What does the name of the palace [written in Chinese characters] mean?" and "When was it constructed?" His students were asked to find the "right" answers to these questions. Mr. Kim assessed their answers based on whether students responded as he intended.

While most questions in the booklet elicited students' factual knowledge, some questions engaged students' thinking skills. However, Mr. Kim actually used the questions to assess student effort, instead of thinking skills. For instance, the booklet asked the students to compare palaces in other countries with one palace they had visited, to think about different historical contexts. When grading the students' completed booklets, Mr. Kim assessed students' answer by looking at how had the students had worked to complete the task, rather than at what the students had thought about the content. If a student provided a long answer, Mr. Kim thought that the student had worked hard to

complete the booklet and he rewarded this with a good score.

Mr. Kim's idea about assessment — rewarding students for their efforts — appeared connected to the primary focus of his instruction: motivating students to enjoy social studies. He said:

I asked students whether today's class was fun. I do not think that test scores are important. Even though a student receives good scores on tests, if the student does not show any interest in learning social studies, my instruction is not successful, I think.

Mr. Kim used performance assessment tasks to assess student effort because he believed that student effort meant student motivation and student motivation was his primary goal. It appeared that his assessment focused more on students' efforts to finish tasks and less on how they thought about the content. For example, students did team projects on social studies topics given by the teacher. After completing the projects, they presented what they had learned from their projects to other students. Mr. Kim gave scores to students observing student presentations and reading student reports. Scores were based on his judgment about whether students invested much effort in completing their projects. As he stated:

I gave 10 points out of 10 if they worked hard when they prepared for their projects. I observed whether they answered my questions and other students' questions well. I can see how hard they worked by the way they present their projects.

Purposes of assessment. Mr. Kim believed that the purpose of assessment is to see whether student learning is progressing, rather than to compare an individual's test score to other students' scores. He said:

Performance assessment was used in past years with traditional paper-and-pencil tests. I think that the performance assessment emphasized in the current reform was different than the performance assessment used in past years. For example, a student got 40 points out of 100 in March and then she got 50 points out of 100 in April because she studied hard. Even though her score is below average, her score has improved. How should I assess this student? I want to give a good grade to this student because her learning has improved.

Another purpose for assessment he emphasized by Mr. Kim was to motivate student learning. He believed that tests would intensify the pressure to succeed, and, as a result, students would try to study harder and learn more. He said that, “To increase learning, we should increase student anxiety. Comparison with more successful peers will motivate low performers to do better.”

Mr. Kim’s assessment practices mixed these two purposes – motivating students and considering learning progress. Since he cared so much about motivating students, he chose to encourage a team competition. After a team’s presentation had finished, as described above, all students in the class had time to study what the team had presented, by looking at their notes. Each team studied together to help their members earn higher scores on the test. Then, students had to take a timed, short, open-ended quiz. Mr. Kim evaluated students’ answers and each team received a team score that combined both an average score of all team members and the team member’s individual score. How Mr. Kim scored the test reflected another of his assessment purposes: considering students’ learning progress. An individual score was a gain score calculated by subtracting the previous test score from the current test score. Based on how many scores had improved, a student derived his or her progress score. Mr. Kim expressed that assessing in this way was helpful in motivating all students to study hard to earn higher test scores:

Students listen to the presentation very carefully. They carefully read the materials provided by the presentation team and make notes so as not to forget what the team presented. While the team is presenting, I emphasize points that students have to know. After the presentation ends, each team studies together. If there is a student who has low academic achievement in the team, another student with high achievement in the team will help that student to learn. This is peer learning. Students with low achievement feel a lot of pressure to study because their team's score will decrease if they get low scores. Also, students with high achievement feel pressure to study because their team's score will decrease if their scores on the current test are lower than their previous test scores. Thus, all students study hard.

Pattern 3: Superficial Implementation

Ms. Lee exemplified the superficial implementation. She brought reform-minded views of instruction and assessment, but could only implement the reform at the surface level. Whereas she could integrate new format of assessment into her classroom, she failed in assessing higher order thinking using the performance assessment formats as well as in using assessment results for improving teaching and learning.

Ms. Lee has been teaching for ten years total in public elementary schools. She is currently teaching a sixth grade classroom. She reported that most students in her classroom performed below average. She received her bachelor's degree in elementary education. She expressed that she liked learning social studies when she was a student, but she was frustrated with teaching social studies in her classrooms: "I like history, political science, and economics when I was student. It was fun to know that there are connections among historical knowledge I knew in fragments. I want my students to know that so they are getting interested in learning social studies. However, I find it so challenging. I have been so frustrated teaching social studies."

Assessment formats. Ms. Lee used more new formats of assessment, as intended by the reform, to replace traditional assessment formats. Her beliefs about assessment seemed to influence the decrease of traditional paper-and-pencil tests. She said that she did not use traditional paper-and-pencil tests because she thought that they decreased students' interests in social studies learning by requiring recall of knowledge. She stated:

I didn't use multiple-choice tests. I don't have enough time to employ paper-and-pencil tests. Also, I don't want to test recall of knowledge using these types of tests. My students don't have sufficient readings so they lack basic historical knowledge. Thus, if I require students to recall knowledge by employing paper-and-pencil tests, students seems to lose their interest in social studies. My students hate to memorize something."

Another reason for not using multiple-choice tests was that content of the tests did not fit what her students learned. She said:

Sometimes my school provides me with samples of traditional paper-and-pencil test developed by publishing companies. These tests consist of multiple-choice, matching, or short-answer questions. The national assessment reform did not allow teachers to use this type of test developed by outside class but teachers sometimes tried to use it because it was easy to employ. I also used this type of test in my classroom just once. My students' test scores were terrible. I found that the focus of the test was not the focus of my instruction. This type of test required students to recall facts such as what are the most important exports of China? I did not ask my students to memorize this kind of knowledge in my classroom. So I disposed of the results without showing students and parents. After then, I never used this type of the test.

The most often used new format was report-based inquiry projects that took at least one week Topics or problems for inquiry projects were given by her, but students could choose what they wanted to investigate. After completing their projects, students wrote reports and presented in class. She assessed both students' reports and presentations.

Another new format of assessment used often by the teacher was tasks with open-

ended questions. These tasks were used approximately ten minutes before finishing a lesson. After students did the presentation, she distributed a task to the students. The students answered the questions based on what they listened to in the presentation or social studies textbooks.

Cognitive demands of assessment. While Ms. Lee chose more employment of performance assessment formats, she failed in assessing higher order thinking such as inquiry skills or application of their knowledge to solve a real world problem. For example, she chose to use many assessment tasks with open-ended questions that are formats of performance assessment. However, these assessment tasks asked students to find right answers from the social studies textbooks.

However, Ms. Lee wanted her students to know more than factual knowledge. Her description about teaching on a unit showed that her instruction focuses on understanding big ideas or concepts in history by asking why and how questions. She said:

The first unit is about different kingdoms that existed in what is now Korea. The main focus of instruction was not to tell facts to students. Knowing just names of different kingdoms or events that happened is not enough for learning history. I asked, “Why and how were kingdoms unified?” I asked, “What were the contexts in which a kingdom declined”? I talked about these kinds of questions with my students. “What kinds of victories occurred?”, “what person was famous in those days?” Students can see answers to these kinds of questions in social studies textbooks and I do not think that these questions are very important for students to know. Instead, I think that my students should understand the process of the rise and fall of kingdoms. This concept can be connected to current situations our country faces.

Despite her attention to understanding big ideas or concepts through student thinking, Ms. Lee undermined her own instructional purposes when assessing student learning in several ways. First, there was a mismatch between her instructional purposes

and her assessment tasks — what she was asking her students to do in her history class differed from what she was asking them to do in the assessment tasks. While she emphasized big ideas or concepts, rather than factual knowledge in her history class, the assessment tasks with open-ended questions appeared to ask right answers from the social studies textbook because she used only the assessment tasks developed by other teachers. That is, although she chose a performance format, the assessment task did not elicit students' higher thinking to solve the questions. Thus, she wanted her students to learn more than factual knowledge by asking 'why and how' questions, but could not assess what she emphasized in her classroom. Although she said that, "In order to do this task, students do not need to memorize. This is a sort of open-book test", there were right answers in the social studies textbook and the students were asked to search the answers from the social studies textbook. That is, although she wanted to assess understanding the big ideas and not memorizing (e.g., open book quiz), what she asked students to do was to find right answers from the social studies textbook.

Second, Ms. Lee's assessment criteria were mismatched with her purposes of instruction and assessment. The way in which she assessed the new formats of assessment such as reports and presentation appeared to assess how much information the student knows. The assessment tasks provided opportunities for students to engage in thinking processes such as collecting information via Internet and summarizing what they found, but what the teacher actually assessed was not thinking skills. She merely looked at mastery of facts by assessing how much information the students wrote. Also, her assessment focused more on assessing how hard students work and less on how they think. If a student wrote many information items about the topic, it would be understood that the

student knew the content well. She stated:

I asked to examine the unification process of three regions. When assessing the reports, I looked at whether students provided more information about this process. For example, a student wrote about the unification process: a region was first unified with another region, a person played an important role in this unification process, and then the region was unified with the rest of the country. This student showed breadth of knowledge. I expect students to write down all the information they know about this process. If another student wrote just one or two sentences about this, I would understand that she lacks content knowledge.

Third, her beliefs about the role of the textbook were mismatched with her goals of instruction and assessment. She appeared to believe that she had to cover all content in the social studies textbook and students needed to know that content. With a given, limited time frame, she had to choose between covering much content and promoting student thinking. Since Ms. Lee believed that all content had to be covered, she sacrificed depth (promoting student thinking) for breadth (covering all of content). She said:

Assessment tasks focused only on content. Since time for assessing content is not enough, I cannot think about assessing thinking skills. I should not think that students who do have content knowledge can have good thinking skills. However, it is very hard for me to make time for assessing thinking skills. Actually, we do not have time to cover content in the social studies textbook. Since I want to teach more content, I cannot think about teaching and assessing thinking skills.

While Ms. Lee's assessment practices showed a shift from over-reliance on traditional paper-and-pencil tests to more employment of new formats of assessment, what she actually assessed was mainly content knowledge represented in the social studies textbook, rather than how students use content knowledge to solve a problem that has been emphasized by proponents of performance assessments.

Purposes of assessment. Ms. Lee failed in using assessment results for the new purpose of assessment: *assessment for learning*. Her main use of the assessment results was to provide students and parents with report cards, her summary judgment about each student. However, she showed her dissatisfaction with her current use of assessment.

I feel that I am doing assessment for assessment's sake. I think that in order to improve student learning, I need to provide more details about an individual student's learning, to parents. For example, a student showed interest when talking about social issues, but the student lacks historical knowledge. I also think that only one issue of report cards per semester is not enough. I ask myself if assessment [of learning] I am now doing is really needed.

Ms. Lee applied the new purpose of assessment (assessment for learning) to a disabled student. She assessed the student, looking at her progress of learning. Ms. Lee believed that assessment should consider both students' starting points and current points. She believed that, assessment should be based on how the student's learning improved, rather than comparing a student with another student. This belief about assessment was reflected when she assessed a disabled student. As she stated:

I had a disabled student. I knew how much the student's learning improved from the beginning to the end of the semester. My assessment for this student was very informative because I observed the student very carefully. The different way in which I assessed the student was reflected in what I wrote in his report card. In the report card, I could provide the student with more details about assessment. The assessment results were not based on comparison with other students, but focused on his improvement of learning.

Although Ms. Lee was less successful in implementing reform purposes of assessment (assessment for learning) for most students, the fact that she considered assessing for student learning when she assessed the disabled student, is a positive sign.

This indicates good possibility for moving toward the profound implementation if she has more competence or more support.

Pattern 4: Reluctant Implementation

Ms. Park exemplified Reluctant Implementation. She was similar to Ms. Lee in that she only implemented the surface element of the reform (formats of assessment), but she differed from Ms. Lee and Messrs. Kim, and Choi in that she saw the content to be memorized and had no dissatisfaction with traditional assessments. This traditional view was aligned with her assessment practices (e.g., heavily relying on traditional formats to assess recall of knowledge).

Ms. Park has been teaching for 11 years total in public elementary schools. She is currently teaching a fifth grade classroom. She reported that most students in her classroom performed averagely. She received her bachelor's degree in elementary education. She expressed that she liked learning social studies when she was a student. She seemed to resist implementing performance assessment because she did not feel need of change in her assessment practices. However, she accepted the implementation of performance assessment because of the mandate.

Formats of assessment. While Ms. Park could integrate new formats of assessment into her classroom, she relied heavily on traditional assessment formats such as quizzes or unit tests. Traditional paper-and-pencil tests appeared related to her belief about assessment. She believed that assessment was to check whether students find right answers. She expressed difficulties when she used new formats of assessment with the new curriculum. She stated:

The new curriculum is different from the previous curriculum. For example, students were asked to answer problems of the city. Students' answers would be insufficient houses or a lot of cars. They do not say why houses are insufficient. If I ask why satellite cities appeared, students do not say anything. In the previous curriculum teachers can tell this and students accept what teachers said. Tests were also easy to employ in the previous curriculum. However, assessment for the new curriculum requires teachers to use open-ended questions that do not have right answers. All students' answers are right in this kind of assessment. This caused a problem. I said that there was no wrong answer when students answered the questions. After that, students took paper-and-pencil tests. I graded the tests based on the number of right answers. After receiving test scores, students complained that their scores were unfair. I think that there are right answers and students should know that right answers are represented in social studies textbooks.

Based on such a belief about assessment, Ms. Park employed paper-and-pencil tests with right answers after finishing every unit. Also she sometimes used quizzes when students did not pay attention to the teacher's lecture.

Ms. Park's use of new formats of assessment seemed to be symbolic compliance with the assessment reform, since all elementary teachers have been required to incorporate new formats of assessment into their assessment practices. She said: "I know that I cannot avoid using performance assessment because the national curriculum and assessment reform require me to use it." Ms. Park's use of portfolios showed that she used new formats of assessment, but her assessments did not reflect what the reform emphasized. She described how she used portfolios:

I asked my students to collect anything they have done and file it in a scrapbook. I also asked them to put unit tests into the scrapbook. When the students planned to take mid-term and final tests, I asked them to look at their results on the unit tests. And I said, "the test will be based on the questions I gave in the unit tests. Try to look at your wrong answers and remember the right answers".

Cognitive demands of assessments. Major focus of assessment was on recalling facts. She believed that there students should memorize many facts when they learn social studies. She said: “When I was a student, I memorized a lot in learning social studies. My students were slow with memorizing facts. That is a problem.”

To assess recall of knowledge, Ms. Lee used many paper-and-pencil tests. She believed that traditional paper-and pencil-tests were the definitive assessment for showing what they know by checking whether they find right answers. Another reason why she preferred traditional paper-and-pencil tests is that new formats of assessment assess too narrow areas of knowledge. She said:

Performance assessments ask too specific topics or issues within a unit but traditional paper-and-pencil tests can assess broader areas. I use performance assessments because I was required to use them but I think that quizzes or unit tests are better to check what students know.

Ms. Park’s reason for using performance assessment formats was not because of assessing student thinking. She used the new formats because the performance assessment tasks seemed interesting activities for students. For example, the teacher asks students to design an advertisement about Kimchi (a traditional Korean food) after learning about a unit on Korean weather. She mentioned that students had fun doing this, but she did not have any assessment targets related to the unit.

Another performance assessment task Ms. Park provided was a task with open-ended questions. The task asked students to write about problems of cities and to suggest solutions to one of problems. The tasks required students’ thinking skills such as application of their knowledge to a real world problem. However, what she actually assessed was whether students know problems of cities. She did not assess how students

solve problems using their knowledge.

Ms. Park assessed students' higher order thinking skills using traditional and pencil tests. She seemed to believe that students who received good scores in paper-and-pencil tests could apply to a real problem what the students know. She explained how she assessed thinking skills based on paper-and-pencil tests.

For example, the grade assessment plan suggested I employ a task to assess whether students can draw a chronological table. Since I did not employ this task, I used the information obtained from paper-and-pencil tests for this task's grading. The paper-and-pencil tests included several questions about a chronological table. If students had good scores in these questions about it, I assumed that the students can draw a chronological table.

Purposes of assessment. Ms. Park did not use assessment for student learning. She had two prominent purposes for assessment. One purpose is to provide report cards. She expressed that she was busy with assessment at the end of semester because she had to submit her own grade-book and write report cards. She had to employ assessments if she missed assessment areas included in the fifth grade assessment plan. Based on the grade assessment plan, she had to write report cards. Another purpose was to check whether students acquire right content. She said that test scores from traditional paper-and-pencil tests were the main sources for writing report cards because tests scores showed whether students know the right answers or right conclusions.

What does it mean to Implement Performance Assessment?

Looking at implementation patterns would be important to understand that teachers' implementation is in progress. While all teachers have implemented the performance assessment reform as they have been ordered, their implementation would not progress at the same speed. All teachers believed that they were responding to their reform yet their practices differed considerably.

Of all eight teachers in the interview sample, only one showed *Profound Implementation*. The teacher implemented all aspects of the reform successfully. First, his assessment practices shifted from over-reliance on traditional paper-and-pencil tests to more employment of new formats of assessment. Second, new formats of assessment were used to assess students thinking skills that have been emphasized in the new curriculum and assessment. Third, his assessment purpose was not only assessment of learning, but also assessment for learning.

Two teachers showed *Transitional Implementation*. While they attempted to implement the substance of the reform, their practices felt short of implementing one of the deep aspects of the reform (cognitive demands and purposes of assessment).

Five of eight teachers interviewed showed *Symbolic Implementation* that indicates adopting the reform without making substantial changes. The teachers integrated new formats of assessment into their classrooms, but failed in implementing the deep aspects of the reform (cognitive demands and purposes of assessment). However, there were differences in these teachers' implementations.

Two teachers in this case strongly intended to implement substantial aspects of the reform, but appeared to implement the reform superficially, transforming only an easier

aspect of the reform (i.e., formats of assessment). These teachers' implementation patterns were labeled as *Superficial Implementation*. The other three teachers showed *Reluctant Implementation*. The difference between Superficial Implementation and Reluctant Implementation was the nature of compliance with the reform. While the teacher in Superficial Implementation showed substantial (voluntary) compliance with the reform, indicating that the teacher strongly wanted to implement the reform, the teacher in Reluctant Implementation showed merely symbolic compliance with the reform, indicating that the teacher did not want to implement the reform, but had to show his or her compliance, as ordered. In other words, while teachers within the Symbolic Implementation attempted to use new formats of assessment, teachers in the Superficial Implementation employed new formats of assessment because they believed in the need for new assessments for assessing students' higher order thinking, but the teachers in this Reluctant Implementation did so because they were required to, by the mandate of the assessment reform.

These case studies showed that the mandated assessment reform seemed to start teachers to use the new assessments in classrooms in some way, but teachers' responses to the reform varied. While some teachers implemented the reform at the deeper level, some teachers implemented the reform at the surface level. Identifying the different responses to the reform helps us to understand that teachers did not implement the reform at the same speed. If a teacher has sufficient capacity to implement the performance assessment reform, the teacher may reach the Profound Implementation phase rapidly. However, another teacher may take time to move to the Profound Implementation since he or she does not know how to use performance assessment although he or she really wants to

implement the performance assessment reform. If extensive professional development or other support is given, it would accelerate implementation.

Now I turn to the second section examining teachers' implementation of each aspect of performance assessment using the questionnaire data. The case studies imply that many teachers may have failed in implementing the substantial aspects of assessment (assessing higher-level cognitive demands and assessing for student learning), while they could integrate new formats of assessments more easily. These findings from the case studies will be reexamined to test whether they are consistent with findings obtained through the questionnaire data with a larger sample size.

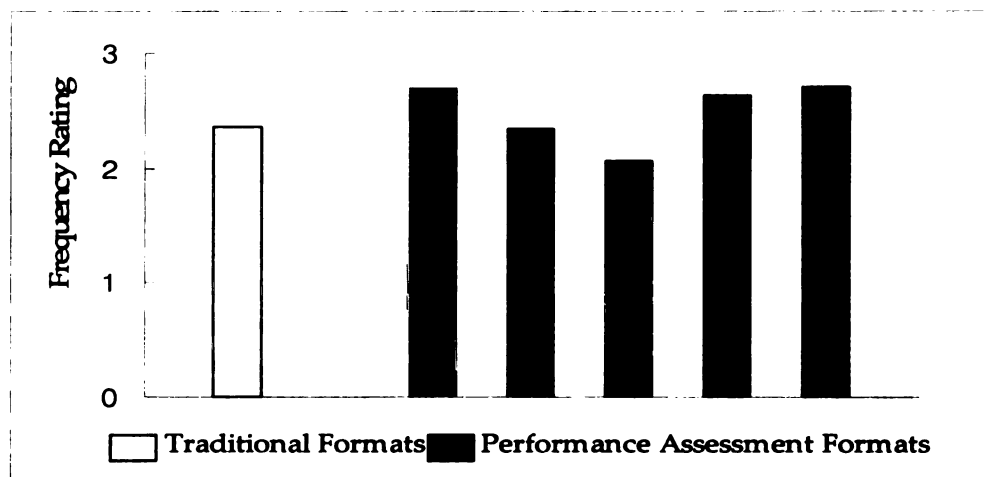
Teachers' Reports on Their Implementation of the Reform

This section presents teachers' reports on their implementation of three aspects of the reform: formats, cognitive demands, and purposes of assessment.

Formats of Assessment

The formats were categorized into two. One type is traditional paper-and-pencil tests consisting of selected response questions such as multiple-choice, true/false, matching, fill-in-the-blank and/or completion. Another type is new formats of assessments such as short-constructed response tasks, extended-response tasks, portfolio, demonstration, and observation. Teachers reported how frequently each type of format was used in the social studies classroom. Rating of frequency was coded as a 1-5 scale⁴.

Figure 2 presents frequency of use of different formats.



Note: TF = Selected responses; PA1=Short-open ended; PA2=Project/Reports; PA3=Portfolio; PA4=Demonstration; PA5=Observation

Figure 2. Frequency of Use of Different Assessment Formats

⁴ 1 = Never; 2= Rarely (once a semester); 3=Sometimes (once every other month); 4=Often (once or twice a month); 5=Very often (once a week)

This figure suggests that teachers' use of new formats has increased. Average frequency of use of each type was between 2 and 3. On average, teachers used both types of assessment formats once a semester or once every other month, on average.

Short-constructed tasks were the oftenest used assessment formats. Since the assessment reform has suggested that teachers reduce use of multiple-choice tests, the assessment reform may have influenced reducing teachers' frequency of multiple-choice tests. However, teachers merely may have shifted from employing multiple-choice tests to short-constructed response tasks because this format would be easier to employ among new formats of assessment.

Among new formats of assessment, demonstration and observation were oftener used than were project and portfolio. When compared to demonstration and observation, project and portfolio assessment would not be easy to employ since these formats recently have been emphasized in assessment reforms, so teachers are unfamiliar with using them in their classrooms

Figure 3 shows which type of formats, traditional and new, each teacher used oftener. Two scores were created. One was to measure frequency for use of traditional assessment formats, the other to measure frequency for use of reform-oriented formats. Each teacher received a score reflecting the extent to which she or he used the two different sets of assessment formats. The score was calculated by subtracting the average score of traditional formats from the average score of using reform-oriented formats to see whether teachers showed a balance between their use of the two. For example, if the average score of using reform-oriented formats for a particular teacher is 3 and his or her average score of using traditional formats was also 3, the difference would be 0,

indicating that the teacher shows a balance between two different types. If the average score of using reform-oriented formats for another teacher was 4 and his or her average score of using traditional formats is 1, the difference would be 3, indicating that the teacher more often used reform-oriented formats.

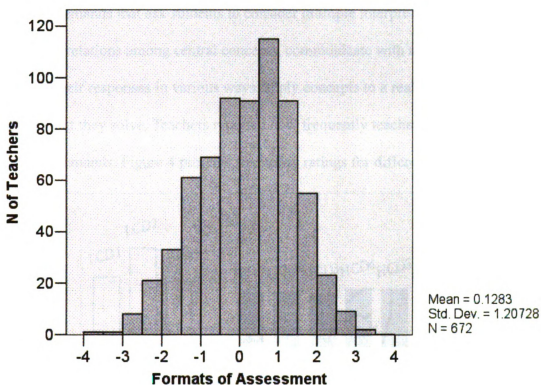
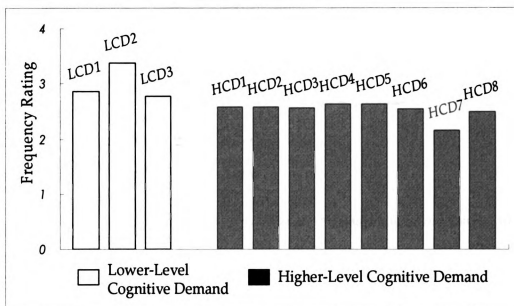


Figure 3. Formats of Assessment

This histogram illustrates that the majority of teachers exhibited balanced use between traditional formats and new formats, because their scores cluster in the center of scales, toward zero. While many teachers had scores toward zero, there is a slight leaning toward the positive end of the continuum, indicating that new formats have an edge in teachers' assessment formats. This finding indicates that teachers have moved from over-reliance on traditional formats to more use of new formats.

Cognitive Demands of Assessment

Cognitive demands were categorized into two types. One type is lower level of cognitive demands that ask students to recall factual knowledge, find answers from social studies textbooks, or find only one right answer/solution. The other is higher level of cognitive demands that ask students to consider multiple interpretations/perspectives, understand relations among central concepts, communicate with teachers or peers, represent their responses in various ways, apply concepts to a real world problem, or explain what they solve. Teachers reported how frequently teachers assessed the following cognitive demands. Figure 4 presents frequency ratings for different cognitive demands.



Note: LCD1= Recall factual knowledge; LCD2 = Find answers from social studies text books; LCD3 = Find only one right answer/solution.; HCD1 = More than one solution; MCD2 = Interpretation; MCD3 = Relations among concepts; HCD4 = Various representations; HCD5 = Application; HCD6 = Explanation; HCD7 = Argument with evidence; and HCD8= Use of inquiry skills

Figure 4. Frequency Ratings for Different Cognitive Demands of Assessment

Figure 5 shows which type of cognitive demands, lower level and higher level, each teacher assessed oftener.

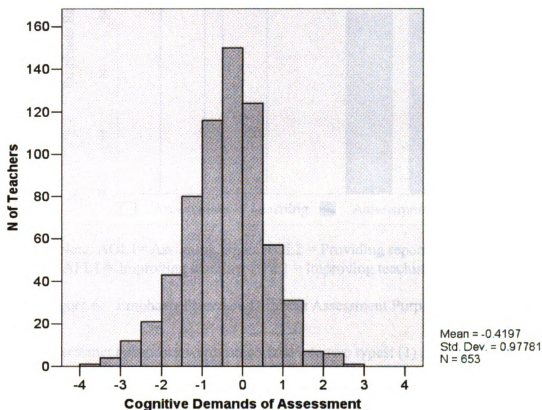
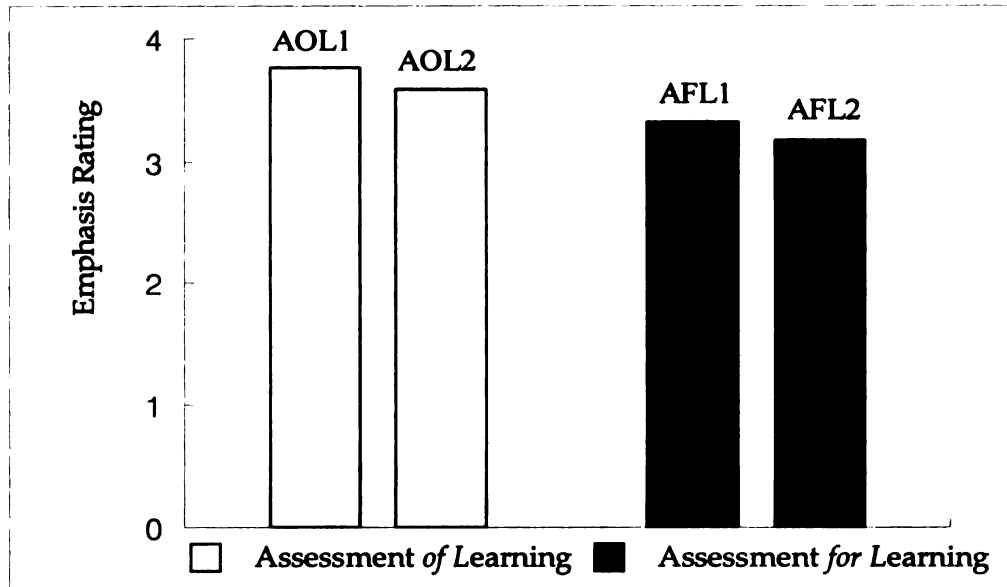


Figure 5. Cognitive Demands of Assessment

While many teachers had scores toward zero, there is a noticeable leaning toward the negative end of the continuum, indicating that lower level cognitive demands have an edge. This result indicates that many teachers still have focused more on assessment of lower-level cognitive demands than on assessment of higher-level cognitive demands.

Purposes of Assessment

Teachers reported how much emphasis is given to the following purposes for their assessment. Figure 6 shows the ratings of emphasis placed on different assessment purposes.



Note: AOL1= Assigning grade; AOL2 = Providing report cards;
AFL1 = Improving learning; AFL2 = Improving teaching

Figure 6. Emphasis Placed on Different Assessment Purposes

Assessment purposes were categorized into two types: (1) *assessment of learning* and (2) *assessment for learning*. *Assessment of learning* means that teachers use assessment for giving grades to students or providing parents with report cards. *Assessment for learning* means that teachers use assessment for improving student learning and teaching. Figure 6 shows a large difference between how much emphasis teachers placed on two different types of purposes: *assessment of learning* and *assessment for learning*. The strongest emphasis was on giving grading and the weakest emphasis on improving teaching.

Figure 7 shows which type of assessment purpose each teacher used oftener. While many teachers had scores toward zero, there is a noticeable leaning toward the negative end of the continuum, indicating that many teachers focused more on *assessment of learning* rather than *assessment for learning*.

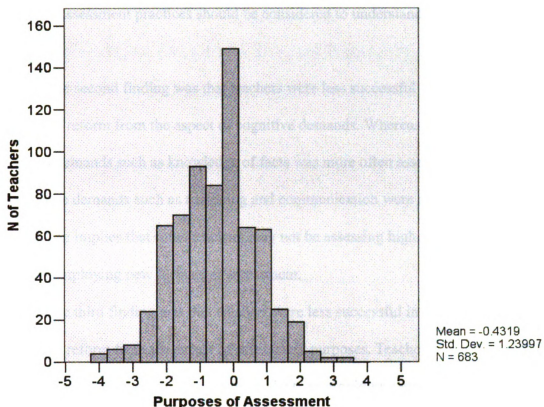


Figure 7. Purposes of Assessment

Summary of the Second Section

Three findings were identified. One was that teachers were successful in implementing the reform from the aspect of formats of assessment since teachers' assessment practices have been shifted from over-reliance on traditional formats and increased use of new formats. However, the frequency of using new assessment formats would not be enough to say that the reform is implementing well. For example, if a teacher used projects often but did not use projects for assessing students' higher order thinking, what can be said about the teachers' implementation? If another teacher used a multiple-choice test only once a semester but students' total scores in their report cards reflected only that test score, what can be said about teachers' implementation? Other

aspects of assessment practices should be considered to understand implementation of the reform.

The second finding was that teachers were less successful in implementing the assessment reform from the aspect of cognitive demands. Whereas the lower level of cognitive demands such as knowledge of facts was more often assessed, the higher level of cognitive demands such as reasoning and communication were not often used to assess. This finding implies that some teachers may not be assessing higher level of cognitive demands employing new formats of assessment.

The third finding was that teachers were less successful in implementing assessment reform from the aspect of assessment purposes. Teachers reported that they gave more emphasis to giving grades to students or providing report cards to parents. Teachers seemed to struggle with using assessment results for the purpose of assessment for learning.

In conclusion, teachers' implementation of the reform seems unsuccessful from these three aspects. Although the mandated performance assessment reform pushed teachers to use more new formats, it has not been successful in eliciting more fundamental changes in their assessment practices because the mandate did not make teachers change more critical aspects of assessment practices, i.e., cognitive demands and purposes of assessment. These findings were consistent with what I found from the case studies because more than half the teachers were less successful in implementing substantial aspects of assessment (cognitive demands and purposes of assessment), but could integrate performance assessment formats into their classrooms. The next chapter will examine factors influencing teachers' implementation of the reform.

CHAPTER 5

FACTORS INFLUENCING IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT

This chapter presents findings on factors influencing teachers' implementation of the reform, and comprises three sections. This chapter begins with results of preliminary analyses which contain descriptive statistics and the correlation matrix. The second section presents findings from the hierarchical regression analysis exploring the relationships between teachers' implementation of the performance assessment reform and factors: school context factors, teacher learning opportunities factors, and teacher capacity & will factors. The chapter concludes with a report of the findings from the structural equation modeling to examine direct and indirect effects of factors influencing implementation.

Preliminary Analyses

Before examining the specific research questions, some preliminary analyses were conducted. The first step in those preliminary analyses was to examine descriptive statistics for each variable included in this present study to describe basic features of data: minimum value, maximum value, mean, and standard deviation. The second step was to examine correlations among the independent variables, as listed in Table 7, and their correlations with the outcome measure, teachers' implementation of the mandated assessment reform, as listed in Table 8. Tables 7 and 8 display descriptive statistics for the variables.

Table 7

Descriptive Statistics for the Independent Variables

Category	Factors/Variables	Variable Symbol	N	Min.	Max.	Mean	SD
School Contexts	<i>School Community</i>						
	Parent support	SC1	688	1	6	4.32	1.02
	Student support	SC2	691	1	6	4.07	1.13
	Principal support	SC3	686	2	6	5.11	.89
	Colleague support	SC4	685	1	6	4.40	1.02
	<i>School Climate</i>						
	Encouraging new ideas	SCL1	678	1	6	3.74	1.26
	Innovative teachers	SCL2	679	1	6	3.68	1.16
	Discussion with new ideas	SCL3	678	1	6	3.72	1.18
	Keep learning new ideas	SCL4	675	1	6	3.66	1.32
	PD for new ideas	SCL5	678	1	6	4.35	1.01
	<i>Resources</i>						
	Time	R1	693	1	6	2.63	1.30
	Resource	R2	687	1	6	2.77	1.24
	Class Size	CSIZE	684	24	53	36.67	4.72
Teacher Learning Opportunities	<i>Collaboration</i>						
	Discussing	CO1	678	1	5	2.47	.87
	Working together	CO2	682	1	5	2.06	.76
	<i>Involvement</i>	Involvement	679	1	6	1.72	1.24
	<i>Specific PD Hours</i>						
	Social studies assessment	PD1	650	1	5	1.54	.87
	Social studies PA	PD2	651	1	5	1.55	.89
	General curriculum	PD3	667	1	5	2.99	1.55
	General PA	PD4	653	1	5	2.03	1.20
	Social studies curriculum	PD5	659	1	5	1.87	1.10
	Social studies instruction	PD6	654	1	5	1.67	1.02
Teacher Capacity & Will	<i>Teacher Beliefs</i>						
	Learning	TB1	688	1	7	4.98	1.58
	Curriculum	TB2	686	1	7	4.95	1.50
	Role of teacher	TB3	689	1	7	4.26	1.84
	Student activity	TB4	687	1	7	3.77	1.85
	<i>Teacher Knowledge</i>						
	What to assess	TK1	690	1	5	3.23	.82
	Assessment tasks	TK2	693	1	5	2.77	.87
	Rubrics	TK3	693	1	5	2.71	.83
	How to score	TK4	691	1	5	2.95	.85
	Use of PA in social studies	TK5	692	1	5	2.97	.82
	<i>Teacher Willingness</i>	Willingness	696	1	7	4.42	1.20

Table 8

Descriptive Statistics for the Outcome Variables Related to Teacher Implementation

Factors/Variables	Variable Symbol	N	Min.	Max.	Mean	SD
<i>Formats of Assessment</i>						
Project	AF1	699	1	5	2.34	1.01
Portfolio	AF2	695	1	5	2.07	1.08
Demonstration	AF3	695	1	5	2.64	1.14
Observation	AF4	699	1	5	2.70	1.07
<i>Cognitive Demands</i>						
More than one solution	CD1	686	1	5	2.57	1.07
Interpretation	CD2	689	1	5	2.57	1.09
Relations among concepts	CD3	690	1	5	2.56	1.12
Various representation	CD4	691	1	5	2.63	1.03
Application	CD5	687	1	5	2.63	1.03
Explanation	CD6	692	1	5	2.54	1.14
Argument with evidence	CD7	691	1	5	2.15	1.05
Use of inquiry skills	CD8	687	1	5	2.49	.98
<i>Purpose of Assessment</i>						
Improving Learning	AP1	691	1	5	3.32	.98
Improving Teaching	AP2	687	1	5	3.17	.69

The means and ranges of all of the variables of interest were inspected first for signs of data entry errors and outliers. Then, normality, i.e., the assumption that each variable should be normally distributed, was assessed by evaluating the skewness and kurtosis of each variable in the study. When the distribution of a variable is normal, the values for skewness and kurtosis are zero. The examination of statistics for skewness and

kurtosis indicated that all values for univariate skewness and kurtosis were inside the acceptable range (-3 to 3 for skewness and -8 to 8 for kurtosis, Kline, 1998).

The correlations among all independent variables and their correlations with the dependent variable (teachers' implementation of the mandated assessment reform) are presented in Appendix C. Correlation coefficients of the independent variables with the dependent variable ranged from .04 to .49. A major task with a large database such as this is finding a way to reduce the number of variables into a small number of coherent constructs. The sections below describe my approach to this problem.

School Contexts and Teacher Implementation

Most school context variables, as listed in Table 7, were significantly related to teachers' implementation ($p < .05$). The correlation coefficients of the independent variables with the dependent variable ranged from .04 to .23. The correlations are presented in Appendix C, Table C1. Among the contextual factors, parents' positive attitude toward the reform had the highest correlation coefficients. This finding reflects that parents play an important role in South Korean education and teachers' practices are influenced strongly by parents.

Two variables were found to be insignificantly related to implementation. One variable is class size, but there was complication: in my survey data, there were no small classes to compare to large classes. Table 9 shows that no one reported a class size less than 20. Only 2.9 % of teachers taught fewer than 30 students. More than half of the teachers each taught more than 35 students. With so few teachers who had a small class size, the small and insignificant correlation between class size and teachers'

implementation of the reform shown in the correlation matrix may be meaningless. With this limitation, it would be hard to say that class size is unrelated to teachers' implementation, in the present study. Since all teachers in the case study complained about large numbers of students in their classrooms, class size would be an important factor influencing teachers' implementation of the reform, but this study could not test the relationship between class size and teachers' implementation with the range restriction problem. Thus, because of the limitation of the sample, class size will not be included as a variable in this present study.

Table 9

Frequency for Class Size

Class size	Frequency	Percent	Cumulative percent
24-29	20	2.9	2.9
30-34	240	35.1	38
35-39	220	32.2	70.2
40-45	173	25.3	95.5
Over 45	31	4.5	100.0

Another variable that did not have a significant relationship with implementation was principals' attitudes toward the reform. There are two possible explanations for this finding. The first explanation is that there are few principals who have a negative attitude toward the reform. Only 3.7 % of teachers reported that their principals did not have a positive attitude toward the reform. This is unsurprising since under the centralized system, an important role of principals is to distribute policy to teachers: in this case, the

mandated assessment reform.

Given a principal's influential role in a school, the second explanation is that it is not so much the principal's attitude toward the reform that influences teachers' implementation as it is the principal's understanding of the reform. For example, even if a principal has a positive attitude toward the reform, if the principal pushed teachers to go in a wrong direction based on his misunderstanding of the reform, teachers would not implement the reform as intended. That is, when a principal with positive attitude toward the reform had the capacity for helping teachers to implement the reform as intended, teachers would better implement the reform.

In preparation for future analyses, these school context variables were grouped into three constructs, which are the factors influencing teachers' implementation of the mandated assessment reform: *School Community*, *School Climate*, and *School Resources*. The first is School Community combining four variables: the principal's, fellow teachers', parents', and students' positive attitude toward the reform. All four variables representing School Community were positively and significantly correlated with each other, so I grouped them together as a construct. The second is School Climate combining five variables, as listed in Table 7, that encouraged teachers to new ideas. All five variables representing this construct were positively and significantly correlated with each other, so I combined these variables into a construct. The third is School Resources combining two variables, available time and resources for implementing the reform. These variables were combined into School Climate. Available time and resources were positively and significant correlated with each other, so I grouped these variables as a construct.

Teacher Learning Opportunities and Teacher Implementation

All learning opportunity variables, as listed in Table 1, except one variable related to professional development (PD), were significantly related to teachers' implementation ($p < .05$). The correlation coefficients of all learning opportunities variables with the dependent variable ranged from .06 to .20. The correlations are presented in Appendix C, Table C2. Among them, variables measuring teachers' collaboration had the highest correlation coefficients. Since these two variables were positively and significantly correlated with each other, these variables were combined into *Collaboration* construct.

Among the six PD variables relating to hour spent on PD, as listed in Table 7, PD hours spent on social studies and social studies performance assessment had higher correlations to teachers' implementation of the reform than those spent on social studies curriculum and teaching. PD hours spent on general curriculum, which did not focus on social studies were not significantly correlated with teachers' implementation of the reform. This finding indicates that teachers who took PD that focused on the subject teachers taught (social studies), and that related more closely to performance assessment, had higher teacher Implementation scores. While the strength of the correlations between Teacher Implementation and six PD variables differed, all PD variables were positively and significantly correlated with each other. Although this study was interested in influences on Teacher Implementation of PD hours spent on different contents, these PD variables could not be included separately in future analyses, since all PD variables were highly correlated with each other. Since two variables, hours spent on social studies assessment PD and hours spent on social studies PA had high correlations with Teacher

Implementation, I selected them to compose the *Specific PD Hours* construct for future analysis

Based on the examination of correlations, learning opportunities variables were grouped into three constructs, which are the factors influencing teachers' implementation of the performance assessment reform: *Collaboration*, *Specific PD Hours*, and *Involvement*. Since teachers' involvement in school activities was measured by only one item, this variable was considered as a factor.

Teacher Capacity & Will and Teacher Implementation

Teachers' implementation of the reform was significantly related to all teacher capacity & will variables measuring teacher knowledge about performance assessment, beliefs about teaching and learning, and willingness to implement the reform, as listed in Table 7. The correlation coefficients of all teacher capacity & will variables with teachers' implementation of the reform ranged from .12 to .32. The correlations are presented in Appendix C, Table C3. When compared to the correlations with contextual variables measuring School Community, Climate and Resources and learning opportunities variables measuring Specific PD Hours, Collaboration, and Involvement, correlations between teacher capacity & will variables and teachers' implementation of the reform were higher. Variables measuring teacher knowledge and teachers' willingness to implement the reform were more highly correlated with teachers' implementation than were variables measuring Teacher Beliefs about teaching and learning.

All teacher knowledge variables were found to be positively and significantly correlated with each other. Thus, these variables were combined into the *Teacher*

Knowledge construct. Also, all Teacher Beliefs variables were positively and highly correlated with each other so I grouped them as the *Teacher Beliefs* Construct. Since teachers' willingness to implement the reform was measured by only one item, this variable was considered as a factor, *Teacher Willingness*.

Correlations among Factors

Table 10 presents correlations among all factors: three school context factors (School Community, Climate, School Resources), three teacher learning opportunities factors (Specific PD Hours, Collaboration, Involvement), and three teacher capacity & will factors (Teacher Knowledge, Beliefs, and Willingness).

Table 10

Correlations among Constructs / Factors

	1	2	3	4	5	6	7	8	9	10
1. Teacher Implementation	1									
2. School Community	.23**	1								
3. School Climate	.23**	.19**	1							
4. Resources	.20**	.14**	.23**	1						
5. Specific PD Hours	.18**	.07	.11**	.09*	1					
6. Collaboration	.21**	.13**	.22**	.05	.11**	1				
7. Involvement	.13**	.09*	.02	.08*	.10*	.19**	1			
8. Teacher Knowledge	.38**	.29**	.26**	.29**	.24**	.24**	.22**	1		
9. Teacher Beliefs	.23**	.06	.09*	.00	.08*	.04	.06	.10*	1	
10. Teacher Willingness	.31**	.42**	.13**	.16**	.00	.15**	.15**	.26**	.10**	1

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

Table 10 shows that there were interrelationships between school context factors and other factors including learning opportunities factors and teacher capacity & will factors. All contextual factors were correlated significantly with learning opportunities factors even if most correlation coefficients were small. Resource was significantly correlated with PD and Involvement. School Climate was significantly related to Collaboration. School Community was significantly correlated with Collaboration and Involvement. All contextual factors were significantly associated with Teacher Knowledge, Willingness to Implement, and Instructional Practices. Correlations between School Community and Willingness were high when compared to correlations between other contextual factors and Willingness.

All learning opportunity factors were correlated significantly with Teacher Knowledge. While Teacher Willingness was correlated with Collaboration and Involvement, it was not significantly correlated with PD. Among three learning opportunities factors, only Specific PD Hours was significantly correlated with Teacher Belief.

Based on the above examination of correlations between factors and teachers' implementation of the assessment reform, this study first hypothesizes that the factors directly influence teachers' implementation since the factors were significantly correlated with implementation. However, there appeared to be indirect relationships between the factors and teachers' implementation of the reform, since interrelationships among factors were found. Thus, this examination led me to think about testing both the direct and indirect influences of independent variables on teachers' implementation of the reform.

To test both the direct and indirect influences of the factors, I first conducted a

hierarchical regression analysis for two purposes: (1) to test the relative direct effects of the factors on teachers' implementation of the reform, and (2) to examine whether there are potential mediators on teachers' implementation of the reform primarily for an exploratory purpose. Then, I employed the Structural Equation Modeling (SEM) to confirm the hypothesized indirect effects of the factors on teachers' implementation of the reform. Since it was found that many factors correlated with each other, the dataset of this study is especially suited to Structural Equation Modeling (SEM), because SEM analysis could help the model sort all these relationships and put them into some theoretical order. That is, the SEM analysis allows me to examine the pathways in which the factors influence implementation.

The next sections describe results of both regression analysis and structural equation modeling conducted to examine factors influencing teachers' implementation of the reform. These different analyses provide a more comprehensive understanding about the relationships between the factors and teachers' implementation of the reform.

Hierarchical Regression Analysis

Before conducting hierarchical regression analysis, reliability analysis using Cronbach's alpha was employed to test internal consistency of the scales measuring constructs. Based on the literature review and the examination of correlation matrix, seven independent constructs and one outcome construct were created. These are School Community, School Climate, School Resources, Specific PD, Collaboration, Teacher Knowledge, and Teacher Belief. A composite reliability estimates for each construct was

listed in table 11.

Table 11

Results of reliability analysis

Group	Constructs (Factors)	N of Items	Mini.	Maxi.	Mean	SD	Alpha
School Contexts	School Community	4	2	6	4.48	.75	.72
	School Climate	5	1	6	3.70	.93	.79
	School Resources	2	1	6	2.70	1.15	.79
Learning Opportunities	Collaboration	2	1	5	2.27	.71	.69
	Specific PD Hours	2	1	5	1.55	.84	.92
Capacity & Will	Teacher belief	4	1	7	4.49	1.14	.59
	Teacher knowledge	5	1	5	2.94	.69	.89
	Implementation	14	1.32	4.93	2.78	.65	.84

After the variables were organized into constructs (factors), the hierarchical regression analysis was run for two purposes: (1) to examine relative direct effects of the three groups of factors on teachers' implementation of the reform (*Teacher Implementation*), and (2) to test whether there are potential mediators on Teacher Implementation. The three groups of factors are: (1) School Context factors, (2) Learning Opportunities factors, and (3) Teacher Capacity & Will factors. These groups of factors were entered in three blocks to examine the direct relationships of those factors to Teacher Implementation. Figure 8 presents the regression model.

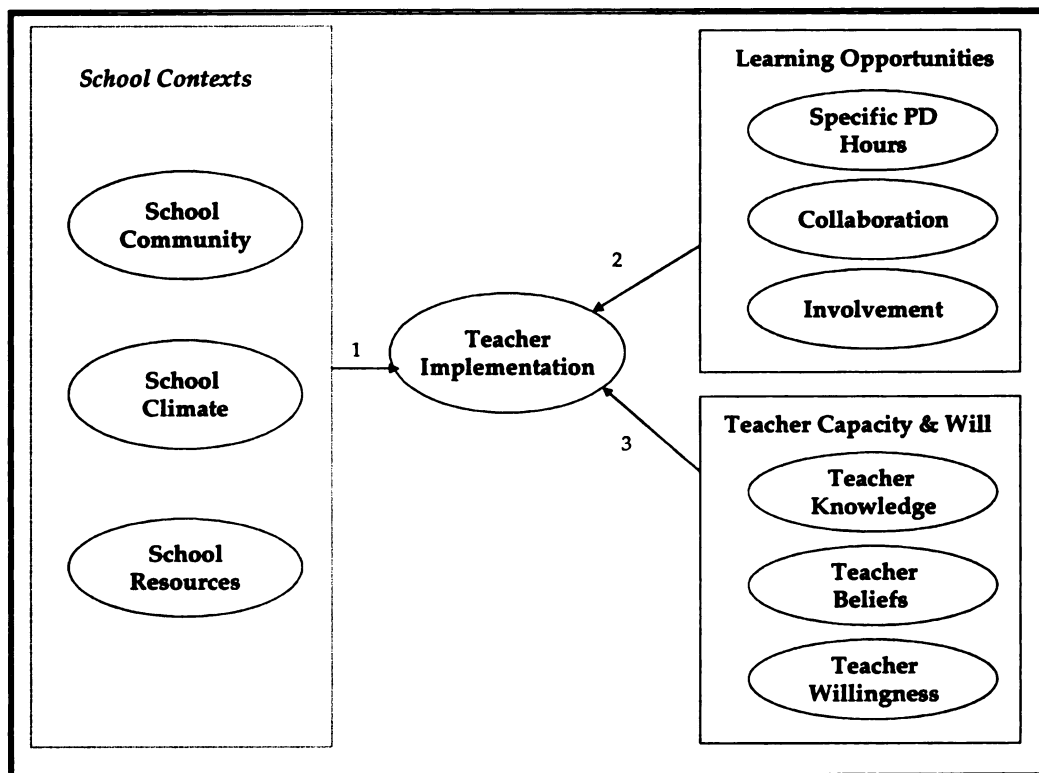


Figure 8. Regression Model

Model 1 was a baseline model, containing only factors related to the context in which teachers work (See Model 1 on Table 12). Each successive model adds one more group of factors, so that we can see how much each group contributes independently to Teacher Implementation. In the second step, learning opportunities factors were added to Model 1 (See Model 2 in Table 12). Then teacher capacity & will factors were added, to create Model 3, a full model that explains Teacher Implementation (See Model 3 in Table 12). Several teachers' demographic and classroom characteristics (gender, level of education, and average level of student's abilities) were entered into all models as control variables. Results of the hierarchical regression analysis are presented in Table 12.

Table 12

Results of the Hierarchical Regression Analysis

Variable	Model 1			Model 2			Model 3		
	School Contexts			Teacher Learning Opportunities			Teacher Capacity & Will		
	B(SE)	Beta	t	B(SE)	Beta	t	B(SE)	Beta	t
Intercept	1.66***(.20)		8.56	1.41***(.20)		7.09	.84***(.20)		4.11
<i>Demographic</i>									
Teacher Gender	-.04 (.07)	-.02	-.56	-.04(.07)	-.03	-.57	-.05(.06)	-.30	-.75
Has MA	.11(.06)	.07	1.80	.09(.06)	.06	1.44	.03(.06)	.02	.67
<i>School Contexts</i>									
Student Ability	-.19**(.05)	-.15	-3.50	-.18*** (.05)	-.15	-3.51	-.13** (.05)	-.10	-2.62
School Community	.14***(.04)	.17	4.00	.13***(.03)	.16	3.65	.04(.04)	.05	1.17
School Climate	.09** (.03)	.14	3.09	.07*(.03)	.10	2.33	.04(.03)	.05	1.68
School Resources	.08***(.02)	.16	3.66	.08*** (.02)	.16	3.60	.05*(.02)	.09	2.15
<i>Learning Opportunities</i>									
Specific PD Hours				.07*(.03)	.10	2.33	.05(.03)	.07	1.66
Collaboration				.11**(.04)	.13	2.93	.07(.03)	.08	1.90
Involvement				.03(.02)	.06	1.33	.01(.02)	.01	.34
<i>Teacher Capacity & Will</i>									
Teacher Knowledge							.15***(.03)	.20	4.52
Teacher Belief							.10***(.02)	.19	4.78
Teacher Willingness							.08***(.02)	.16	3.58
F statistics (df)			12.90(6,479)***			11.01(9,476)***			14.79(12,473)***
R ² (change in R ²)			.139 (.139***)			.172(.033***)			.273(.101***)

* p < .05; ** p < .01; *** p < .001

Effects of School Contexts Factors

Model 1 contained the school context factors: *School Climate*, *School Resources*, and *School Climate*. These three factors were entered as the first block of independent variables to test whether these contextual variables and some demographic variables had an impact on Teacher Implementation. The overall fit of Model 1 was statistically significant ($F(6, 470) = 12.90, p < .001$). The school context model explained 13.9% of variance in the dependent variable, Teacher Implementation.

The results of hierarchical regression showed that students' average ability, School Community, School Climate, and School Resources had direct and significant effects on Teacher Implementation, but that teachers' demographic variables, such as gender and level of education, did not have direct effects. When teachers perceived that the school community had a positive attitude toward the reform and the school climate encouraged teachers to try new ideas, and more resources were available, they reported higher implementation scores. Students' average ability level was associated significantly with Teacher Implementation. Teachers working with students of lower ability levels had difficulty implementing the assessment reform. Results of the analyses are presented in Model 1 in Table 12.

Effects of Teacher Learning Opportunities Factors

In the second step of the hierarchical regression analysis, learning opportunities factors were added to the school context model. Since I found earlier that hours teachers spent on professional development (PD) closely tied to the assessment reform were more highly correlated with Teacher Implementation, and since this finding is also supported by other research (Cohen & Hill, 2000; Garet et al., 2001), I decided to separate that type of PD from other types. I wanted to compare the effects of PD hours focusing on specific

reform-related content with the effects of PD hours focusing on general content, but I could not test this distinction since all the PD variables were highly correlated. Thus, instead of combining all PD variables, I selected two PD variables focusing on social studies assessment and social studies performance assessment and created one PD measure combining these two variables, since I wanted to examine the influences of PD focusing on specific reform-related content. The overall fit of Model 2 was found to be statistically significant ($F(9, 476)=11.01, P<.001$), and Model 2 explained 17.2% of variance in Teacher Implementation. The addition of learning opportunities factors, such as Specific PD hours, Collaboration, and Involvement, slightly improved the prediction of teachers' implementation ($R^2 \text{ change} = .03, p<.001$).

Collaboration with fellow teachers (Collaboration) and hours teachers spent on specific reform-related content (Specific PD Hours) had a significant effect on Teacher Implementation, but teachers' involvement in school activities (Involvement) did not have a significant effect when the contextual variables were controlled. The number of school activities in which teachers were involved was not shown to be important in improving teachers' implementation of the reform. Instead, the kinds of activities in which teachers participated or the nature of their participation may be more important. All school context factors that had significant direct effect on Teacher Implementation continued to show significant direct effects in this model, but the relative strength of these effects slightly decreased when learning opportunity variables were controlled. This finding indicates that there would be possibility of indirect effects of the contextual factors on Teacher Implementation since beta for all of the contextual factors dropped when learning opportunities factors were entered into Model 1 and these contextual factors were correlated significantly with the learning opportunities factors as examined in the first

section of this chapter. That is, the effects of contextual factors on Teacher implementation may be mediated by learning opportunities variables

Effects of Teacher Capacity & Will Factors

In the third step of the hierarchical regression, factors related to teacher capacity & will, such as teacher knowledge about performance assessment (Teacher Knowledge), Beliefs about teaching and learning (Teacher Beliefs), and willingness to implement the reform (Willingness) were entered into Model 2. The results of Model 3 are represented in Table 12. The overall fit of Model 3 was found to be statistically significant ($F(12, 473) = 14.79, P < .001$), and 27.3% of variance in Teacher Implementation was explained by Model 3. The addition of Teacher Capacity & Will factors improved the prediction of teachers' implementation (R^2 change = .10, $p < .001$).

Teacher Capacity & Will factors were found to be statistically significant when controlling context factors and learning opportunities factors. The results of Model 3 indicated that teachers with more knowledge of performance assessment, more reform-oriented beliefs about teaching and learning, or a stronger willingness to implement the reform had higher levels of implementation. Among the contextual factors that had significant effects on Model 2, *School Climate and School Community* were found to have insignificant effects on Teacher Implementation and the relative strength of these effects decreased when teacher capacity and will variables were entered. This result was the same with learning opportunities factors. All learning opportunities factors that had significant effects in the previous model showed insignificant effects in this, and the relative strength of these effects decreased when teacher capacity & will factors were controlled. Based on these results, I suspect that contextual and learning opportunities factors had indirect relationships with Teacher Implementation. That is, the effects of these variables on

Teacher Implementation might be mediated by teacher capacity & will factors since these factors no longer affected Teacher Implementation and were found to be significantly correlated with teacher capacity & will factors as examined in the first section of this chapter.

In sum, school context factors in Model 1 contributed significantly to the prediction of the implementation. In Model 2, the addition of factors related to learning opportunities slightly improved the prediction of the implementation. PD hours on specific reform-related content and collaboration with fellow teachers were significantly associated with teachers' implementation of the reform, but the effects of contextual factors on Teacher Implementation slightly decreased when learning opportunities factors were controlled. In Model 3, when teacher capacity & will variables were entered into Model 2, all contextual and learning opportunities factors except Resources became insignificant and beta for all contextual and learning opportunities decreased.

These findings prompted further analyses because previous research on educational reform, as evidenced in the literature review, has emphasized the importance of contextual and learning opportunities factors in changing teachers' practices. One possible explanation of why the full model does not show direct effects of the contextual variables and learning opportunities is that these variables may have indirect effects on teachers' implementation of the reform. For instance, collaboration with other colleagues may affect teachers' implementation indirectly by influencing teacher knowledge. That is, it can be hypothesized that contextual variables and learning opportunities have indirect effects on implementation through teacher capacity. Thus, to investigate both direct and indirect effects, structural equation modeling (SEM) analyses were conducted. These results are presented in the next section.

Structural Equation Modeling (SEM)

While the hierarchical regression analysis showed which variables influenced teachers' implementation of the mandated assessment reform, this analysis did not allow for examination of the interrelationships between independent variables. SEM analysis enabled me to find the indirect effects of factors that had no significant direct effects on teachers' implementation in the hierarchical regression analyses. The direct effects of the factors were also tested, conducting SEM analysis, to confirm the results of the regression analysis. Figure 9 presents the hypothesized model.

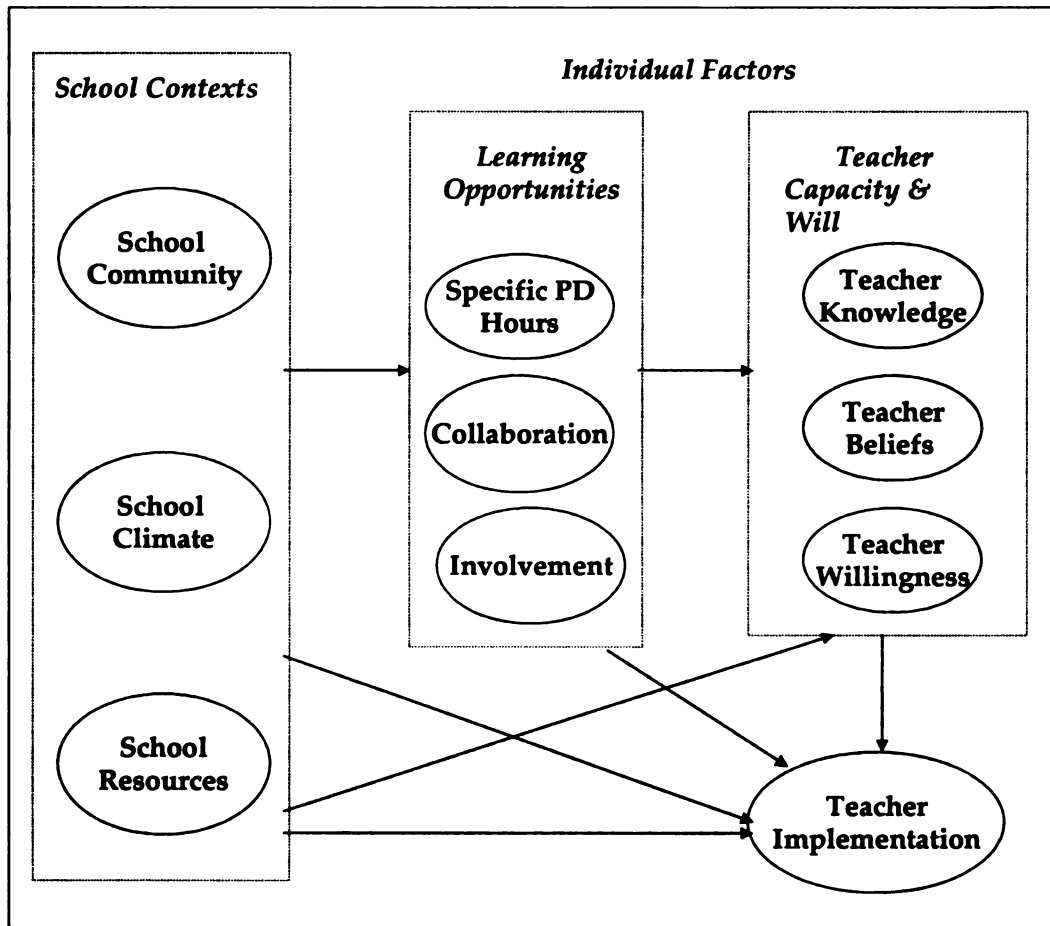


Figure 9. Hypothesized Model

This hypothesized model includes three groups of factors influencing teachers' implementation of the reform. The first group includes three contextual factors: the school community's attitude toward performance assessment (PA) (School Community), (2) the school climate for innovation (School Climate); and the available time and resources for implementing PA (School Resources). These factors are shown on the left side of Figure 9. The literature review establishes the importance of these School Context factors in teachers' implementation of the reform (Teacher Implementation), but no significant direct effects of the School Context factors were found in the earlier regression analyses. Indirect influences of these factors on Teacher Implementation were suspected because the School Context factors were found to be correlated with other factors, such as Learning Opportunities and Teacher Capacity & Will factors. Three hypotheses concerning both direct and indirect influences can be generated to explain indirect effects of the School Context factors on Teacher Implementation: (1) School Context factors indirectly influence Teacher Implementation through Learning Opportunities factors; (2) School Context factors indirectly influence implementation through Teacher Capacity & Will factors; and (3) School Context factors directly influence Teacher Implementation.

The second group consists of Learning Opportunities factors. These include: (1) hours teachers spent on PD specific for the reform (Specific PD Hours), (2) collaboration for implementing performance assessment (Collaboration), and (3) involvement in school activities related to the implementation of PA (Involvement). These factors are shown in the center of Figure 9. My earlier regression analysis could not find significant direct effects of these Learning Opportunities factors on implementation when the School Context and Teacher Capacity & Will factors were controlled. Based on the examination of correlations between Learning Opportunities and Teacher Capacity & Will factors, it

was hypothesized that Learning Opportunities factors indirectly would influence Teacher Implementation through Teacher Capacity & Will factors. Also, direct relationships between Learning Opportunities factors and Teacher Implementation were examined to see whether SEM analysis supports the results of the previous regression analysis.

The third group includes Teacher Capacity & Will factors: (1) teacher knowledge about performance assessment (Teacher Knowledge), (2) teacher belief about constructivist teaching (Teacher Beliefs), (3) and willingness to implement the reform (Teacher Willingness). These factors are shown on the right side of Figure 9. Based on the literature review and my previous regression analysis, only direct effects of teacher capacity & will factors on Teacher Implementation were hypothesized.

The hypothesized model in Figure 10 was specified in the form of a structural equation model. Figure 10 depicts the specified SEM model. In this figure, variables represented by rectangles are observed variables that were directly measured as part of the study, while the variables represented by ovals are unmeasured, or latent variables, which are hypothetical constructs (factors). For example, the School Community factor (SC), represented as a circle was measured by four observed variables, SC1 through SC4 represented by rectangles. This hypothesized model comprises both a measurement model and a structural model. The measurement model defines relations between observed variables and the underlying latent variables that the observed variables were designed to measure. The structural model defines relations among the latent variables. Comprising both a measurement model and a structural model, this SEM model is called a hybrid model (Kline, 1998).

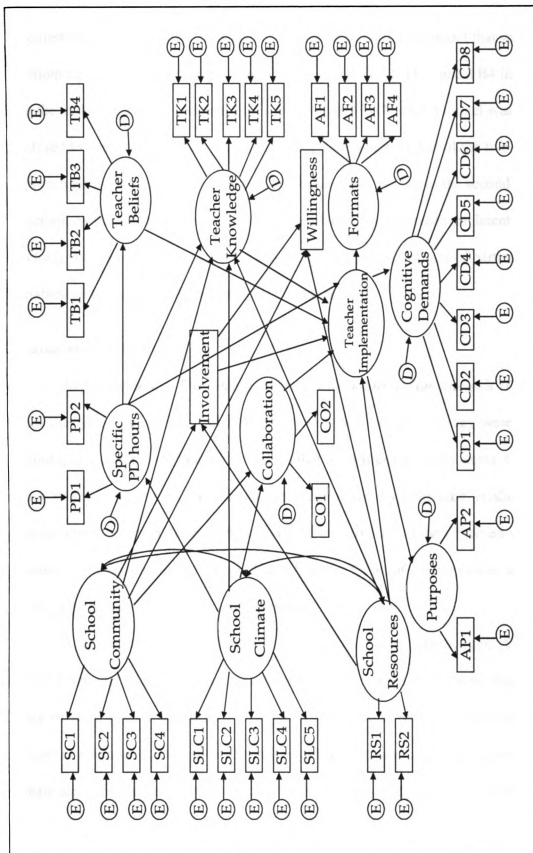


Figure 10. Specified SEM model

The specified hybrid model was analyzed using a two-step modeling approach. The first step was to create a confirmatory factor analysis (CFA) model that specified the relationships between the observed variables (i.e. TB1, TB2, TB3, and TB4 in Figure 10) and the underlying latent variables (i.e. TB in Figure 10). The CFA model was then analyzed to determine whether the model fit the data. The first part of the two-step approach aimed to find an acceptable CFA measurement model. In the second step, the structural model was analyzed to examine the relationships between the latent variables (factors). This structural model tested both direct and indirect effects of particular latent variables on teachers' implementation of the reform.

Measurement Model

Implementation: The second-order confirmatory factor analysis (CFA) model.

Since the latent variables are unobserved hypothetical constructs, they were represented by multiple indicators or observed variables that are scores on survey items. Confirmatory factor analysis (CFA) was conducted to confirm the hypothesized relations between selected observed variables and the underlying constructs that the observed variables are presumed to measure. Before testing the CFA model for all of the constructs, a CFA model for the implementation variable was examined.

A second-order factor model for implementation was hypothesized. A second-order CFA model is present when first-order factors are explained by some higher order factor structure (Schumacker, 2004). Based on the literature review, three dimensions—formats of assessment, cognitive demands of assessment, and purposes assessment—can explain teachers' implementation of the reform. The hypothesized second-order factor model is shown in Figure 11.

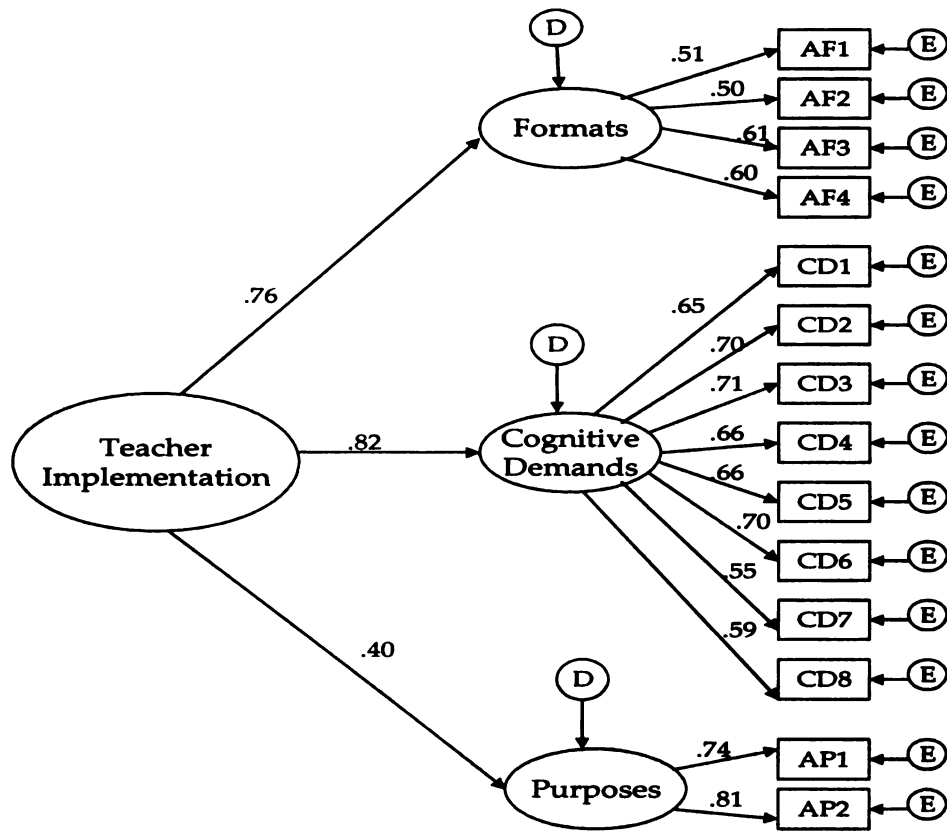


Figure 11. Hypothesized Second Order CFA for Teacher Implementation

The hypothesized second-order CFA was tested using AMOS 5.0. Several goodness-of-fit indices were used to see if the proposed second-order CFA model fitted the data. Table 13 presents the goodness-of-fit-indices for the hypothesized second-order CFA model. The χ^2 statistic was equal to 307.18 with 74 degrees of freedom and a p value less than .001. While the significant value of χ^2 indicated that the fit of the model was not acceptable, if the sample size is large, χ^2 statistics may be significant (Kline, 1998). To reduce the sensitivity of χ^2 statistic to sample size, the value of χ^2/df was examined as to whether the model fitted the data. The χ^2/df less than 3 indicates that the fit of the model is

acceptable. Since the value of χ^2/df was over 3, the hypothesized second-order model was not adequate.

Several fit indices, such as values of TLI, IFI, CFI, and RMSEA, were also examined to determine the adequacy of the model fit. Values of TLI, IFI, and CFI over .90 and values of RMSEA less than .05 indicate an acceptable fit. The indices in Table 13 provided mixed support for the initial second-order CFA model because the CFI and IFI values are greater than .90, but the values of χ^2/df and RMSEA were greater than the acceptable level. It thus was concluded that there was a problem with the model's fit.

To identify the problem, significant tests of parameter estimates first were examined. Inspection of the standardized parameter estimates shows that all indicators loaded significantly on their respective constructs. Standardized estimates showed that cognitive demands of assessment (.82) was indicated as the strongest measure of Teacher Implementation followed by formats of assessment(.76) and purposes of assessment (.40) with all three being statistically significant. The examination of parameter significant tests indicates that the indicators were good measures of the underlying latent variables.

Then, standardized residuals, which are differences between observed and model-implied correlations, were examined. A rule of thumb is that correlation residuals with absolute values greater than .10 are favorable (Kline, 1998). Four standardized residuals were greater than .10. Among these four, three were related to CD8, which is the variable "Inquiry skills" measuring cognitive demands, so CD8 was eliminated from the measurement model.

One more residual greater than .10 was found between CD5 (Application to a real world problem) and CD6 (Explanation of solutions). These variables were kept, but the covariance between error terms of these variables were added to the initial measurement

model because these two variables were important to measure dimensions of cognitive demands and these variables seemed correlated with each other. Teachers would ask students to explain how they solve a real world problem. Thus, I consider it appropriate to re-estimate the model with the error covariance between CD5 and CD6. The modified model is presented in Figure 12.

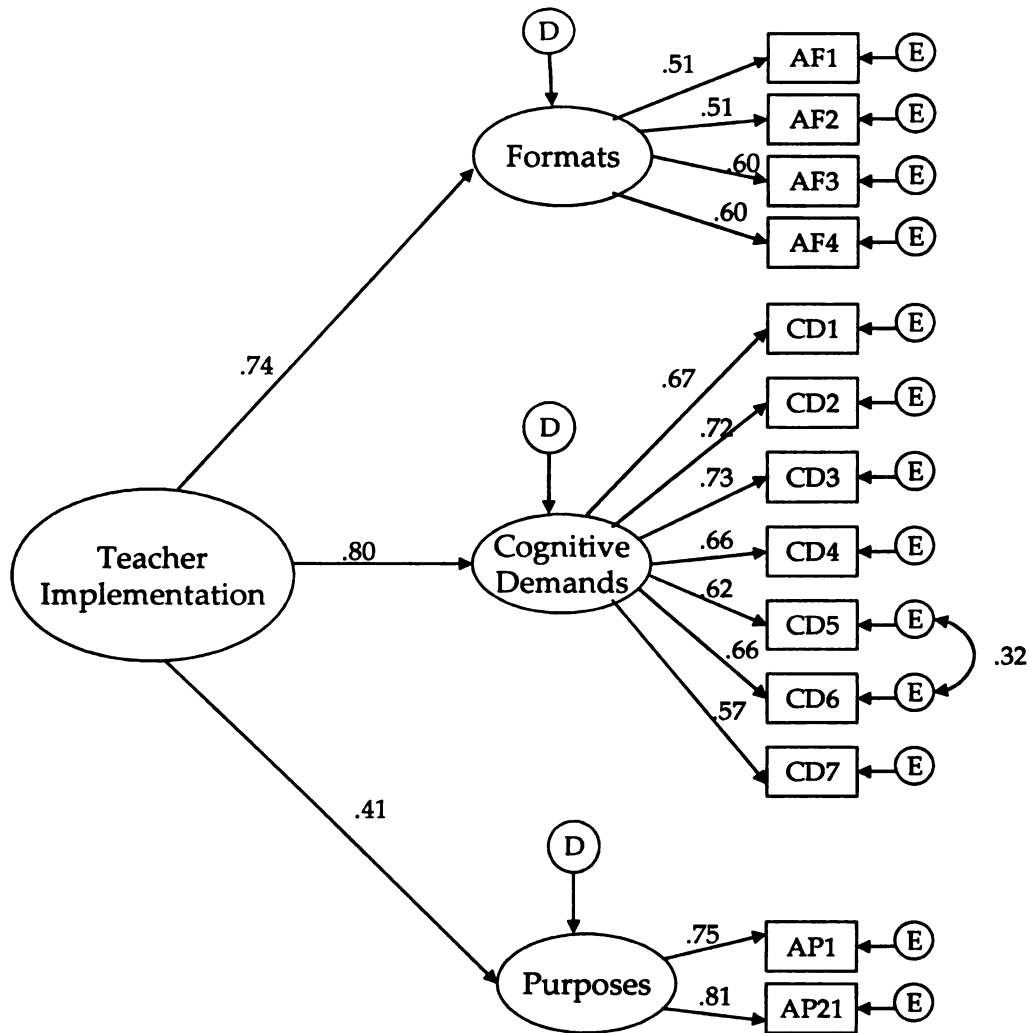


Figure 12. Modified Second Order CFA model for Teacher Implementation

Goodness-of-fit indices for the modified measurement model are also presented in Table 13. The value of TLI now exceeded .90, and values of χ^2/df and RMSEA were smaller than 3 and .05 respectively. The change in the overall χ^2 statistic was 163.10 and statistically significant ($p < .001$). This means that there was significant improvement by modifying the initial measurement model and the modified measurement model represents a substantively reasonable fit to the data. Thus, this modified model was considered to represent best the structure of Teacher Implementation.

Table 13

Comparison of the Initial and Modified Second Order CFA Models

	χ^2	df	χ^2/df	χ^2/df	df _{diff.}	TLI	IFI	CFI	RMSEA
Initial Model	307.18***	74	4.15	—	—	.88	.92	.92	.07
Modified Model	144.98***	61	2.36	163.10***	13	.95	.97	.97	.04

*** $P < .001$

The measurement model for all constructs. The measurement model was estimated using the maximum likelihood method to examine the hypothesized relations between observed variables and the underlying constructs. Inspection of the standardized parameter estimates indicated that all indicators loaded significantly on their respective constructs and correlations between constructs were not excessively high (less than .85 by Kline, 1998). The results of CFA analysis were presented in Tables 14 and 15.

Table 14

Parameter Estimates of the Measurement Model

	Paths	Standardized Estimates
Formats	← Teacher Implementation	.78***
Cognitive Demands	← Teacher Implementation	.68***
Purposes	← Teacher Implementation	.52***
AF1	← Formats of Assessment	.51***
AF2	← Formats of Assessment	.53***
AF3	← Formats of Assessment	.57***
AF4	← Formats of Assessment	.61***
CD1	← Cognitive Demands of Assessment	.75***
CD2	← Cognitive Demands of Assessment	.67***
CD3	← Cognitive Demands of Assessment	.72***
CD4	← Cognitive Demands of Assessment	.67***
CD5	← Cognitive Demands of Assessment	.62***
CD6	← Cognitive Demands of Assessment	.66***
CD7	← Cognitive Demands of Assessment	.57***
AP1	← Purposes of Assessment	.81***
AP2	← Purposes of Assessment	.75***
SC1	← School Community	.75***
SC2	← School Community	.63***
SC3	← School Community	.48***
SC4	← School Community	.66***
SCL1	← School Climate	.64***
SCL2	← School Climate	.62***
SCL3	← School Climate	.76***
SCL4	← School Climate	.61***
RS1	← School Resources	.73***
RS2	← School Resources	.90***
PD1	← Specific PD Hours	.86***
PD2	← Specific PD Hours	.97***
CO1	← Collaboration	.77***
CO2	← Collaboration	.68***
TK1	← Teacher Knowledge	.71***
TK2	← Teacher Knowledge	.88***
TK3	← Teacher Knowledge	.88***
TK4	← Teacher Knowledge	.83***
TK5	← Teacher Knowledge	.64***
TB1	← Teacher Beliefs	.51***
TB2	← Teacher Beliefs	.54***
TB3	← Teacher Beliefs	.58***
TB4	← Teacher Beliefs	.44***

*** p <.001

Table 15

Correlation Analysis among Latent Variables

Correlations	Estimates
School Community < ---- > School Climate	.28***
School Community < ---- > School Resource	.19***
School Community < ---- > Collaboration	.18**
School Community < ---- > Specific PD hours	.10*
School Community < ---- > Teacher Beliefs	.08
School Community < ---- > Teacher Knowledge	.34***
Teacher Implementation < ---- > School Community	.36***
Teacher Implementation < ---- > School Climate	.31***
Teacher Implementation < ---- > School Resources	.29***
Teacher Implementation < ---- > Collaboration	.31***
Teacher Implementation < ---- > Specific PD hours	.24***
Teacher Implementation < ---- > Teacher Knowledge	.43***
Teacher Implementation < ---- > Teacher Beliefs	.31*
School Climate < ---- > Collaboration	.36***
School Climate < ---- > Resources	.23***
School Climate < ---- > Teacher Knowledge	.33***
School Climate < ---- > Teacher Beliefs	.12*
School Climate < ---- > Specific PD hours	.12*
Collaboration < ---- > School Resources	.06
Collaboration < ---- > Teacher Knowledge	.29***
Collaboration < ---- > Teacher Beliefs	.05
Specific PD Hours < ---- > Resources	.10*
Specific PD Hours < ---- > Collaboration	.14**
Specific PD Hours < ---- > Teacher Knowledge	.27***
Specific PD Hours < ---- > Teacher Beliefs	.11*
Teacher Belief < ---- > School Resources	-.01
Teacher Belief < ---- > Teacher Knowledge	.09**
Teacher Knowledge < ---- > School Resources	.36***
AFN8 < ---- > AFN9	.32***

* p <.05; ** p <.01; *** p <.001

Several fit indices were used to see if the fit of the model was good. The χ^2 statistic was equal to 1691.14, with 862 degrees of freedom, and a p value less than .001. A smaller and insignificant χ^2 statistic indicates a better model fit, but this measurement was statistically significant. However, since χ^2 statistic can be sensitive to sample size, other fit indices, such as values of χ^2/df , CFI, IFI, and RMSEA, were utilized as alternative fit indices to determine the adequacy of model fit. The values of CFI, IFI, and TLI over .90, the values of RMSEA less than .05, and the value of χ^2/df less than 3 indicate an acceptable fit. The results of these fit indices were as follows: χ^2/df =1.96, TLI=.91, IFI=.92, CFI=.92, and RMSEA=.04. Thus, across this particular set of model fit indices, the measurement model represents a substantively reasonable fit to the data. This modified measurement model was accepted as the final measurement model for testing the hybrid model.

The Structural Model

Since the measurement model fit was satisfactory, the modified hybrid model that combined the final measurement and structural models was tested to examine relationships among latent variables (factors), using the maximum likelihood method. Figure 13 presents the modified hybrid model.

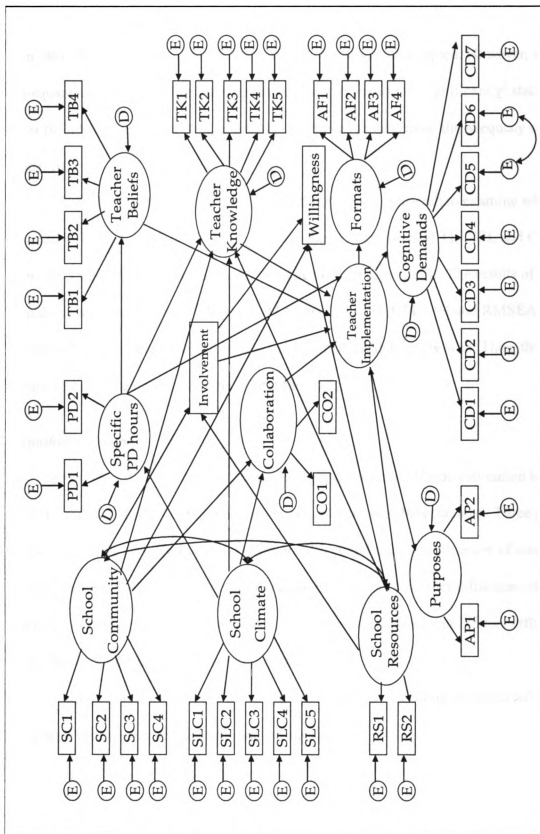


Figure 13. Modified Hybrid Model

Several goodness-of-fit indices were used to see if the hybrid model fitted the data. The χ^2 statistic was equal to 1204.98 with 661 degrees of freedom and a p value less than .001. When the χ^2 statistics is insignificant, the model is accepted. However, since chi-square statistics based on a large sample size may result in a significant χ^2 statistic, other fit indices were utilized as alternative fit indices to determine the adequacy of the model fit.

Values of χ^2/df , TL, IFI, CFI, and RMSEA were considered to examine whether the model fit the data. The values of χ^2/df less than 3, the values of TLI, IFI, and CFI over .90, values of RMSEA less than .05, indicate an acceptable fit. The results of these fit indices were as follows: $\chi^2/\text{df}=1.82$, TLI= .93, IFI= .94, CFI= .94 and RMSEA=.04. All fit indices supported that this hybrid model fitted the data fairly well. Thus, this model was accepted as the final model for the study.

Hypotheses Testing

The SEM analysis of the final model using maximum likelihood estimation tested both the direct and indirect relationships between latent variables (factors). Three groups of the hypotheses were proposed: (1) both the direct and indirect influences of contextual variables on Teacher Implementation, (2) both the direct and indirect influences of learning opportunities on implementation, and (3) both the direct and indirect influences of teacher capacity & will on implementation.

Table 16 presents the results of the SEM analysis, including standardized parameter estimates, critical ratios, and p values.

Table 16

Results of the SEM Analysis

Exogenous Variables ⁵	Endogenous Variables	Standardized Estimate	C.R.	P value
<i>School Contexts</i>				
School Community	Specific PD hours	.61	1.21	.23
	Collaboration	.11	1.90	.06
	Involvement	.10	2.00	.05
	Teacher Knowledge	.21	4.51	.00
	Teacher Beliefs	.07	1.07	.29
	Teacher Willingness	.46	9.76	.00
	Teacher Implementation	.11	1.21	.23
School Climate	Specific PD hours	.09	1.90	.06
	Collaboration	.34	5.63	.00
	Involvement	.00	.10	.92
	Teacher Knowledge	.15	3.08	.00
	Teacher Beliefs	.11	1.64	.10
	Teacher Willingness	.01	.13	.90
	Teacher Implementation	.07	1.12	.26
School Resources	Specific PD hours	.07	1.46	.14
	Collaboration	-.03	-.60	.55
	Involvement	.09	2.08	.04
	Teacher Knowledge	.25	5.55	.00
	Teacher Beliefs	-.06	-1.00	.32
	Teacher Willingness	.09	2.26	.02
	Teacher Implementation	.13	2.30	.02

⁵ Exogenous variables are synonymous with independent variables or predictors. These variables are causes of endogenous variables. Endogenous variables are dependent variables influenced by the exogenous variables (Byrne, 2000). Mediators are always endogenous variables.

<i>Teacher</i>				
<i>Learning Opportunities</i>				
Specific PD Hours	Teacher Knowledge	.18	4.75	.00
	Teacher Beliefs	.09	1.72	.08
	Teacher Willingness	-.06	-1.57	.12
	Teacher Implementation	.10	2.00	.05
Collaboration	Teacher Knowledge	.14	2.77	.00
	Teacher Beliefs	-.02	-.25	.80
	Teacher Willingness	.08	2.78	.00
	Teacher Implementation	.13	2.09	.04
Involvement	Teacher Knowledge	.11	2.93	.00
	Teacher Beliefs	.07	1.28	.20
	Teacher Willingness	.08	2.34	.02
	Teacher Implementation	.02	.49	.62
<i>Teacher</i>				
<i>Capacity & Will</i>				
Teacher Knowledge	Teacher Implementation	.20	3.28	.00
Teacher Beliefs	Teacher Implementation	.23	3.39	.00
Teacher Willingness	Teacher Implementation	.19	3.41	.00

A path diagram is also presented in Figure 14. For ease of presentation, the path diagram shows only significant paths with standardized parameter coefficients that represented the strength of the direct influences of an exogenous variable on an endogenous variable.

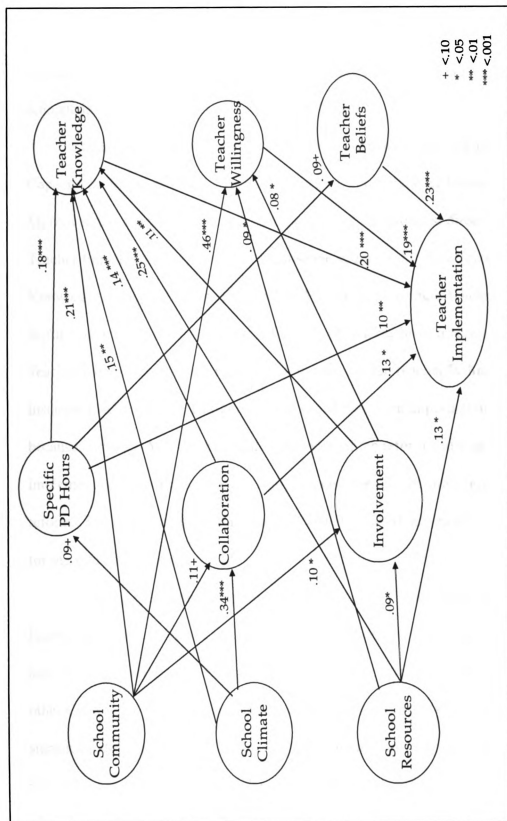


Figure 14. Results of Hypothesis Testing

Direct Effects on Teacher Implementations. Among the nine factors, six showed significant direct effects on teachers' implementation of the mandated assessment reform. These are School Resources, Collaboration, Specific PD Hours, Teacher Knowledge, Teacher Beliefs, and Teacher Willingness.

Unlike the School Contexts and Learning Opportunities factors, all the Teacher Capacity & Will factors had significant direct effects on Teacher Implementation. Moreover, the Teacher Capacity & Will factors had the strongest direct influences on Teacher Implementation. However, because the strength of these effects for Teacher Knowledge, Teacher Beliefs, and Teacher Willingness, no single Teacher Capacity & Will factor was most responsible for the influences of Teacher Implementation ($\beta = .20$ for Teacher Knowledge, .23 for Teacher Beliefs, and .19 for Teacher Willingness). For instance, teacher beliefs, willingness, and knowledge were important in the Ms. Lee case, because, although Ms. Lee believed in the value of the reform and was willing to implement, she did not do it successfully because she did not know how to implement the reform,. The results indicate that Teacher Capacity & Will factors are equally important for the success of the reform.

Among the learning opportunities factors, only Collaboration and Specific PD Hours had significant direct effects on implementation ($\beta=.10$ for Specific PD Hours, and .13 for Collaboration). However, the effect of Specific PD Hours was smaller than all other Contextual, Learning Opportunities, and Teacher Capacity & Will factors that had significant direct effects on Teacher Implementation. This finding contradicts previous research, as evidenced in the literature review, that professional development is a critical influence on changing teachers' practices. One possible explanation for this discrepancy is

that Specific PD Hours may have both direct and indirect effects on teachers' practices, and that this indirect effect may have had a significant impact on Teacher Implementation. For example, as evidenced in Figure 14, Specific PD Hours indirectly may have influenced Teacher Implementation through Teacher Knowledge. This hypothesis of Specific PD Hours' indirect effect on Teacher Implementation will be examined after my discussion of the direct effects.

Among the School Context factors, only School Resources was found to have a significant direct effect on Teacher Implementation. This finding was consistent with what four teachers in the case studies reported about barriers to implementation of the reform. Those four teachers complained that there was insufficient time for designing assessment tasks or evaluating students' reports. No significant direct effects of School Climate and School Community were found. However, several of eight teachers interviewed mentioned the importance of parents' attitudes toward performance assessment. For example, one teacher mentioned that parents wanted to know test scores or their child's class standing in relation to other students since they were familiar with traditional assessment. Thus, this example suggests that School Community does influence teachers' practices. Indeed, my analysis of survey data, as evidenced in Figure 14, shows that School Community and School Climate indirectly influenced Teacher Implementation.

Indirect effects on Teacher Implementation. Table 17 presents standardized direct, indirect, and total effects of all endogenous and exogenous variables in the final model.

Table 17

Standardized Estimates for Direct, Indirect, and Total Effects

Exogenous Variables	Endogenous Variables	Standardized Direct Effect	Standardized Indirect Effect	Standardized Total Effect
<i>Contextual variables</i>				
School Community	Specific PD Hours	.06	.	.06
	Collaboration	.11		.11
	Involvement	.10		.10
	Teacher knowledge	.21	.04	.25
	Teacher belief	.07	.01	.08
	Teacher Willingness	.46	.01	.47
	Implementation	.11	.18	.29
School Climate	Specific PD hours	.09		.09
	Collaboration	.34		.34
	Involvement	.01		.01
	Teacher knowledge	.15	.07	.22
	Teacher belief	.11		.11
	Teacher Willingness	.01	.02	.03
	Implementation	.07	.13	.20
School Resources	Specific PD hours	.07		.07
	Collaboration	-.03		-.03
	Involvement	.09		.09
	Teacher knowledge	.25	.02	.27
	Teacher belief	-.06	.01	-.05
	Teacher Willingness	.09		.09
	Implementation	.13	.07	.20

<i>Teacher Learning opportunities</i>				
Specific PD hours	Teacher knowledge	.18		.18
	Teacher belief	.09		.09
	Teacher Willingness	-.06		-.06
	Implementation	.10	.05	.15
Collaboration	Teacher knowledge	.14		.14
	Teacher belief	-.02		-.02
	Teacher Willingness	.08		.08
	Implementation	.13	.04	.17
Involvement	Teacher knowledge	.11		.11
	Teacher belief	.07		.07
	Teacher Willingness	.08		.08
	Implementation	.02	.05	.07
<i>Teacher Capacity & Will</i>				
Teacher knowledge	Implementation	.20		.20
Teacher Beliefs	Implementation	.23		.23
Teacher Willingness	Implementation	.19		.19

School Context factors were found to have insignificant direct effects on implementation, but the results of SEM analysis showed that contextual factors indirectly influenced implementation through two routes. One route was that contextual constructs indirectly influenced Teacher Implementation through Learning Opportunities factors. For instance, School Climate had strong direct effect on Collaboration ($\beta=.34$); Collaboration,

in turn, directly influenced Teacher Implementation ($\beta=.13$).

Another route was that School Context factors indirectly influenced Teacher Implementation through Teacher Capacity & Will factors. For instance, colleagues', students' and parents' positive attitude toward the reform (School Community) had strong, direct effect on teachers' openness (Teacher Willingness) to implement the reform ($\beta=.46$). Since this motivation to establish the reform was found to have a significant direct effect on teachers' implementation of the reform, School Community indirectly influenced Teacher Implementation through Teacher Willingness. When both direct and indirect effects of School Context factors on implementation were summed, the effects of School Context factors on implementation (total effects) became larger. Among the School Context constructs, school community had the strongest total effect on Teacher Implementation. The standardized total effects of School Community, Climate, and Resources were .29, .20, and .20, respectively.

However, the total effects on Teacher Implementation for each learning opportunities factor were smaller than the total effects of School Context and Teacher Capacity & Will factors. Additionally, the finding that Specific PD Hours had a smaller total effect on implementation than expected was surprising, since this finding contradicts previous research that professional development plays an important role in changing teachers' practices, and challenged my hypothesis that professional development makes this strong impact by supplementing its direct influence with indirect influence. One possible explanation for this small total effect of Specific PD Hours on implementation is that analysis of time spent on particular content areas did not account for the ways in which this content was learned. In the literature review I have argued that both quantity

and quality of professional development should be considered; in Chapter 6, I will expand my notion of PD quality to include the PD pedagogy, *authentic teacher learning*, or learning how to implement performance assessment through experiencing reform-oriented instructional practices. For example, when teachers, as learners, had activities in PD programs that were consistent with student activities emphasized in school reforms, they were more likely to change their practices. In the next chapter, the effects of PD on Teacher Implementation are reexamined, considering this aspect of PD programs.

Overall, 39.3 % of the variance in implementation was explained by the structural equation model, which tested both the direct and indirect effects of School Context, Learning Opportunities, and Teacher Capacity & Will factors on implementation, thereby explaining a larger proportion of variance in implementation than the regression model, which tested only the direct effects of these factors on teachers' implementation of performance assessment.

Three main findings were identified from the SEM analysis. First, all three groups of factors (School Context, Learning Opportunities, and Teacher Capacity & Will) were important for the success of the reform since these factors have substantial total effects on Teacher Implementation. This finding shows the complexity of implementing reform when considering that no single factor had a dominant influence on Teacher Implementation.

Second, the success of the reform was directly related to what teachers think and can do, since Teacher Capacity & Will factors had the strongest direct effects on teachers' implementation of the reform. The result shows that most School Context factors and Learning Opportunities factors indirectly influenced teachers' implementation of the

reform through Teacher Capacity & Will Factors. That is, the effects of these two factors were mediated by Teacher Capacity & Will factors. While the School Context and Learning Opportunities factors played important roles in Teacher Implementation, these factors' effects did not necessarily help teachers' implementation of the reform. When these factors promoted Teacher Capacity & Will factors, these factors could influence teachers' implementation of the reform.

Finally, Learning Opportunities factors appeared to play an important role in implementation since these factors had both direct effects on Teacher Capacity & Will factors, and mediating effects between the School Context and Teacher Capacity Factors. All the Learning Opportunities factors had substantial influences on teacher knowledge, and, among all factors, only Specific PD hours had a direct effect on Teacher Beliefs,. When considering the important role of Teacher Beliefs on Teacher Implementation, professional development appeared to be highly critical for the success of the reform, even if the strength of the effect was not large. In the next chapter, I will reexamine more the influences of Learning Opportunities factors on teachers' implementation of the reform, adding more learning opportunities that cannot be included in the model shown in this chapter.

CHAPTER 6

TEACHERS' LEARNING OPPORTUNITIES AND THEIR IMPLEMENTATION OF CURRICULUM-EMBEDDED PERFORMANCE ASSESSMENT

This chapter consists of two sections. The first section describes the relationships between the surveyed teachers' implementation of the mandated assessment reform (Teacher Implementation) and their experiences with professional development, collaboration, and involvement in school activities. The second section presents findings from the SEM analysis that reexamined both the direct and indirect effects of learning opportunities on Teacher Implementation, including other variables not tested in the previous analysis. For example, I could not analyze an important feature of PD, namely whether teachers in PD programs learned in the same ways in which the mandated reform encouraged them to teach their students. It seemed important to examine the role of active learning activities in teachers' professional development, because these activities have been shown to facilitate learning and model reform-oriented instructional practices. I could not include this measure because, to test the influence of this measure, I had to select only teachers who took PD programs. Thus, the new SEM model for this chapter included only teachers who had at least one PD program related to social studies.

Teachers' Learning Opportunities for Implementing the Reform

The previous chapter examined the influences of learning opportunities on Teacher Implementation, combining related variables into constructors (factors), but, in this

chapter, combined variables are separated to compare the influence of each variable on Teacher Implementation. For example, six professional development (PD) variables indicating hours spent on different PD content were included in the survey originally, because I wanted to examine the influences of hours spent on different PD content areas on Teacher Implementation. However, I could not differentiate influences of the six variables in the previous analyses because these variables are highly correlated and they led to multicollinearity in regression and SEM models. As a result, I formed a single measure combining reports of PD hours spent on social studies assessment and social studies performance assessment that excluded the other four PD variables discussed in Chapter Five in the previous analyses. This strategy did not permit examination to compare the independent effects of different PD content areas.

In separating the learning opportunities variables combined in Chapter Five, the first section of Chapter Six will determine which learning opportunities promoted teachers' implementation of the reform and will compare their relative influences. Thus, the first section of this chapter aims to compare influences of learning opportunities by separating variables combined in the previous analyses, in order to identify what kinds of learning opportunities would help teachers' implementation. In this section, I first examined the content areas teachers learned and the activities in which teachers engaged in PD programs, to identify the features of PD programs that most improved teachers' implementation of the reform. Then, I examined how teachers collaborated and identified the forms of collaboration that most encouraged teachers' implementation of the reform. Finally, I examined teachers' engagement in school activities and whether this involvement promoted teachers' implementation of the reform.

Professional Development

While many educators have pointed out that professional development (PD) of teachers is a critical influence in the implementation of reforms, the total effect of PD (Specific PD Hours in Chapter Four) was found to be smaller in the previous analysis when compared to other variables influencing implementation. In the previous analyses, I used a single measure of PD combining both hours on social studies assessment and hours on social studies performance assessment. While this measure combined both a structural aspect (total number of contact hours) and a core feature (Content of PD) identified from previous research (Cohen & Hill, 2000; Cohen & Hill, 2001; Desimone et al., 2002; Garet et al., 1999; Garet et al., 2001; Kennedy, 1998) it overlooked the important feature of whether teachers learned in PD programs in the same ways in which the reform mandated them to teach their students (Pedagogy of PD).

As reviewed in Chapter 2, recent research emphasizes that teachers need to engage actively in their learning in PD programs. This feature is important for teacher learning for the same reason active learning is important in student learning. In other words, because both content (what to learn) and pedagogy (how to learn) are important for students, both aspects should be considered for teacher learning. When teachers experience the same learning opportunities emphasized for student learning in educational reform proposals, teachers may be able to provide the same learning opportunities to their students. Thus, this feature of PD should be included in an analysis of the influences of PD on teachers' implementation of the mandated assessment reform.

Content of professional development (PD). As part of the survey, teachers were asked to report how many hours they spent on PD that addressed each of six different

content areas in the past three years: (1) new curriculum not focusing on social studies (general curriculum), (2) performance assessment not focusing on social studies (general performance assessment), (3) social studies curriculum, (4) social studies instruction, (5) social studies assessment, and (6) social studies performance assessment (PA). Table 18 shows that many teachers had few PD programs across the six content areas.

Table 18

Total PD Hours on Content Areas

Content of PD	Never N(%)	1- 5 N(%)	6-10 N(%)	11-20 N(%)	Over 20 N(%)
General curriculum	152(22.8)	156(23.4)	98(14.7)	67(10.0)	194(29.1)
General performance assessment	273(41.8)	224(34.3)	73(11.2)	29(4.4)	54(8.3)
Social studies curriculum	308(46.7)	229(34.7)	61(9.3)	21(3.2)	40(6.1)
Social studies assessment	404(62.2)	185(28.5)	32(4.9)	14(2.2)	15(2.3)
Social studies PA	404(62.1)	185(28.4)	32(4.9)	11(1.7)	19(2.9)/
Social studies instruction	378(57.8)	186(28.4)	41(6.3)	22(3.4)	27(4.1)

Percentages of teachers who attended no PD programs ranged from 22.8 to 62. 2.

Two largest percentages of teachers who did not attend PD programs (62.1%) were for sessions focusing on learning social studies assessment (62.2%) and social studies performance assessment (62.1%). While 22.8 % of teachers did not take any PD programs on general curriculum, more than 60 % of teachers reported they did not take any PD programs on social studies assessment and social studies performance assessment. Fewer than 10% of teachers reported they spent more than six hours learning social studies assessment and social studies performance assessment.

Figure 15 shows the relationship between total PD hours spent and Teacher Implementation across the six content areas. First, it shows that there was no difference in implementation between teachers who reported zero PD hours and teachers who reported less than five hours regardless of PD content. Second, the figure indicates that, when teachers attended more than 6 hours of PD, differences across PD content areas emerged. Three different patterns were found.

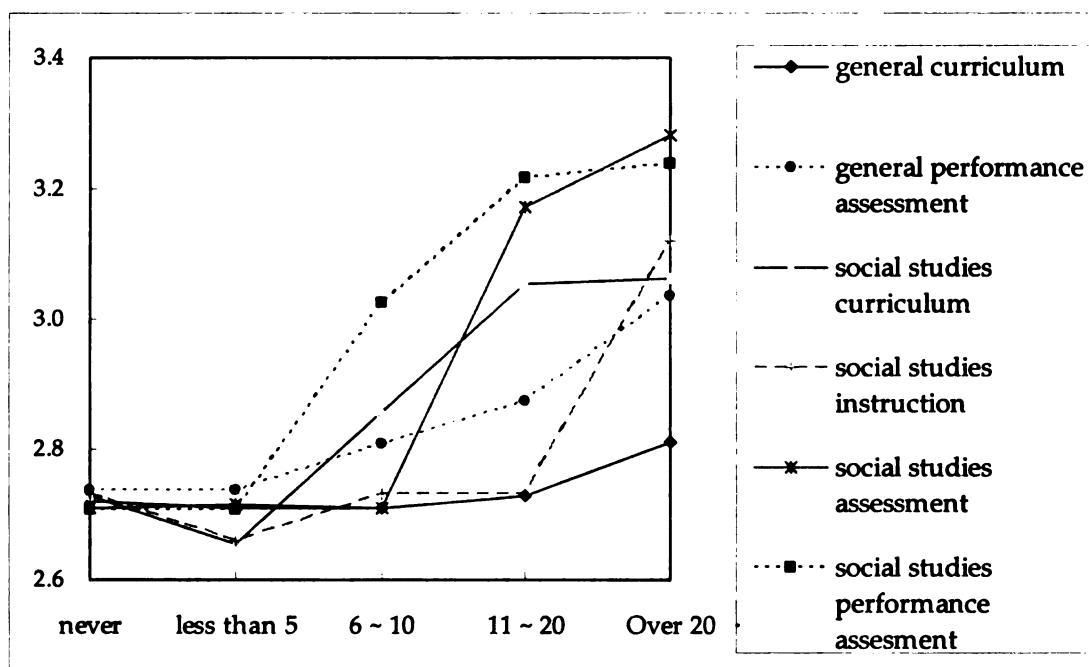


Figure 15. The Relationship Between PD Hours on Content Areas and Teacher Implementation

The first pattern is that time spent on a PD content area that did not focus both on social studies and assessment was less beneficial in improving teachers' implementation of the mandated assessment reform. As figure 15 illustrates, PD hours spent on general curriculum had little influence on Teacher Implementation. The second pattern is that hours invested in PD content areas that focused on either social studies or assessment showed some increases in Teacher Implementation. These content areas included social

studies curriculum, social studies instruction, and general performance assessment. While there were some differences in implementation when teachers attended 6-10 or 11-20 hours of PD, implementation scores were similar when teachers spent more than 20 hours on these three content areas.

The third pattern is that larger increases in implementation scores were found when teachers spent more hours on PD content areas focusing on both social studies and assessment. These content areas included social studies assessment and social studies performance assessment. When teachers spent more hours on these two content areas, they showed substantial increases in their implementation scores. While the influences on Teacher Implementation of social studies assessment PD hours strengthened when teachers took more than 11 hours, the influence of social studies performance assessment PD hours showed the greatest consistency when teachers invested more than six hours.

Figure 16 provides box plots of the influence on Teacher Implementation of the hours spent on the six PD content areas. The top of each box represents the 75th percentile, and the bottom of each box represents the 25th percentile. The horizontal line within each box represents the median, center of the distribution spread, that was used to mark the 50th percentile of teacher responses. The length of each box shows the spread of the distribution - "The longer the box, the greater the spread; the shorter the box, the smaller the spread." Vertical lines drawn outside the box represent more extreme cases. The numbers on the left of the box designate teachers' implementation scores, and the numbers under each box indicate the number of hours teachers spent on each PD content area. Ratings of the PD hours were coded on a 1-5 scale⁶.

⁶ 1 = Never; 2 = less than 3 hours; 3 = 3-5 hours; 4 = 6-10 hours; 5 = Over 11 hours

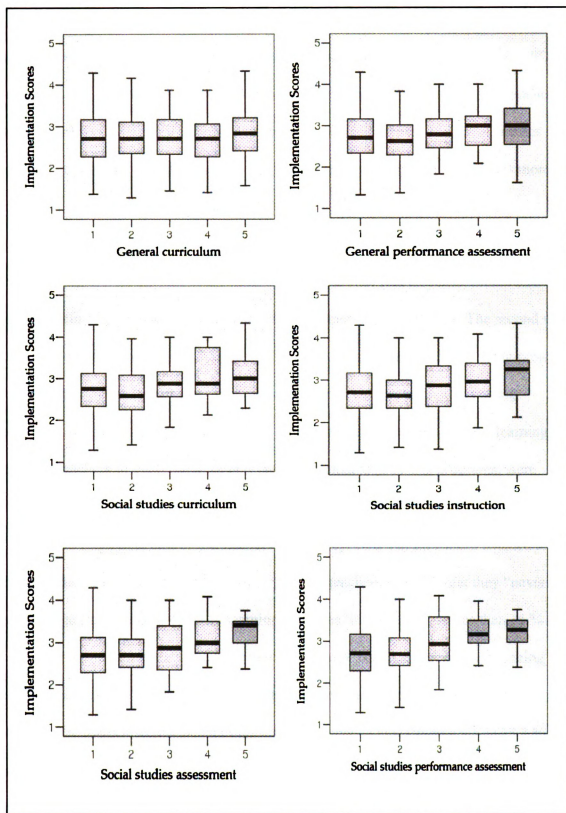


Figure 16. Box Plots of Implementation by PD Hours Spent on Specific Areas

The box plots in Figure 16 show the same patterns found in the previous graphs, but also allow us to see whether teachers differed much within each group. While more PD hours spent on general curriculum did not make a difference in implementation, teachers who reported spending more PD hours on social studies assessment and social studies performance assessment had higher median values, and their implementation was more consistent.

Pedagogy of professional development (PD). The learning activities provided by the PD programs were divided into two categories. The first was more traditional teacher learning, including listening to lectures and taking paper-and-pencil tests. The second was authentic teacher learning, including designing assessment tasks or rubrics, doing projects, making portfolios, undergoing performance assessment, and presenting or leading discussions. Table 19 shows the frequencies of teachers' engagements in these learning activities. Teachers in the original sample, who participated in no PD programs, were excluded from this examination of teachers' learning activities.

Whereas more than half of the teachers studied reported that they often listened to lectures in their PD programs, more than 65% of the teachers reported that they "never or seldom" engaged in reform-oriented learning activities. Interestingly, fewer than 15 % of the teachers reported that they "often or very often" took paper-and-pencil tests during their PD sessions, while more than 40 % of teachers did not take any paper-and-pencil tests. This finding is also true when examining performance assessment: While 48.2 % of teachers did not take any performance assessments during their PD, only 7.7 % reported that teachers often or very often took performance assessment. These results indicate that PD programs seldom assess teachers' learning using either type of assessment. Thus,

taking paper-and-pencil tests does not seem to be a traditional teacher learning activity any more than performance assessments are. In this sense, teachers' PD differs from traditional student learning activities in which students routinely are expected to undergo paper-and-pencil tests to assess their learning.

Table 19

Frequencies of Learning Activities

Pedagogy of PD		Never N (%)	Seldom N (%)	Sometimes N (%)	Often N (%)	Very Often N (%)
Traditional Teacher Learning	Lecture	6 (2.1)	61 (21.3)	68 (23.7)	79 (27.5)	73 (25.4)
	Paper-and-pencil multiple choice test	119(43.3)	55 (20.0)	62 (22.5)	31 (11.3)	8 (2.9)
Authentic Teacher Learning	Designing assessment tasks or rubrics	105 (38.3)	79 (28.8)	56 (20.4)	33 (12.0)	1 (0.4)
	Doing Project	181 (67.0)	37 (13.7)	38 (14.1)	14 (5.2)	0 (0.0)
	Making portfolio	136 (49.5)	75 (27.3)	41 (14.9)	23 (8.4)	0 (0.0)
	Undergoing performance assessment	132 48.2)	70 (25.5)	51 (18.6)	21 (7.7)	0 (0.0)
	Presenting or leading discussions	148 (53.8)	62 (22.5)	44 (16.0)	19 (6.9)	2 (0.7)

Figure 17 shows the relationship between types of PD learning activities teachers engaged in and their implementation of the mandated assessment reform.

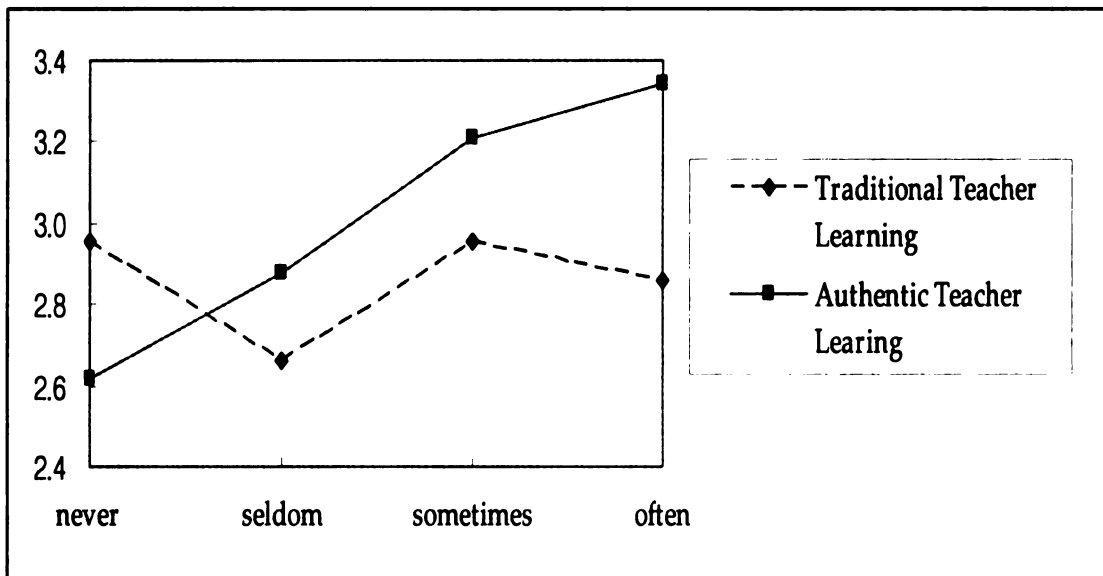
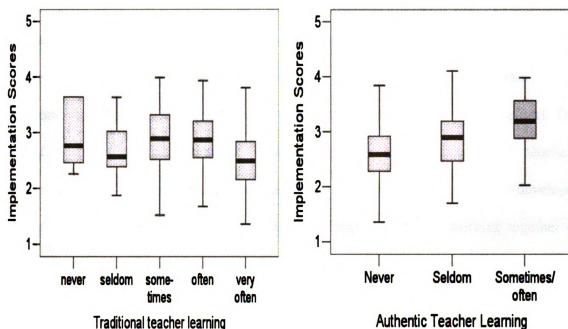


Figure 17. Teacher Learning Activities and Teacher Implementation

The scores on the left of the graph designate teachers' implementation scores. This graph indicates that teachers showed substantial improvement in their implementation of the reform when they engaged more often in reform-oriented learning activities. Engaging in the traditional learning activity of listening to lectures did not consistently improve teachers' implementation.

The box plots in Figure 18 also showed that whereas engaging more often in traditional learning activities did not improve teachers' implementation, teachers who reported that they engaged more often in reform-oriented learning activities had higher median values, and their implementation showed a smaller spread.



Note: When examining reform-oriented learning activities, no teachers reported experiencing these “very often.” Only three teachers responded with the answer, “often.” Therefore, the box plot of reform-oriented learning activities contains three boxes with “never,” “seldom,” and “sometimes/often” categories.

Figure 18. Box Plots of Influence of Frequency of Collaboration on Implementation

These examinations of graphs on both PD hours spent on specific content areas and learning activities, indicated that the content and pedagogy of professional development are important for teachers. To help teachers implement a reform, teachers should have sufficient professional development time to learn content closely tied to the reform, as well to experience learning activities that mirror those will be expected to provide for their students.

Collaboration

Teachers were asked to report how often they collaborated with colleagues. Additionally, two questions measuring different forms of collaboration were included in

the survey. One form of collaboration was discussion with colleague of how to implement the mandated assessment reform. For example, teachers shared problems or ideas that arose as they were trying to implement the mandated assessment reform. Another form of collaboration is actively working together with colleagues to implement the reform. This second form of collaboration is stronger because it better supports teachers' implementation of performance assessment. For example, teachers developed performance assessment tasks or rubrics with colleagues. Actively working together on collaborative tasks for implementation was more beneficial than merely sharing their ideas or problems. Table 20 shows the frequencies of different forms of collaboration

Table 20

Frequencies of Collaboration with Colleagues

	Never	Once or twice a semester	Once or twice a month	Once a week	Almost every day
	N (%)	N (%)	N (%)	N (%)	N (%)
Discussion on implementing performance Assessment (PA)	51 (7.5)	360 (53.1)	176 (26.0)	76 (11.2)	15 (2.2)
Developing PA tasks or rubrics	128 (18.8)	427 (62.6)	90 (13.2)	32 (4.7)	5 (.7)

Table 20 indicates that most teachers worked alone to implement the mandated assessment reform. More than 60 % of teachers reported that they never, or once or twice a semester, “discussed with colleagues” how to implement performance assessment. More than 18 % of teachers “never” collaborated and more than 60 % of teachers collaborated “once or twice a semester” to develop performance assessment tasks or rubrics with

colleagues. While teachers did not collaborate regularly (at least once a week) using either form of collaboration, the percentage of teachers (5.4 %) reporting that they had worked with colleagues on implementation at least “once a week” was smaller than the percentage of those (13.4%) who discussed with colleagues how to implement performance assessment.

Figure 19 shows a positive association between teachers’ collaboration and implementation. The numbers on the left of the graph designate teachers’ implementation scores. The graph indicates that when teachers collaborated more often with colleagues, teachers earned higher scores for implementation. When teachers collaborated with colleagues at least “once a week”, their implementation scores increased substantially. However, there were no differences in implementation scores between teachers who “never” collaborated and teachers who collaborated “once or twice a semester”.

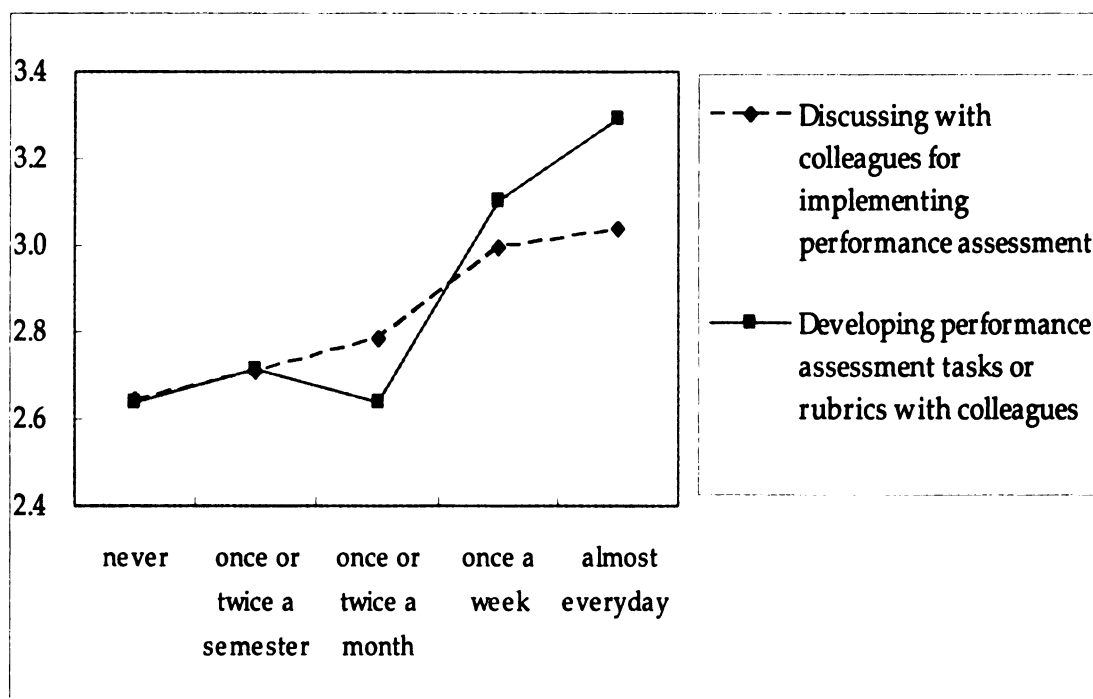


Figure 19. The Relationship between Collaboration and Teacher Implementation

Another finding presented in this graph is that collaborating with colleagues to develop performance assessment tasks or rubrics at least “once a week” showed a stronger influence on Teacher Implementation than merely discussing implementation of the mandated assessment reform with colleagues. This finding indicates that sharing ideas was an important first step in Chapter Six.

Involvement in School Activities

Teachers also were asked to report whether they were involved in any of five school activities; these included the development of: (1) a school curriculum framework, (2) a school assessment system, (3) a school performance assessment system, (4) performance assessment tasks, and (5) performance assessment rubrics. Table 21 shows the percentages of teacher involvement in each of these five.

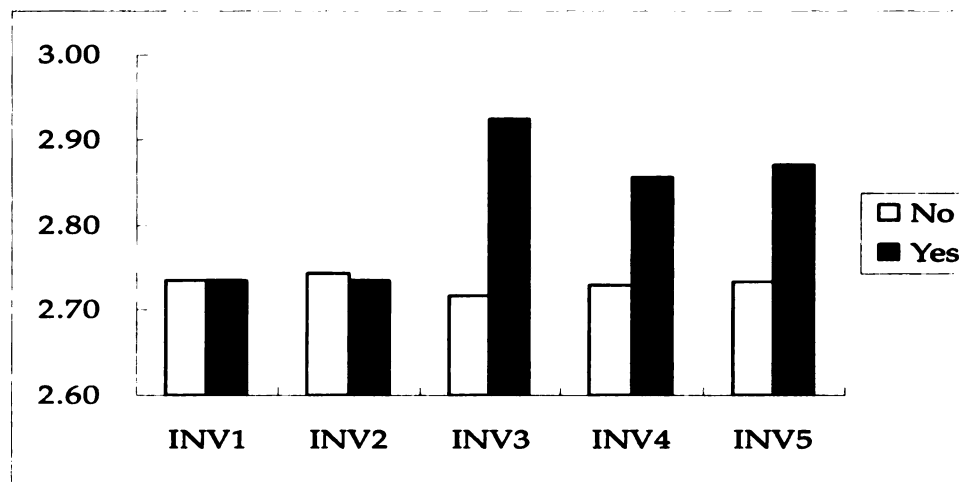
Table 21

Involvement in School Activities.

Involvement in developing:	No N (%)	Yes N (%)
School curriculum framework	579 (85.7)	97 (14.3)
School assessment system	630 (93.3)	45 (6.7)
School performance assessment system	565 (83.3)	113 (16.7)
Performance assessment tasks	545 (80.4)	133 (19.6)
Performance assessment rubrics	580 (85.5)	98 (14.5)

Table 21 shows that most teachers were not involved in the targeted school activities. The percentages of teachers who reported involvement ranged from 6.7 to 19.6. The scores on the left of the graph designate their implementation scores. The graph

presented below illustrates the influences on Teacher Implementation of involvement in the five school activities.



Note: INV1 = Curriculum Framework; INV2 = Assessment System; INV3 = Performance Assessment System; INV4 = Performance Assessment Tasks; INV5 = Performance Assessment Rubrics

Figure 20. Involvement in School Activities and Implementation

This graph indicates that teachers had higher implementation scores when they were involved in three school activities specifically related to performance assessment. Comparisons of the different school activities show that when teachers were involved in those school activities that were more closely related to performance assessment, they reported higher level of implementation.

The following box plot, Figure 21, examines whether involvement in the five school activities influences teachers' implementation of performance assessment. Teacher involvement was calculated by summing the number of school activities in which teachers were involved. For example, if a teacher reported that she or he was involved in all five activities, she or he earned a score of five. If a teacher reported that she or he was involved in none of the activities, she or he earned a score of zero.

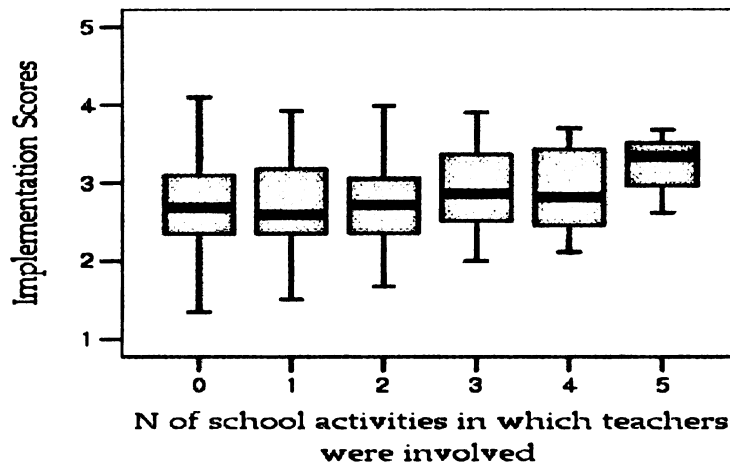


Figure 21. Box Plot of Implementation by Involvement in School Activities

This box plot indicates that teachers who reported being involved in more activities showed slightly higher mean values of implementation. This pattern is most apparent for teachers who reported involvement in all five school activities. When teachers were involved in all five school activities, they had the highest mean value and their implementation was less spread.

Summary of the First Section.

Several important findings for the success of the mandated assessment reform emerged from the graphs and box plots of learning opportunities. First, when the content of PD sessions was closely tied to performance assessment, teachers' implementation was better. Second, when PD programs provided teachers with learning activities that mirrored those they were expected to provide for their students, teachers' implementation improved. Third, when teachers regularly and actively worked together on collaborative tasks, their levels of implementation increased. Finally, when teachers were involved in school activities closely related to the mandated assessment reform, their implementation improved.

Structural Equation Modeling (SEM) Analysis

This section presents the results of SEM analyses that reexamined both the direct and indirect effects of Learning Opportunities on Teacher Implementation. One reason why I reanalyzed the influences of Learning Opportunities on implementation was to add a measure I could not include in the previous analysis. As found in the previous SEM analysis, the effect of hours teachers spent on professional development was small when compared to other factors influencing Teacher Implementation. One possible explanation for this small effect is that the PD variables failed to measure the most salient aspects of teachers' PD. That is, although the PD variables measured hours of PD closely tied to the assessment reform, PD time alone did not necessarily help teachers to implement the mandated assessment reform well. While my research has shown that the quantity of PD influences teachers' implementation of performance assessment, the quality of PD is also important in reforming teachers' practices.

To further define and assess the quality of PD, I hypothesized that quality of PD programs would use the same pedagogical methods as effective classrooms, and then, tested the influence of this factor on Teacher Implementation, using SEM. I could not include this factor in the previous analyses, because I had to select only teachers who took PD programs to test the influence of Classroom Consistent Pedagogy. Thus, the SEM analyses for this chapter included only teachers who had at least one PD program in curriculum, instruction, assessment, and performance assessment in social studies.

Another reason for reanalyzing the influences of Learning Opportunities on Teacher Implementation is to change the measure of Involvement. The first section of

this chapter examined the relationship between Involvement and Teachers' Implementation showing that teachers had higher implementation scores when they reported involvement in school activities closely related to the mandated assessment reform. The previous analyses assessed Involvement by combining different school activities in which teachers might have participated; however, the new involvement measure was created by combining only the three school activities that appeared to have substantial relationships with Teacher Implementation. The new, hypothesized teacher learning model is presented in Figure 22.

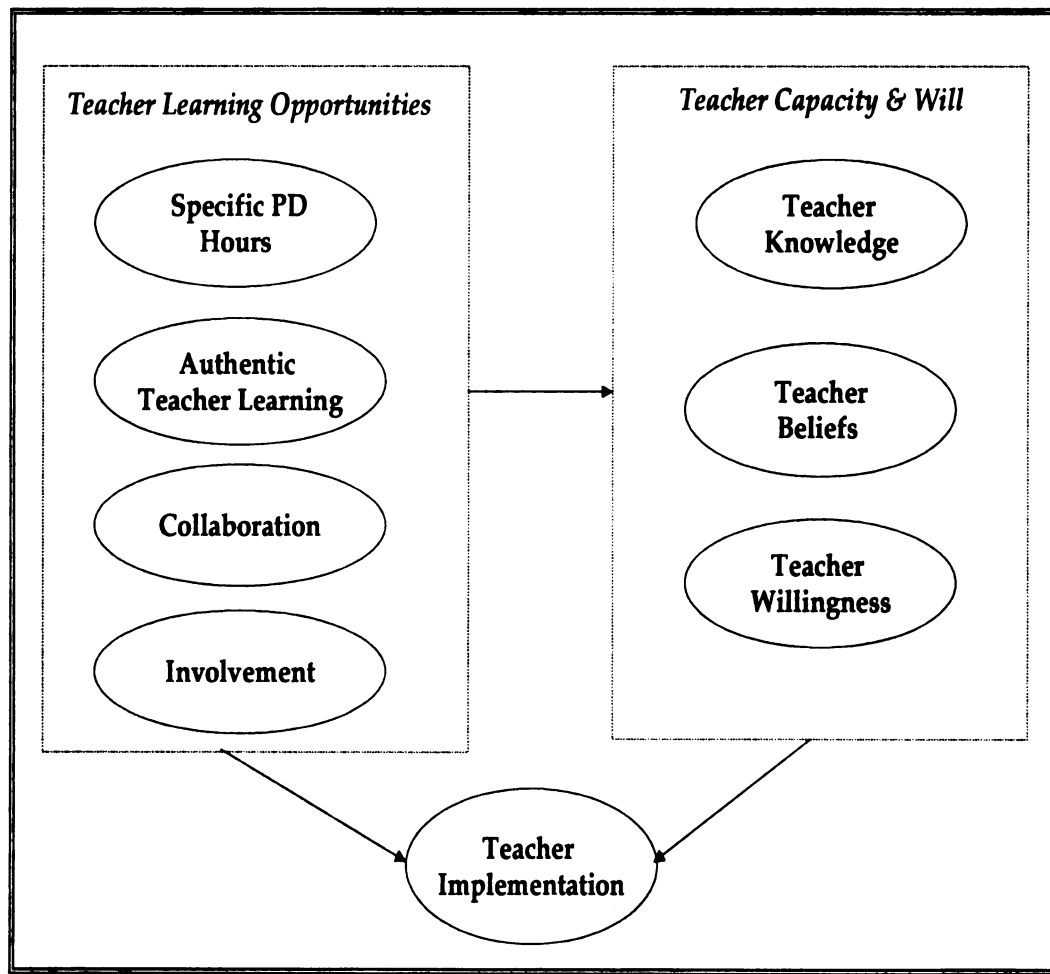


Figure 22. Hypothesized Teacher Learning Model

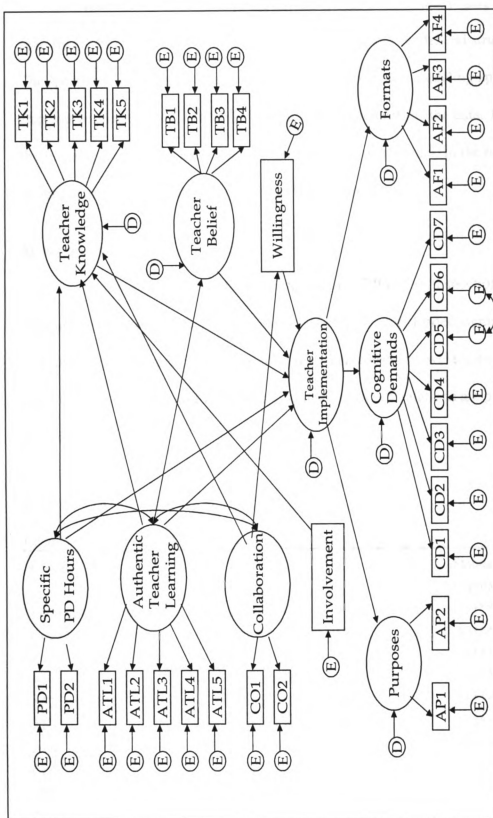


Figure 23. Specified Hybrid Model

The first step in SEM is to specify the model to be tested. Figure 23 depicts the specified hybrid model of SEM combining the measurement model and structural model. The specified hybrid model was analyzed using the two-step modeling approach. The first step was to analyze the CFA model to determine whether it fits the data. The purpose of the first step was to find an acceptable CFA measurement model. In the next step, the structural model was analyzed to examine the constructs.

Measurement Model

The proposed CFA model was tested using AMOS 5.0. Inspection of the standardized parameter estimates revealed that all indicators loaded significantly on their respective constructs, and that correlations between constructs were not excessively high (less than 0.5). The standardized factor loadings and the correlations are presented in Tables 22 and 23.

Table 22

Results of the Proposed CFA Model

Paths		Standardized Estimates
Formats of Assessment	← Implementation	.75***
Cognitive Demands	← Implementation	.67***
Purposes of Assessment	← Implementation	.55***
AF1	← Formats of Assessment	.52***
AF2	← Formats of Assessment	.56***
AF3	← Formats of Assessment	.58***
AF4	← Formats of Assessment	.62***
CD1	← Cognitive Demands	.76***
CD2	← Cognitive Demands	.67***
CD3	← Cognitive Demands	.72***
CD4	← Cognitive Demands	.68***

CD5	← Cognitive Demands	.68***
CD6	← Cognitive Demands	.65***
CD7	← Cognitive Demands	.61***
AP1	← Purposes of Assessment	.81***
AP2	← Purposes of Assessment	.79***
ATL1	← Authentic Teacher Learning	.72***
ATL2	← Authentic Teacher Learning	.78***
ATL3	← Authentic Teacher Learning	.77***
ATL4	← Authentic Teacher Learning	.56***
ATL5	← Authentic Teacher Learning	.77***
PD1	← Specific PD hours	.84***
PD2	← Specific PD hours	.93***
CO1	← Collaboration	.82***
CO2	← Collaboration	.62***
TK1	← Teacher Knowledge	.70***
TK2	← Teacher Knowledge	.89***
TK3	← Teacher Knowledge	.89***
TK4	← Teacher Knowledge	.82***
TK5	← Teacher Knowledge	.62***
TB1	← Teacher Beliefs	.50***
TB2	← Teacher Beliefs	.59***
TB3	← Teacher Beliefs	.54***
TB4	← Teacher Beliefs	.46***

*** p <.001

Table 23

Correlation Analysis of Latent Variables

Correlations	Estimates
Collaboration < ---- > Specific PD hours	.17*
Specific PD hours < ---- > Authentic Teacher Learning	.40***
Collaboration < ---- > Authentic Teacher Learning	.39***
AFN6< ---- > AFN7	.33***

* p <.05; ** p <.01; *** p <.001

Several fit indices were used to see if the fit of the model was acceptable. The χ^2 statistic was equal to 676.31, with 415 degrees of freedom, and a p value of less than .001. Although χ^2 statistic was significant, since χ^2 statistic is sensitive to sample size, other fit indices were utilized as alternative fit indices to verify the results of the chi-square test of model fit. χ^2/df , TLI, CFI, IFI, and RMSEA were considered to determine the adequacy of model fit. In general, values of TLI, CFI and IFI greater than .90, values of χ^2/df less than 3, and a value of RMSEA less than .05 indicate that the overall model fit is acceptable. Since the values of χ^2/df and RMSEA were smaller than the acceptable level (3 and .5, respectively), and the values TLI, IFI, and CFI were greater than the acceptable level (.90), I concluded that the model fit was acceptable. The results of the fit indices are presented in Table 23. Once the measurement model had been satisfied, the evaluation of the structural model could proceed.

Structural Model

Since the measurement model fit was satisfactory, the hybrid model combining the measurement and structural models was tested to examine the relationship among the constructs. The hybrid model was estimated using the maximum likelihood method. To determine whether the model fit the data, several fit indices were examined. The results of the fit indices are displayed in Table 24.

The χ^2 statistic was equal to 795.20, with 471 degrees of freedom, and a p value of less than .001. Although the χ^2 statistic was significant, since χ^2 statistic is sensitive to sample size, other fit indices were utilized as alternative fit indices to verify the results of the chi-square test of model fit. χ^2/df , CFI, IFI, and RMSEA were considered to

determine the adequacy of model fit. Results of fit indices were presented in Table 24. TLI, CFI and IFI were greater than the acceptable level (.9) and χ^2/df and RMSEA were smaller than the acceptable level (3 and .05, respectively). Across this particular set of model fit indices, the conclusion was that the proposed hybrid model fit the data fairly well.

Table 24

Results of Fit Indices in the Measurement Model and Structural Model

	χ^2	df	χ^2/df	TLI	IFI	CFI	RMSEA
Measurement model	676.31***	415	1.63	.93	.94	.94	.04
Structural model	795.20***	471	1.69	.91	.93	.93	.04

*** p <.001

Hypotheses Testing

The SEM analyses tested both the direct and indirect relationships between the constructs factors. Three groups of hypotheses were tested: (1) Learning Opportunities had direct effects on Teacher Implementation; (2) Learning Opportunities had direct effects on Teacher Capacity & Will; and (3) Teacher Capacity & Will factors had direct effects on Teacher Implementation. Table 25 presents the results of SEM analysis, including standardized parameter estimates, critical ratios, and p values.

Table 25

Results of the SEM Analysis

Exogenous Variables	Endogenous Variables	Standardized Estimate	C.R.	P value
<i>Learning opportunities</i>				
Specific PD hours	Teacher Knowledge	.12	1.93	.05
	Teacher Beliefs	-.00	-.01	.99
	Teacher Willingness	-.07	1.06	.29
	Teacher Implementation	.13	3.55	.09
Authentic Teacher Learning	Teacher Knowledge	.29	3.55	.00
	Teacher Beliefs	.23	2.15	.03
	Teacher Willingness	.09	1.11	.27
	Teacher Implementation	.31	2.87	.00
Collaboration	Teacher Knowledge	.23	2.94	.00
	Teacher Beliefs	.01	.12	.91
	Teacher Willingness	.18	2.41	.02
	Teacher Implementation	.04	.42	.68
Involvement	Teacher Knowledge	.08	1.74	.08
	Teacher Beliefs	.05	.69	.49
	Teacher Willingness	.06	1.14	.25
	Teacher Implementation	.05	.76	.45
<i>Teacher capacity & will</i>				
Teacher Knowledge	Teacher Implementation	.24	3.00	.00
Teacher Beliefs	Teacher Implementation	.19	2.18	.03
Teacher Willingness	Teacher Implementation	.24	3.74	.00

A path diagram is also presented in Figure 24. For ease of presentation, the path diagram shows only significant paths with standardized parameter estimates representing the strength of the direct effects of an exogenous variable on an endogenous variable.

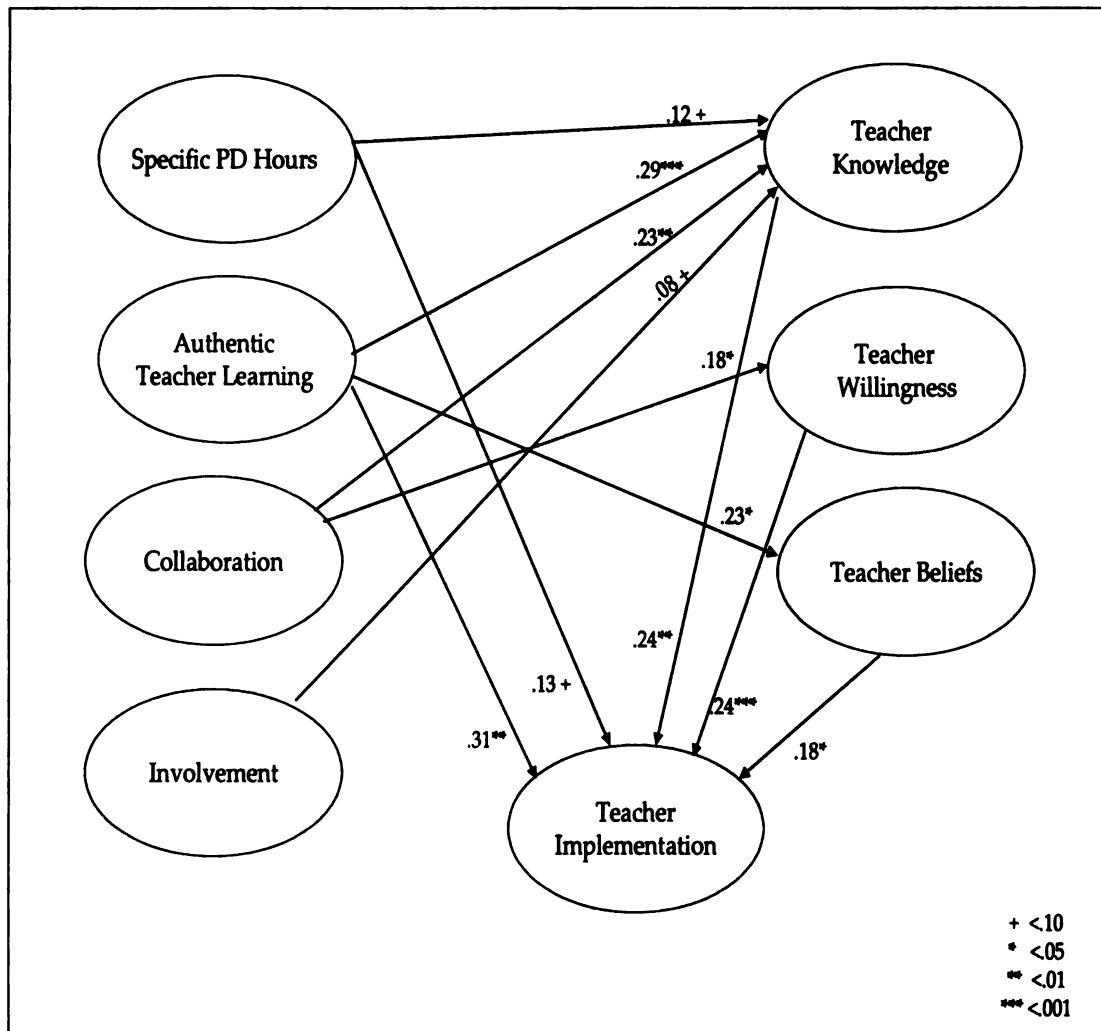


Figure 24. Results of Hypothesis Testing

Direct effects. Among the Learning Opportunities factors analyzed, only reform-oriented teacher learning activities (Authentic Teacher Learning) had a significant and strong direct effect on Teacher Implementation ($\beta = .31$). The direct effect of Authentic Teacher Learning was the strongest of all factors included in the hypothesized teacher

learning model. The standardized direct effects of PD hours spent on specific content was .13, but was significant only at .1 of the significant level. Two other Learning Opportunities factors, Collaboration and Involvement had very small direct effects and insignificant at .05 of significant level.

All Teacher Capacity & Will factors had substantial direct effects on implementation. Teacher Knowledge and Teacher Willingness had the same degree of direct effects ($\beta = .24$), and the effects of these factors were larger than the effect of Teacher Beliefs ($\beta = .18$). Teacher Capacity & Will factors had stronger direct effects on Teacher Implementation than Learning Opportunities factors except for Authentic Teacher Learning.

Learning Opportunities factors had significant direct effects on Teacher Capacity & Will factors, meaning that, through Teacher Knowledge, Teacher Willingness, and Teacher Beliefs, Learning Opportunities factors indirectly influenced Teacher Implementation. Authentic Teacher Learning and Collaboration had significant direct effects on Teacher knowledge ($\beta = .29$ for Authentic Teacher learning and $\beta = .23$ for Collaboration). The standardized effect of Specific PD Hours on Teacher Knowledge was .12, but the effect was significant only at .1 of the significant level.

Only collaboration had a significant direct effect on teachers' willingness to implement ($\beta = .18$). The previous SEM analyses in Chapter Five found that the School Community's positive attitude toward the reform had a direct effect on teachers' willingness to implement the reform ($\beta = .47$). Teachers' willingness to implement might be influenced more by contextual factors.

Among the Learning Opportunities factors, only Authentic Teacher Learning had a

significant direct effect on Teacher Beliefs ($\beta = .23$). This is an important finding since the SEM analysis in Chapter Five, which used a more comprehensive model, but did not include the Authentic Teacher Learning factor, found no factors that had a substantial influence on Teacher Beliefs. Thus, Authentic Teacher Learning was an important factor in the success of the reform since Teacher Beliefs was one of the critical factors directly influencing Teacher Implementation.

Indirect effects and total effects on implementation. The results of both the direct effects of several Learning Opportunities factors on Teacher Capacity & Will factors and the direct effects of Teacher Capacity & Will on Teacher Implementation indicated that Learning Opportunities factors influenced indirectly through Teacher Capacity & Will factors. Table 26 presents the standardized direct, indirect, and total effects of all endogenous and exogenous variables in the model.

Examination of both direct and indirect effects showed that Authentic Teacher Learning was the most critical factor influencing Teacher Implementation. The strength of the total effect on Teacher Implementation was the largest among all factors ($\beta = .44$). The total effect of reform-oriented learning activities was larger than the total effect of Teacher Knowledge, which had the strongest direct effect of the Teacher Capacity & Will factors. Authentic Teacher Learning indirectly influenced Teacher Implementation by promoting teacher knowledge about performance assessment and building reform-oriented teacher belief about teaching and learning.

Table 26

Standardized Estimates for Direct, Indirect, and Total Effects				
Exogenous Variables	Endogenous Variables	Standard- ized Direct effect	Standard- ized Indirect effect	Standard- ized Total effect
<i>Learning opportunities</i>				
Specific PD hours	Teacher Knowledge	.12		.12
	Teacher Beliefs	-.00		-.00
	Teacher Willingness	-.07		-.07
	Teacher Implementation	.13	.01	.14
Authentic Teacher				
Learning	Teacher Knowledge	.29		.29
	Teacher Beliefs	.23		.23
	Teacher Willingness	.09		.09
	Teacher Implementation	.31	.13	.44
Collaboration	Teacher Knowledge	.23		.23
	Teacher Beliefs	.01		.01
	Teacher Willingness	.18		.18
	Teacher Implementation	.04	.10	.14
Involvement	Teacher Knowledge	.08		.08
	Teacher Beliefs	.05		.05
	Teacher Willingness	.06		.06
	Teacher Implementation	.05	.04	.09
<i>Teacher capacity & will</i>				
Teacher Knowledge	Teacher Implementation	.24		.24
Teacher Beliefs	Teacher Implementation	.19		.19
Teacher Willingness	Teacher Implementation	.24		.24

The direct effect of Collaboration was small ($\beta = .04$), but when both direct and indirect effects of Collaboration on Teacher Implementation were summed, the effect of Collaboration became large ($\beta = .14$). Collaboration indirectly influenced Teacher Implementation by increasing teachers' willingness to implement and teachers' knowledge about performance assessment.

However, Specific PD hours still had a smaller total effect on Teacher Implementation ($\beta = .14$) than expected. The effect of Specific PD Hours was much smaller than reform-oriented learning activities (Authentic Teacher Learning). The small effect of specific PD Hours may be explained in several ways. One possible explanation is that the effect of Specific PD Hours depends on the quality of the PD programs. The strong effects of Authentic Teacher Learning indicates that, although teachers spent more hours of PD on content closely related to performance assessment, if teachers did not learn in the same ways in which the reform mandated them to teach their students, more PD hours did not improve teachers' implementation of the reform. This finding suggests that when PD programs are designed considering both content (what to learn) and pedagogy (how to learn) for teacher learning, teachers' implementation improves.

Involvement, measured by whether teachers participated in school activities related to performance assessment, was found to be the insignificant factor influencing implementation. Involvement had a small direct effect on Teacher Knowledge, but the standardized effect was less than .1 and significant only at .1 of the significant level. This finding would be consistent with the results about effects of quantity and quality of PD. While teachers' involvement in school activities related to the reform helped teachers to learn about performance assessment, involvement in school activities did not necessarily

improve teachers' implementation of the reform. The quality of the experiences teachers had during their involvement in school activities should be considered along with quantity of their involvement, to test the true effect of teachers' involvement in school activities on teachers' implementation of the reform.

Summary of the Second Section

From the SEM analysis, several important findings for success of the performance assessment reform emerged. First, success of the performance assessment reform was related directly to teacher capacity and will to implement the reform. This means teachers' implementation of the reform depended on what teachers think, and what teachers want and can do. Second, among learning opportunities factors, only authentic teacher learning was related directly to teachers' implementation. This means that when teachers actively engaged in their learning during professional development sessions, their levels of implementation increased. Third, influences of teachers' opportunities to learn about the reform were mediated by influences of teacher capacity and will to implement the reform. While all the learning opportunities factors appeared to influence teachers' knowledge about performance assessment, only authentic teacher learning had a direct and significant effect on teachers' beliefs about teaching and learning. Fourth, among all factors, including teacher capacity and will factor, authentic teacher learning has the strongest total effect on teachers' implementation of the performance assessment. The finding shows that the ways in which teachers learned were highly influential in promoting teacher capacity and will as well as their implementation.

CHAPTER 7

DISCUSSION, IMPLICATIONS, AND CONCLUSION

The purpose of this study was to examine how South Korean teachers responded to the performance assessment reform mandated by the Ministry of Education and to identify what factors influenced their implementation of the reform. Specifically, this study examined the influences of different features of teachers' learning opportunities on teachers' implementation of the reform, considering their learning opportunities as an important influence on changing teachers' practices.

This chapter comprises three sections. It begins with a discussion of significant findings, based on the research questions. Then, some implications of the findings from this study are presented. In the third and last section, I conclude this study by providing final thoughts about changing teachers' practices.

Discussion

This section provides a discussion of the findings regarding teachers' different responses to the assessment reform and influences on their levels of implementation of the reform. In interpreting the results, two methodological caveats, discussed in Chapter 3, should be considered: (1) my conclusions are based on self-report data and (2) my study does not draw causal inferences. However, for reasons laid out in Chapter 3, the study findings seem credible and plausible.

Teachers' Different Responses to the Performance Assessment Reform

The mandated performance assessment reform pushed teachers to start using new assessments in their classrooms. However, merely mandating reform did not cause teachers to fully implement that reform. Case studies show that all teachers believed that they were responding to the reform, yet their practices were quite different. Based on what teachers thought and what they did in response to the performance assessment reform, I identified four patterns in implementing the performance assessment reform.

Pattern 1, *Profound Implementation*, refers to implementation of the assessment reform at the deepest level: Teachers successfully implement the three aspects of performance assessment (formats, cognitive demands, and purposes) with reform-oriented views about instruction and assessment. Pattern 2, *Transitional Implementation* refers to implementation that mixes traditional and reform-oriented practices, moving toward Profound Implementation. Teachers' views about instruction and assessment seem to be close to reform ideals, but their assessment practices are not as closely aligned with the performance assessment reform. For example, teachers revise traditional paper-and-pencil tests to assess thinking skills, but do not assess higher-order thinking when employing performance assessment formats. Pattern 3, *Superficial Implementation*, refers to implementation of just the surface element of the reform (formats of assessment), even if teachers have views about instruction and assessment that seem close to reform ideals. Pattern 4, *Reluctant Implementation*, refers to implementation of just the surface elements of the reform, even though teachers express very traditional views about instruction and assessment.

The questionnaire data, based on a larger sample size (700 teachers) supported the

finding that teachers struggled to implement the performance assessment reform at a deep level. While many teachers could integrate surface elements of the reform (assessment formats) in their classrooms without difficulty, they were less successful in implementing the deep aspects of the reform (cognitive demands and purposes of assessment).

This study expanded upon previous research in that it attempted to assess the progress of the reform, considering both what teachers think or intend to do and what they actually do in response to assessment reform. If teachers believed the vision of the reform, this may mean that they had rationales or reasons for implementing the reform. While some research has examined teachers' views about instruction and assessment as an influence on their practices (Cronin Jones, 1991; Mintrop, 1999; Prawat, 1992; Wilson, 1990; Woodbury & Gess-Newsome, 2002), this study considered what teachers think about reform ideals both as an influence on teachers' practices and indicators of progress of the reform.

From my analyses of the case study and questionnaire data, three noticeable findings about policy and practice emerged. First, teachers responded to the reform in different ways. Second, while teachers easily could implement surface aspects of the reform, they struggled with implementing substantial aspects of the reform. These two findings are consistent with previous research (Olsen & Kirtman, 2002; Spillane & Jennings, 1997; Spillane & Zeuli, 1999), although this study examined a different context (South Korea), a different subject (social studies), and different reform practices (performance assessment).

The third important finding is that teachers' practices are influenced by their own knowledge and values. Examination of the case studies allowed me to think about the

importance of what teachers think about reform ideals as an indicator of their progress in implementing a reform. For example, even if two teachers, Ms. Lee and Ms. Park, implemented the reform at merely the surface level, we cannot say that these two teachers showed the same progress in implementing the reform. Whereas Ms. Lee showed more reform-oriented views about instruction and assessment, Ms. Park had traditional views about instruction and assessment. If we looked only at teachers' practices, we might say that Ms. Lee and Ms. Park were in the same stage of reform implementation. However, I argue that the teachers' progress in implementing the reform was not the same. Ms. Lee's views were aligned with the reform and she wanted to implement the reform, but she could not do so at the deeper level. Despite this inconsistency, Ms. Lee showed positive signs of implementation. Teachers like Ms. Lee may be able to implement substantial aspects of the reform if they receive appropriate support or have learning opportunities that help them attain competence for implementing the reform, when compared to teachers like Ms. Park. Since teachers like Ms. Park have no interest in implementing performance assessment reforms, it would be more difficult for them to change their assessment practices. Thus, this study emphasizes the importance of having an understanding of, and a commitment to, rationales for reform and this may be able to be an indicator to see the progress of the reform.

The examination on teachers' practices from both the cases and questionnaire allowed me to understand that teachers do not implement reform at the same speed, because they are not at the same starting point. While some teachers can implement both surface and deep aspects of the reform, some teachers cannot implement only the surface aspects of the reform. We should not celebrate the success of the reform only by looking

at the implementation of the surface aspects of a reform. Also, we need not grieve the failure of the reform because teachers cannot implement the reform at the deeper level. Teachers' different responses to reform should be understood as the developmental progress of the reform, and we need to attend to how we help teachers achieve further progress in their implementation of the reform.

Factors Influencing Teachers' Implementation of Performance Assessment

This study proposed a theoretical model that includes three groups of factors influencing teachers' implementation of the performance assessment reform. These three groups of factors are School Context factors, Teacher Learning Opportunities factors, and Teacher Capacity & Will factors. All factors mediated the influence of the mandated assessment policy (intended reform) on teachers' implementation of the reform (enacted reform). The proposed structural equation model (SEM) attempted to identify a process whereby these three groups of factors shaped teachers' implementation of the assessment reform. The SEM model shows possible pathways in which these factors influence teachers' implementation of the assessment reform. Since my data is cross-sectional, I cannot rule out alternative causal interpretations. Although I specified an ordering of the influences of these three factors, we should consider the possibility of two-way effects.

Influences of teacher capacity and will. My discussion on factors influencing teachers' implementation of the reform starts with teachers' capacity and will factors. These appeared to be the most critical factors shaping teachers' assessment practices. Teachers' knowledge about performance assessment, teachers' beliefs about instruction and assessment, and teachers' willingness to implement the reform directly influenced

teachers' implementation of the reform, while other groups of factors mostly had indirect influence on teachers' implementation of the reform. Without promoting or changing teachers' capacity and will to implement to reform, influences of other groups of factors may be ineffective. Thus, this study highlights that success of a reform is mediated by teachers' capacity and will to implement the reform. This pattern helps us make sense of the difference between Ms. Park and Ms. Lee.

Influences of teachers' learning opportunities. While teachers' capacity and willingness to implement the reform were critical factors, since these factors directly influenced teachers' implementation of the reform, teachers' opportunities to learn about performance assessment were important factors because teachers' learning opportunities played an important role in building teachers' capacity and will to implement the performance assessment reform. This study showed teachers' learning opportunities factors indirectly influenced their implementation of the reform through teachers' capacity and will. If we think that teacher capacity and will are highly important factors influencing teachers' implementation of reform, providing learning opportunities to teachers would be a good strategy to help teachers to better implement the reform. Four learning opportunities factors were examined.

Among the teacher learning opportunities factors, authentic teacher learning activities were the strongest influence on teachers' implementation of the reform. I argued that teachers may implement the reform more successfully when they experience learning activities similar to those they were expected to provide for students. The results of this study support this argument. This study showed that when teachers actively participated in their professional development programs (e.g., designing performance assessment tasks,

doing projects, and presentation/leading discussion), they were likely to learn more knowledge about performance assessment and change their beliefs about teaching and learning. Especially, authentic teacher learning was a unique influence on teachers' beliefs about teaching and learning.

This finding about the importance of authentic teacher learning is consistent with previous studies that examine the importance of teachers being actively engaged in their learning during professional development activities (Desimone et al., 2002; Garet et al., 2001; Ingvarson et al., 2005). Ingvarson et al. (2005) found consistent significant direct effects of active learning⁷ both on teacher knowledge and teaching practices across four professional development programs.

While the finding that there is a significant influence of authentic teacher learning on teacher knowledge and practices was consistent with the previous research, findings from my study extended the literature. First, my study found that authentic teacher learning was the only factor that directly influenced teacher belief. While previous studies found significant effects of active learning on teacher knowledge, they did not examine the importance of active learning on teacher beliefs about teaching and learning. When we consider that it is hard to change views, this study shows that authentic teacher learning is an important aspect of learning opportunities.

Second, this study found a greater effect of authentic teacher learning on teachers' assessment practices than did previous studies. One possible explanation for the larger effect would be that my study focused specifically on teachers' learning activities about

⁷ Since my questionnaire items were not the same with items used in previous study, I used a different name of the factor. However, the concept of authentic teacher learning was similar as active learning used in the previous studies.

assessment and their assessment practices, while the previous studies used more general questionnaire items. For example, Garet et al. (2001) used broader outcomes such as a composite score on teacher knowledge and skills in six areas: curriculum, assessment, instructional method.

Another possible explanation for the larger effect of authentic teacher learning is that my study examined social studies, while the previous research studied mathematics or did not focus on only one subject. However, there is nothing inherent in these content area differences that would lead one to expect different patterns of influence. The amount of professional development time spent in specific content designed for learning performance assessment also significantly influenced both teacher knowledge and teachers' implementation of the reform. Comparing the influences of time teachers spent on different content areas, I found that when teachers spent more hours specifically on performance assessment in social studies, they had higher implementation scores.

Professional development programs that did not focus on a specific subject or specific content about the reform did not have a substantial influence on teachers' implementation of the reform. However, the effects of this factor were smaller than the effects of authentic teacher learning activities. This result indicates that learning content is important, but if teachers do not learn the content while engaging in authentic teacher learning, their learning will not necessarily help them build teacher capacity or will to implement the reform.

No significant direct effect of specific content was consistent with the study by Ingvarson et al. (2005). While PD programs focusing on content had a significant direct effect on teacher knowledge, they did not show significant direct effects on practices in

two professional development programs and only a small direct effect in one professional development. However, another study conducted by Garet et al.(2001) found that professional development focusing on content had strong direct effect on teacher knowledge and practices, while active learning did not have significant effect on practices. I suspect that different studies examined different subjects. Garet et al. (2001) found the larger effect of professional development focusing on content investigated mathematics classroom practices. While mathematics teachers may struggle with understanding the content, social studies teachers may have more benefits when professional development programs provide more active learning activities to teachers because they may learn more about how to teach the content by experiencing the ways in which they are expected to provide for students.

Another learning opportunities factor that significantly influenced teachers' implementation was collaboration with fellow teachers. Such collaboration directly influenced teachers' implementation of reform as well as indirectly influenced implementation by enhancing their knowledge about performance assessment. I compared influences of different forms of collaboration: (1) discussing with fellow teachers about implementing the reform, and (2) working together on collaborative tasks for implementing the reform. This study found that when teachers regularly worked together to develop assessment tasks or assessment rubrics, their implementation improved.

This finding may be explained by Ball and Cohen (1999) who noted that situating professional discussion in concrete tasks is important because these tasks can ground teachers' conversations in intellectual work. Based on Ball and Cohen's study, I reason that teachers in my study may have improved their implementation while working

together to develop assessment tasks or rubrics by discussing what the goals of social studies instruction should be, what kinds of learning should be assessed, based on the goals of social studies, and how they should plan lessons to use the assessment tasks. I speculate that when teachers work together with specific tasks, their discussion is deep and rich.

Influences of school contexts in which teachers work. While teacher-level factors including teachers' learning opportunities and teacher capacity and will to implement the reform are crucial in implementing the reform, school contexts in which teachers work also need to be considered as factors indirectly influencing teachers' implementation through either teachers' learning opportunities or teacher capacity and will to implement the reform. Several significant findings related to influences of school contexts emerged.

One significant finding is that school community's positive attitude toward the reform had the strongest effect on teachers' willingness to implement. Among school community members, parents' attitude toward the reform appeared the most influential on teachers' willingness to implement. This finding makes sense because influences of parents on education in South Korea are extremely powerful.

Another significant finding is about the importance of school climate that encouraged teachers to try new ideas. School climate exerted a strong influence on teachers' collaboration with fellow teachers. This finding is consistent with my own experience as a teacher. One of the problems with implementing reform ideas in South Korea is that school culture has not encouraged teachers to bring new ideas to their classrooms. Since more than 90% of South Korean teachers work in public schools as

government officials, their jobs are secure. Under this situation, teachers may want to stay inside their classrooms without changes.

School Resources (time and available resources for implementing the reform) are the only school context factor that directly influenced teachers' implementation of the assessment reform. All teachers in the case studies complained it was hard to find time for evaluating performance assessments. Since teachers had many students, performance assessment tasks that required extended responses could be highly time-consuming. With limited time, teachers may prefer using multiple-choice tests or tasks requiring short answers, because evaluating these formats of assessment is easy and quick.

Previous literature suggests that contextual factors alone make a significant difference in changing teachers' practices, and their influences are mediated by knowledge and skills of teachers (Elmore, 1995; Gess-Newsome et al., 2003). My study supports the previous studies, showing that contextual factors indirectly influence teachers' implementation of the reform via teacher learning opportunities or teacher capacity and will to implement the reform. I agree that changes in school contexts do not guarantee changes in teachers' practices, but this does not mean that the government or school administrators need not pay attention to the contextual barriers to implementing the reform. They should remove these barriers in order to help teachers better implement the reform; removal of contextual barriers cannot be a sufficient condition, but should be considered as necessary condition for changing teachers' practices (Gess-Newsome et al., 2003).

Implications

The findings from this study allow us to ask what teachers need to bring about fundamental changes in their assessment practices from their current point. How can we get teachers' practices to go one step further? In Figure 25, I try to diagram the progress of implementing the reform.

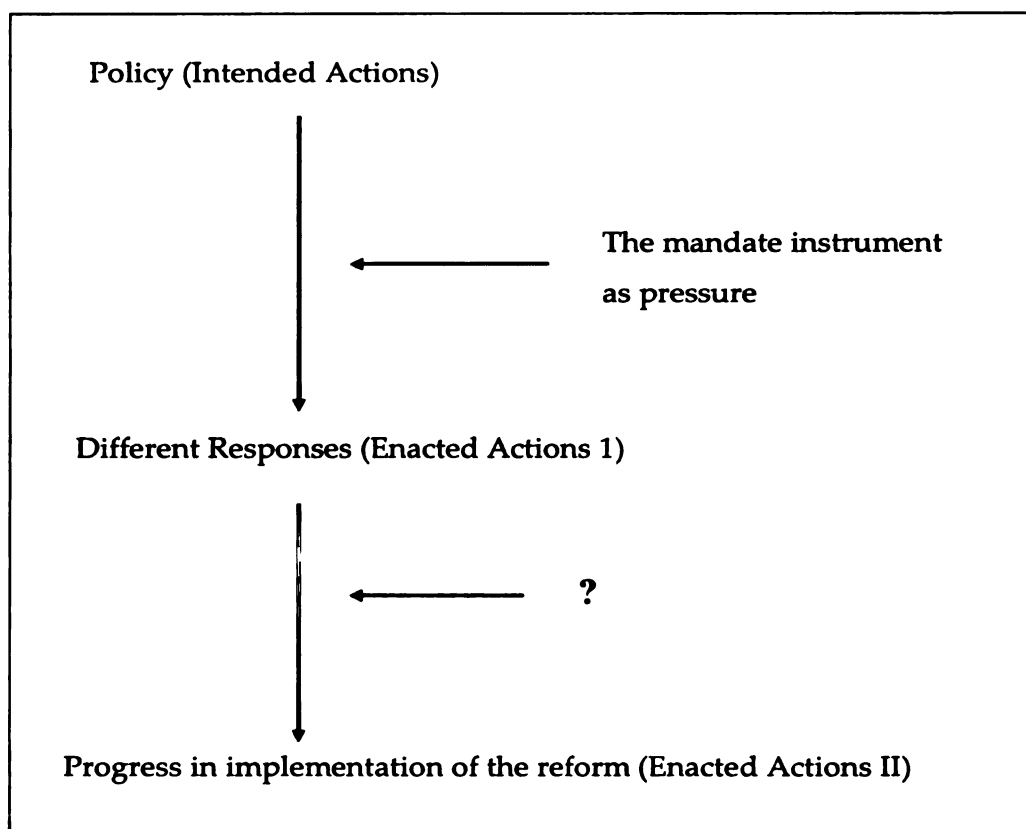


Figure 25. Progress of Implementing the Reform

Figure 25 suggests that the impetus for change begins at the top with a new policy that defines a set of intended actions. Employing the mandate as a policy instrument, the government brought rapid changes to teachers' assessment practices across the nation even if these changes were at the surface level. All teachers made some sort of change in assessment practice. That is, the top-down strategy (e.g., mandate) does not have only

disadvantages. The top-down strategy for reforming schools seemed to have a strong power to bring attention to new assessments and to initiate changes on assessment practices.

However, the problem of using only the top-down strategy is that it cannot bring about deep changes in practices that have been emphasized by reforms at the local level (Firestone, Monfils, & Camilli, 2001; Fullan, 1994; McLaughlin, 1987). The reasons many reform efforts have been unsuccessful in South Korea is that the government has relied heavily on this top-down strategy. Teachers, as evidenced in the case studies, believed they were responding to the reform, since they have been ordered, but their responses to the assessment reform differed. That is, even if the mandate can cause teachers to initiate the reform in their classrooms, teachers implemented the reform in different ways, directly depending on their capacity and will to implement the reform.

Merely mandating the reform cannot increase or change teachers' knowledge about the reform, their beliefs about instruction and assessment, or their willingness to implement, as McLaughlin (1987) noted that policy "cannot mandate what matters" (p.173). If this top-down strategy did not work, should we now consider the bottom-up strategy for success of the reform? Instead of giving up the mandate instrument, my suggestion is to consider another instrument to supplement the top-down strategy. Fullan (1994) found that both top-down and bottom-strategy were weak in changing practices, reviewing studies that used different reform strategies. Thus, it would be better to find ways in which both strategies can be balanced.

What we are missing (*the question mark in Figure 25*) is a better understanding of how to change teachers' implementation, from superficial compliance to substantial and

deep change in practice. I will call it “scaffolding instruments” that operate as supports for teacher learning and change in practice. Even if the reform operates as pressure for changing practices, merely mandating a reform will not guarantee the kind of changes in practices ultimately desired. Firestone et al. (2001) suggest that when pressure is well harmonized with support, teachers will adopt reforms successfully. Thus, while the assessment reform mandated by the South Korean Ministry of Education made it possible for teachers to move from their previous assessment practices to somewhere (e.g., reluctant, profound implementation), teachers need scaffoldings to move the next step toward profound implementation. In Figure 26, black arrows indicate the effect of the mandate and dotted arrows indicate the imaginary effects of scaffoldings if teachers receive appropriate scaffoldings based on their current points. The diagnosis on how teachers implemented the reform is a way to decide the right kinds of scaffoldings for helping teachers to make a progress in their implementation of the reform.

As I step back from my study, and as I look ahead toward future reform efforts, it seems reasonable to ask what kind of scaffoldings we need to provide for teachers’ more fundamental changes in their practices. For example, how might we help Ms. Park change her beliefs about assessment practices? How might we help Ms. Lee implement reform at the deeper level? This study suggests that the quality of learning opportunities is the best scaffolding instrument for changing teachers’ practices by promoting teacher capacity and will

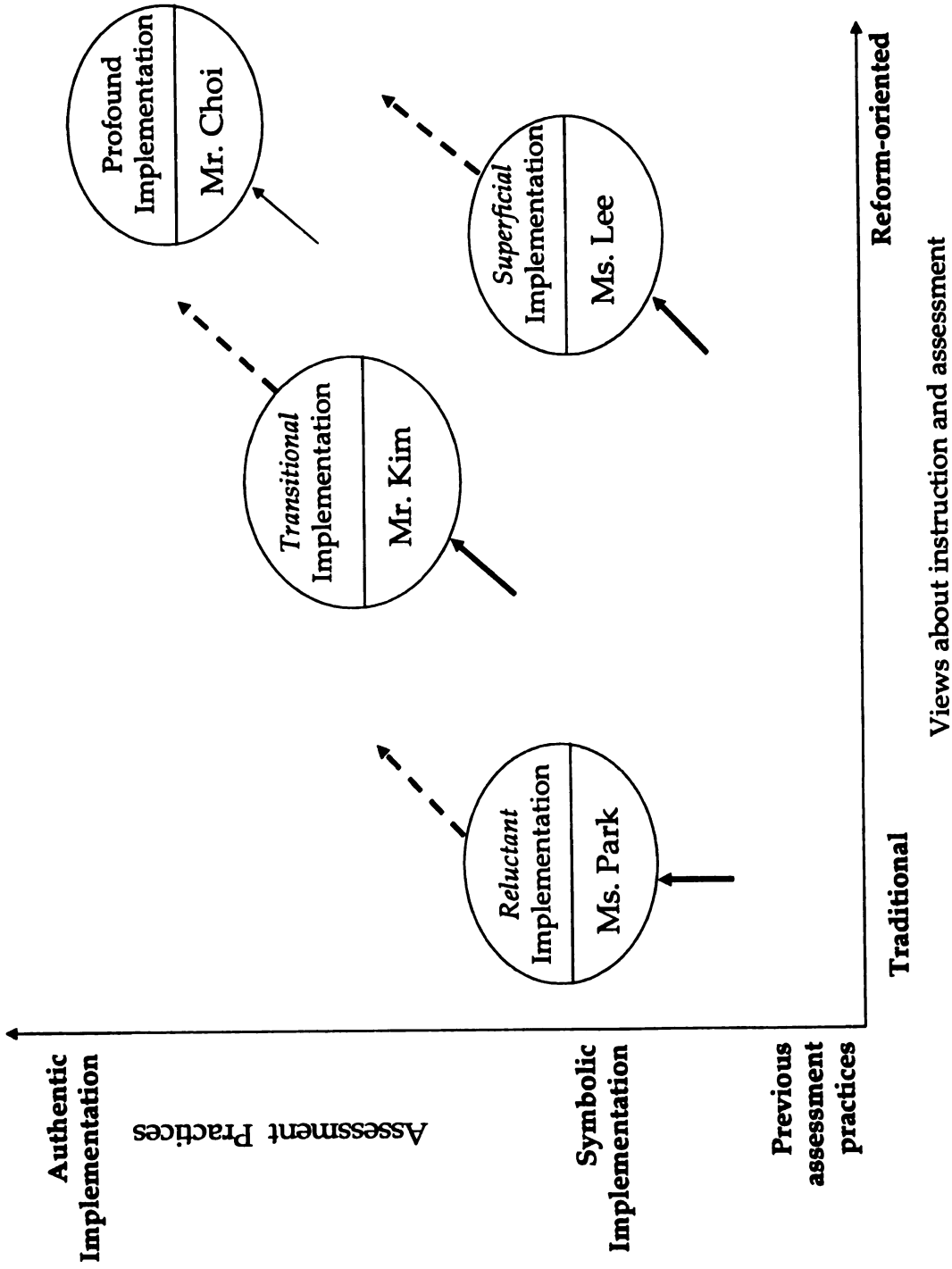


Figure 26. A Model for Combining Mandates with Scaffolding Instruments

Since current reform efforts require teachers to do something new or unfamiliar (uncomfortable) to them, teacher learning is necessary for them to implement the reforms successfully. To integrate new ideas into classrooms, teachers need to understand new ideas. For example, to successfully implement performance assessment reform, teachers should understand what and how to assess using performance assessment formats. Also, new assessments need to change their old beliefs about instruction and learning. However, many reforms in South Korea were mandated without giving sufficient learning opportunities for teachers to realize the reform visions in their classrooms, and these reform efforts have been unsuccessful. Thus, to move from policy changes toward real changes on practices, teachers should have sufficient and high-quality learning opportunities. These learning opportunities can promote teacher capacity and will for implementing reforms, as well as help teachers to change their practices at the deeper level.

The quality of learning opportunities should incorporate three aspects related to three dimensions of learning: content, pedagogy, and context. It would be better for these aspects to integrate mutually when creating teachers' opportunities of learning. While my suggestions are based on these three aspects of learning opportunities that have been emphasized in both research on teacher learning, and my study emphasized the importance of these three aspects of learning opportunities, South Korean contexts also need to be considered to make the suggestions more realistic.

The first aspect of learning opportunities is that *content should focus on a specific reform and subject*. For example, for success of the performance assessment reform, social studies teachers had professional development programs designed specifically for

performance assessment and social studies. By participating in the professional development program that focused on content designed specific for a reform, teachers can learn what and how to teach for their students.

The case of Ms. Lee illustrates the importance of helping teachers to learn about performance assessment. Even if her beliefs seemed to align with the principles of the reform, she lacked competence for implementing the performance assessment reform. The urgent task for her would be learning about performance assessment. She needs to learn how to design and score performance assessment tasks to assess what she intended (higher-order thinking). While she needs to understand the strategies of performance assessment, Ms. Lee needs to learn what and how to teach in social studies, since assessment cannot be separated from instruction, so both instruction and assessment must be changed together.

The second aspect of learning opportunities is that teachers should engage in authentic learning activities. Even if learning opportunities focused on content specific for a reform, but if teachers learn the content only in traditional ways (e.g., listen to lecture, memorize content), these learning opportunities would not help teachers to have deep understanding about the content or to change their practices. This situation would be similar when we think about teachers' instruction and assessment in the case studies. In the case studies, teachers like Ms. Lee or Ms. Park focused on content, but they did not encourage students to think about the content.

Thus, both content and pedagogy should be considered together when professional development programs are designed. When both aspects of learning opportunities are integrated into professional development programs, the effects of the

professional development programs would be powerful for changing practices as well as promoting teacher capacity and will. When teachers engage in authentic teacher learning activities, they may experience the importance of changes on their practices as well as gain more practical knowledge that can be applied to their classrooms.

The importance of authentic teacher learning can be illustrated by considering the case of Ms. Park. Since her belief about instruction and assessment did not fit the reform visions, she was reluctant to implement the performance assessment. For teachers like Ms. Park, just providing more resources or more times would not be a matter for helping them to fully implement the reform from their current point. Without a rationale for implementing the performance assessment reform, teachers may not make vigorous efforts to implement the reform. Authentic teacher learning can help teachers to change their beliefs about instruction and assessment. While teachers experience the ways in which they were expected to provide for student learning, they may realize the importance of changes on their practices.

Creating professional development programs that integrate both specific content and learning opportunities may need more resources (e.g., money, experts, space). Thus, given limited resources, formal professional development programs provided by the top (e.g., government or region) would be reasonable. Even it were difficult for government to provide many quality development programs designed for a specific school, because these programs target teachers from different schools, it would be more beneficial if the professional development programs can be designed for same-grade teachers because teachers can easily and immediately integrate what they learn into their classrooms. Since all same-grade teachers, even if from different schools, have the same national curriculum

and the same social studies textbook, and work under a similar educational system, the national curriculum can be an important means of helping them to work together with a high degree of commonality.

While I suggest formal professional development programs focusing on both content and authentic teacher learning, these formal programs should be connected to the third aspect of learning opportunities. Because formal professional programs target teachers from different schools and different grades, what teachers learn during professional development programs would not work for specific contexts. The third aspect of learning opportunities can help teachers to revise what they learn from the formal professional development programs based on their contexts.

The third aspect of learning opportunities is collaboration with fellow teachers. This aspect of collaboration is related closely to authentic teacher learning because authentic teacher learning can encourage teachers to collaborate with other teachers during PD sessions. While authentic teacher learning could incorporate the collaborative aspect into the formal professional development, collaboration in schools is also helpful for teachers because these learning opportunities are situated in the actual contexts where the new ideas are implemented. Combining both formal professional development that the government controls, and collaborations with fellow teachers, would be a way to balance between the top-down and bottom-up strategies.

Teacher collaboration as a strategy for supporting teachers is not new. Elementary teachers in South Korea have had opportunities to meet and talk with fellow teachers because schools set up regular meetings for teachers and some schools have attempted to provide spaces for teachers teaching the same grades. However, whether they are

regularly meeting with, or talking to, each other does necessarily help teachers change their practices. I suggest that we consider both substantial aspects of collaboration (e.g., what teachers discuss or what they do during meetings) and structural aspects of collaborations (e.g., time and space).

I suggest considering three aspects of collaboration for helping teachers to change their practices. First, working together for specific tasks that are closely related to their instructional practices would be more effective than merely discussing with other teachers. For example, teachers can work together to design performance assessment tasks and rubrics for scoring the assessment tasks for better assessment of students. Working together with other teachers naturally embraces teachers' discussion about ideas, problems, and challenges. When teachers work together with collaborative tasks for improving their instructional practices, they have clear targets to collaborate with other teachers and real benefits of collaboration.

Second, collaborations with fellow teachers teaching the same grade would be more beneficial, as I suggested for designing formal professional development programs. Even if this study focuses on social studies classes, my suggestions may be applied to professional development programs for other subjects. Since elementary teachers teach many subjects, their instructional concerns are not restricted only to social studies classes. Some problems may occur when they teach or assess other subjects. Solutions to problems in one may well be applicable or at least partially applicable to problems in another. Elementary teachers teaching the same grade can discuss their problems and challenges as well as work together to develop solutions with a high degree of commonality.

Third, collaboration with fellow teachers needs to be supported by mentors or change agents who may be expert teachers like Mr. Choi in my case studies. While collaboration with fellow teachers may offer emotional support when they have had frustration from implementing reform ideas, as well as contrive solutions to problems that they had, collaborations with teachers may have limitations. We can image that all teachers who collaborate are like Ms. Park in the case studies .If teachers do not have any rationales for implementing reform ideas, how can teacher collaboration help them change their practices? Based on my experience, collaboration with fellow teachers in schools was very helpful to me to try new ideas, but when I faced some challenges or problems, while attempting new ideas, and when other teachers could not provide any ideas to solve problems, I needed help from experts. However, it was extremely hard to find someone appropriate to help me. I suggest that schools provide change agents who can support teacher collaboration.

However, we should remember it is important to remove contextual barriers blocking teachers' implementation of reform ideas, even if this does not necessarily help teachers to implement the reform fully if teachers do not have strong capacity and will. Even if teacher capacity and will influenced by teacher learning are critical factors for changing their practices, teachers cannot be freed by school contexts.

The case of Mr. Choi in profound implementation illustrates the importance of school contexts for sustained changes. Even if he showed profound implementation at the current point, we do not know whether his practice is durable, because he mentioned that because of class size he could not use the portfolio as one of the new assessment formats. Also he mentioned that he was losing his energy for implementing the performance

assessment since he had to spend extra time after school to provide feedback to the large number of students. To sustain his changes on assessment practices, policymakers should remove contextual barriers, such as by providing time for implementing the reform.

Concluding Remarks

This study shows the complexity of implementing reform, since many interrelated factors influence teachers' practices. It indicates how difficult it is to change teachers' practices. Even if a policy mandates a reform, assuming that the mandate would change their practices, many teachers only would implement the reform at the surface level. To reduce the discrepancies between policy (intended reform) and practice (enacted reform), policy-makers should consider adding scaffolding instruments to complement the mandate of the reform. I suggest that the quality of learning opportunities is the best scaffolding instrument to support teachers. While teacher learning can play a critical role in connecting policy and practices, we should not expect that teachers' practices will change within a short period. We should wait until teachers learn new ideas and implement new ideas through trial and error, providing teachers with the right kinds of support for their implementation. Fullan(1993) said that, "problems are our friends.". Starting the problems teachers faced, policy-makers have to plan what they need to help them to move toward fundamental changes. When teachers feel comfortable facing problems, and searching and applying possible solutions to the problems they have, they can make progress in changing their practices.

APPENDICES

Appendix A

Questionnaire

Background Information

1. Location of School (Please mark (V) only one box):

☐ Seoul ☐ Incheon ☐ Kyungi

2. Type of school: ☐ Public ☐ Private

3. School Size (# of classes in your school): _____

4. Gender: ☐ Female ☐ Male

5. Grade level you teach this year: ☐ 3 ☐ 4 ☐ 5 ☐ 6

6. Number of students in your classroom: _____

7. Achievement level of students in your classroom this year:

☐ Mostly high achieving

☐ Mostly average achieving

☐ Mostly low achieving

☐ Students at a range of achievement levels

8. Teaching experience you have had as a full time teacher (Please include this year):

_____ years, _____ months

9. Please indicate your highest degrees:

☐ BA/BS ☐ MA/MS ☐ Ed.D/Ph.D ☐ Other (Please specify): _____

Assessment Practices in Social Studies

10. This semester, how often did you use each of the following in your social studies class? Circle only one in each row.

Assessment formats	Never	Once or twice per semester	Less than once per month but more than twice per semester	Once or twice a month	Once a week
a. Teachers' own paper-and-pencil tests including multiple choice, true and false, short answer fill-in tests	1	2	3	4	5
b. Open-ended exercises that require students to construct response to prompts or problems within a short time	1	2	3	4	5
c. Extended tasks that last longer than open-ended exercises such as projects	1	2	3	4	5
d. Portfolio	1	2	3	4	5
e. Demonstration that takes the form of student presentations of their work	1	2	3	4	5
f. Observation	1	2	3	4	5

11. This semester, how did you use the results of assessment in social studies? To what extent did you use each of the following? Please circle only one in each row.

Purposes of Assessment	None	A little	Somewhat	Moderately	A great deal
a. To Assign grades to students	1	2	3	4	5
b. To give report cards to students and parents	1	2	3	4	5
c. To identify areas where students need improvement	1	2	3	4	5
d. To improve teaching practices or plans for future	1	2	3	4	5

12. This semester, how often did you give each of the following kinds of problems for assessing student learning in social studies? Circle only one in each row.

Cognitive Demands of Assessment	Never	Once or twice per semester	Less than once per month but more than twice per semester	Once or twice a month	Once a week
a. Problems that require students to recall facts and concept	1	2	3	4	5
b. Problems that require students to use methods of inquiry (e.g. collecting and analyzing data or information).	1	2	3	4	5
c. Problems that allow students to criticize a position and justify their own position using evidence.	1	2	3	4	5
d. Problems that find answers from social studies textbook	1	2	3	4	5
e. Problems that require students to develop their own interpretations and stories, not merely take from social studies textbooks or what teachers told.	1	2	3	4	5
f. Problems emphasizing relationships among social studies concepts.	1	2	3	4	5
g. Problems with more than one possible interpretation or many plausible answers	1	2	3	4	5
h. Problems that require students to show their understanding by using various representations (e.g. writing, tables, graphs)	1	2	3	4	5
i. Problems with only one correct answer.	1	2	3	4	5
j. Problems that require students to apply social studies concepts to real world problems.	1	2	3	4	5
k. Problems that require students to explain how they solve the problem and reasons in support of their answers	1	2	3	4	5

Teacher Capacity and Will

13. Different teachers have different philosophies. For each of the following pairs of statements, check the box that best shows how closely your own beliefs match each of the statements in a given pair. The closer your beliefs to a particular statement, the closer the box you check. Please check only one for each set.

a. "I mainly see my role as a facilitator. My job is to provide opportunities and resources for students to discover or construct concepts for themselves."

☐ ☐ ☐ ☐ ☐ ☐ ☐

"That's nice, but students really won't learn the subject unless you go over the material in a structured way. It is my job to explain, to show students how to do the work, and to assign specific practices."

b. "The most important part of instruction is the content of the curriculum. That content is the community's judgment about what children need to be able to know and do."

☐ ☐ ☐ ☐ ☐ ☐ ☐

"The most important part of instruction is that it encourages thinking among students. Content is secondary."

c. "It is good idea to have many activities going on in the class at the same time. Students learn doing different activities"

☐ ☐ ☐ ☐ ☐ ☐ ☐

"It's more practical to give the whole class the same assignment, one that has clear directions, and one that can be done in short intervals that match students' attention spans and the daily class schedule."

d. "Students learn more when the teacher gives background and directly teaches concepts."

☐ ☐ ☐ ☐ ☐ ☐ ☐

"Students learn more when the teachers use active approaches like student discussion, projects, and presentations."

14. How would you characterize your willingness to implement performance assessment in your class?

None				Some			Extensive
1	2	3	4	5	6	7	

15. How would you rate your knowledge about each of the following? Please circle only one in each row.

Content	Very poor	Poor	Adequate	Good	Excellent
a. Performance assessment in social studies	1	2	3	4	5
b. What to assess by performance assessment	1	2	3	4	5
c. Selecting/Developing good assessment tasks	1	2	3	4	5
d. Selecting/Developing good rubrics	1	2	3	4	5
e. Scoring student responses	1	2	3	4	5

16. Indicate how much time you have spent, in 2003-2004, in professional development activities devoted to each of the following areas.

Content	Hours						
	Never	Less than 3 hours	3-5 hours	6-10 hours	11-20 hours	21-30 hours	Over 30 hours
a. 7 th curriculum in general	1	2	3	4	5	6	7
b. 7 th social studies curriculum	1	2	3	4	5	6	7
c. Social studies instruction	1	2	3	4	5	6	7
d. Social studies assessment	1	2	3	4	5	6	7
e. Performance assessment in general	1	2	3	4	5	6	7
f. Social studies performance assessment	1	2	3	4	5	6	7

Teacher Learning Opportunities

17. Indicate how often you did the following during the PD activities?

Learning Activity	Never	Seldom	Sometimes	Often	Very often
a. Listened to a lecture	1	2	3	4	5
b. Did actual work (e.g., developing assessment plans, assessment tasks, or rubrics)	1	2	3	4	5
c. Engaged in projects that take longer time	1	2	3	4	5
d. Completed paper-and-pencil tests	1	2	3	4	5
e. Made portfolio	1	2	3	4	5
f. Presentation/Leading discussion	1	2	3	4	5

18. In your own school, have you been involved in developing the following? Please check all which apply.

- ☐ None
- ☐ Developing school curriculum framework or content standards
- ☐ Developing school assessment system
- ☐ Developing performance standards
- ☐ Developing assessment tasks
- ☐ Developing assessment rubrics

19. To what extent did you have opportunities for discussing with fellow teachers to implement performance assessment?

None Some Extensive
 1 2 3 4 5 6 7

20. To what extent did you have opportunities for working together with fellow teachers to develop performance assessment tasks or rubrics?

None Some Extensive
 1 2 3 4 5 6 7

School Contexts in which Teachers Work

21. The following statements describe teachers' work environments. Please indicate how much each statement agrees or disagrees with your own work environment.

	Strongly Disagree	Moderately Disagree	Slightly Disagree	Slightly Agree	Moderately Agree	Strongly Agree
a. Other teachers encourage me to try new ideas.	1	2	3	4	5	6
b. Teachers who successfully introduce a major innovation in their teaching are given public recognition among other teachers.	1	2	3	4	5	6
c. New ideas presented at in-services are discussed afterwards by teachers in this school.	1	2	3	4	5	6
d. Teachers in this school are continually learning and seeking new ideas.	1	2	3	4	5	6
e. Major professional development activities are followed by support to help teachers implement new practices.	1	2	3	4	5	6

22. Indicate how much you agree or disagree with each of the following statements. Circle only one in each row. PA= performance assessment

	Strongly disagree	Moderately disagree	Slightly disagree	Slightly agree	Moderately agree	Strongly agree
a. My principal supports using PA.	1	2	3	4	5	6
b. Students support using PA.	1	2	3	4	5	6
c. Parents support using PA.	1	2	3	4	5	6
d. My colleagues support using PA.	1	2	3	4	5	6
e. Time available to prepare and implement PA is sufficient.	1	2	3	4	5	6
f. Resources for PA are sufficient.	1	2	3	4	5	6

(Optional) The researcher needs to know which teachers have completed the survey, so please put your name, school and address, on this page. Published reports of the research will not refer to the actual names of schools and individuals participating in the study, and the research will not reveal to anyone your individual response.

Name: _____ School: _____

Home Address: _____

Email Address (If available): _____

THANK YOU for your thought, time, and effort in completing this questionnaire.

Appendix B

Descriptions of Questionnaire Items

Table 27

Descriptions of Questionnaire Items

Categories	Measures	Items (question number)
Teacher Implementation	Formats of Assessment	
	Traditional	10-a
	Performance*	10-b through f
	Cognitive Demands of Assessment	
	Lower-level	12-a, d, i
	Higher-level*	12-b, c, e, f, g, h, j, k
	Purposes of Assessment	
	Assessment of Learning	11-a, b
	Assessment for Learning*	11-c, d
Teacher Capacity & Will	Teacher Knowledge*	15-a through e
	Teacher Belief*	13-a through d
	Teacher Willingness*	14
Teacher Learning Opportunities	Content of Professional Development	16-e and f
	Specific content*	16-a through d
	Other content areas	
	Pedagogy of Professional development	17-a
	Traditional Teacher Learning	17-b through f
	Authentic Teacher Learning*	19, 20
	Collaboration*	18
	Involvement*	

Categories	Measures	Items (question number)
School Contexts	School Climate	21-a through e
	School Communities' Attitude	22-a through d
	School Resources*	22-e, f
	Class Size*	6

* Measures used in regression statistical analyses

Appendix C

Interview Protocol

Before the interview:

A week prior, I will list the topics I plan to cover in the interviews. The topics are: (a) your background information, (b) your knowledge and beliefs about teaching, learning, and assessment in social studies, (c) your understanding about performance assessment, (d) your implementation of performance assessment, (e) professional development programs you had, and (f) factors that influence your assessment practices. I will ask teachers to bring artifacts such as samples of assessment tasks, rubrics, and report cards in your social studies classrooms and your assessment plans.

Introduction of the interview:

I will ask for permission to tape the interview. Describe general purpose: I am interested in how you implement performance assessment and what factors influence your implementation but I would like to start with a few more general questions.

Before I begin the interview, I will have the teachers read the consent form and then obtain their signature on the form: "I assure you that your identity is confidential and nothing you say will be associated with you or your identity. Please read carefully the consent form and sign on the form if you wish to participate in this study. If you have questions for me at any time during the interview, please ask. If at any point you would like the tape recording turned off, the off switch is immediately available to you. You also can withdraw from the study at any time. Do you have any questions before we begin?"

Set # I: Background Information

Purpose: The purpose of the first set of questions is to gather information about you as a teacher. This will include questions on teaching background, grades taught, years taught, the reason you entered teaching, and your classroom.

- How long have you been teaching?
- What prompted your decision to be a teacher?
- In what did you major?
- What kind of certification do you hold?
- What grades do you teach this year? Describe your classroom this year.
- What do you like/dislike about teaching in social studies?
- What is your elementary social studies background?
- Did you like social studies subjects when you were a student?

Set # 2: Beliefs and Knowledge

Purpose: The second set of questions is designed to gather information about your understanding of the current curriculum and assessment reform in social studies, views/attitude toward the current curriculum and assessment reform in social studies.

Knowledge about social studies

- Do you believe you have in-depth knowledge about social studies? Why?

Before the interview, I will select a unit for each grade, then ask the following:

- Can you tell me what the unit is about?
- What important social studies concepts or big ideas should be learned in this unit? Why?
- What have you learned to teach this unit from your education? Do you think you are well prepared to teach this unit?

Beliefs about teaching and learning in social studies

- What are the important goals of teaching and learning in social studies? What do you want your students to know and be able to do? Why do you think these are important?
- What do you think would be good teaching to achieve your mentioned goals? What learning experiences do students need to have?
- Describe for me a good lesson you believe. Include instructional activities, the teacher and students' roles, and the use of the social studies textbook.

Beliefs about assessment

- What would your ideal assessment look like?
Probe: What are the important learning outcomes for assessment?
What are the important purposes of assessment?
What kinds of assessment do you think are the most beneficial to your students?
- Please select the most important assessment you conducted? I will ask the following questions.
Probe: What are the purposes of the assessment and the reason for its importance?
What did you want to measure? (e.g., recall of SS facts)
Describe how you measured (e.g., true-false tests, observations) and why you selected the forms.

Knowledge about performance assessment:

- How do you define performance assessment?
Probe: What are key points on performance assessment?
How does performance assessment differ from other assessments you used?
- Do you feel well prepared to use performance assessment? Include what to assess, how to assess, how to score, and how to use and report the results. Why do you think so?

Attitude toward performance assessment

- What do you think about performance assessment?
Probe: Do you believe that performance assessment is generally a good assessment method? Why?
Is performance assessment a better assessment approach for your students when compared to traditional tests (multiple-choice, false/true, and short answer paper-and-pencil tests)? Why?
- How do you feel about implementing performance assessment in your social studies classroom?
Probe: Do you like/dislike implementing performance assessment in your social studies classroom? Why?

Set # 3: Implementing Performance Assessment

Purpose: The next set of questions was designed to gather information about your assessment practices, implementation of performance assessment in social studies, and factors influencing implementation.

- Tell me how you assess students in your social studies classroom.
- Please show me samples of assessment tasks including performance assessment tasks you used this semester? Looking at the samples, I will ask the following questions.

What specific learning outcome did you assess using these tasks?

How were the assessments aligned with your curriculum?

How were tasks incorporated into your instruction?

How did you assess? Did you use rubrics? If so, please show me the rubrics.

How did you use the results of assessment?

How did you fill out your report cards? Please show me an example of report cards?

Did the tasks work well with your students?

What challenges, problems have you faced using different assessment tasks?

Impact of implementing PA

- Since performance assessment policy has been mandated, what happens in your classrooms?
Are you doing anything differently in your assessment practices since implementing PA?
Have you noticed any changes in your classroom since implementing PA?
For you and for your students, what is/are the good and/or the bad in implementing performance assessments?

Factors influencing implementation:

- What factors supported or constrained your implementation?

- Were you in charge of the implementation of PA, what would be the one or two most important things you would do to ensure that the performance assessment was implemented well?

Set # 4: Professional Development

Purpose: Questions in set # 4 are designed to gather information about professional development activities you had.

- Please describe mandatory professional development programs in which you participated to learn and implement performance assessment.

Probe: What did you learn?

How did you learn in the programs?

What role did you play as learner?

What types of activities were used in the programs?

What did you like/dislike about the programs/activities?

- Please describe professional development programs in which you volunteered to participate to learn and implement performance assessment?

Probe: What did you learn?

How did you learn in the programs?

What role did you play as learner?

What types of activities were used in the programs?

What did you like/dislike about the programs/activities?

- What kinds of learning opportunities were most valuable / least valuable to you? Why?

- Did you have an opportunity to work regularly with other teachers on performance assessment? If so, please describe. How often did you meet? What did you learn? What was useful?

- Have you done anything differently in your instruction and assessment since you learned about PA in PD programs? Has your role and have your students' roles changed? Are you asking students to do anything differently? If so, describe.

- What do you recommend subsequent/next PD sessions should do to help you implement PA well? What kinds of things would you like to see addressed in the next PD sessions?

That is the end of our interview. Are there things you wish to say that you did not have an opportunity to say? Are there questions that should have been asked but were not?

Thank you for your time.

Appendix D

Categories and Examples of Interview Language for Coding Transcripts

Table 28 presents the categories used to code teachers' assessment practices. The categories were selected based upon a review of previous studies that examined teachers' assessment practices. Based on these categories, I coded transcripts, using the computer software program N-Vivo.

Although my main focus in analyzing interview data was to examine how teachers implement the South Korean performance assessment reform, I asked interview questions to understand their practices and coded these interview data, using several categories (e.g., teachers' educational and professional backgrounds, school assessment systems, instructional practices connected with assessment practices, views about instruction and assessment, preparedness to implement performance assessment, and problems and difficulties in implementing performance assessment).

Table 28

Categories for Sorting Data about Teachers' Assessment Practices

Category		Examples of interview language and assessment documents*
Surface Aspect	<i>Formats of Assessment</i>	[Codes here are based on both interview transcripts and assessment tasks teachers shared with me during the interviews]
	Traditional formats	Multiple-choice, quizzes, paper-and pencil tests**
	Performance formats	Observation, project, report, portfolio, demonstration, presentation

Deep Aspect	<i>Cognitive demands</i>	
	Lower-level	<p>[Excerpts from interview transcripts] After that, students took paper-and-pencil tests. I graded the tests based on the number of right answers.</p> <p>[Excerpts from interview transcripts] I asked my students to collect anything they had done and file it in a scrapbook [portfolios]. I also asked them to put unit tests into the scrapbook. When the students planned to take mid-term and final tests, I asked them to look at their results on the unit tests. And I said, "the test will be based on the questions I gave in the unit tests. Try to look at your wrong answers and remember the right answers.</p> <p>[Excerpts from interview transcripts] The grade assessment plan suggested I employ a task to assess whether students could draw a chronological table. Since I did not employ this task, I used the information obtained from paper-and-pencil tests to grade this task. The paper-and-pencil tests included several questions about a map. If students had good scores on these questions about a map, I assumed that the students could draw maps</p> <p>[Excerpts from assessment documents] What does the name of the palace [written in Chinese characters] mean? When was it constructed?</p>
	Higher-level	<p>[Excerpts from interview transcripts] I wanted to assess students' abilities to develop their arguments based on their understanding of content.</p> <p>[Excerpts from assessment documents] <i>Assessment criteria: (1) Did the student understand the arguments of the different applied scientists? (2) Did the student's essay reflect the problems of the time? and (3) Did the student write his or her arguments showing appropriate evidence?</i></p> <p>[Excerpts from assessment documents] <i>How would you write a public statement for reforming people's ideas assuming that you are one of the applied scientists?</i></p> <p>[Excerpts from assessment documents] <i>Compare palaces in other countries with one palace you visited, to think about different historical contexts"</i></p>

Deep Aspect	<i>Purposes of assessment</i>	
	<i>Assessment of learning</i>	<p>[Excepts from interview transcripts] I feel that I am doing assessment for assessment's sake.</p> <p>[Excepts from interview transcripts] "I was busy with assessment at the end of the semester because I had to submit my grade-book and write report cards. I had to employ assessments if I missed assessment areas included in the fifth grade assessment plan.</p>
	<i>Assessment for learning</i>	<p>[Excepts from interview transcripts] I had a disabled student. I knew how much the student's learning improved from the beginning to the end of the semester. My assessment for this student was very informative because I observed the student very carefully. The different way in which I assessed the student was reflected in what I wrote in his report card. In the report card, I could provide the student with more details about assessment. The assessment results were not based on comparison with other students, but focused on his improvement of learning.</p> <p>[Excepts from interview transcripts] I always provide feedback. Students do not like having my feedback because it means they should revise their essays. However, I see their improvement as time goes on.</p>

- * Assessment tasks used by teachers were analyzed with their interview transcripts.
 ** Paper-and-pencil generally refers to tests consisting of selected response items.

Based on these categories, passages in each case were sorted. I read the sorted passages to examine how each teacher implemented the performance assessment in terms of the three aspects of assessment (formats, cognitive demands, and purposes of assessment). Since teachers' interview questions were organized around specific documents, such as assessment plans and assessment tasks, I also analyzed teachers' assessment practices by looking at assessment documents teachers had brought to the interviews. Each teacher's responses to the performance assessment reform were

examined in terms of the three aspects of assessment. Then, teachers' responses from these three aspects were compared, to find similarities and differences among the cases, and the three patterns (profound, transitional, and symbolic implementation) were initially identified.

The first pattern (profound) was implementation of all three aspects of performance assessment: (1) used performance assessment formats, (2) assessed higher-level cognitive demands, and (3) used assessment *for* learning. The second pattern (transitional) was implementation of the surface aspect of the assessment (formats), but failure to implement one of the deep aspects of the assessment (cognitive demands or purposes of assessment). For example, a teacher used performance assessment formats to assess higher-level cognitive demands, but did not use assessment *for* learning. The third pattern (symbolic) was implementation of only the surface aspect of the assessment (formats). For example, a teacher used performance assessment formats, but focused on assessing lower-level cognitive demands and used assessment only for assigning grades or providing report cards.

After identifying these three implementation patterns, I sorted transcripts according to them to see whether any differences emerged within an identified pattern. Carefully rereading the sorted transcripts allowed me to find that teachers in the third pattern, who implemented only the surface aspect of the assessment, had different views about the reform. While one teacher believed strongly in the reform, another had traditional views about instruction and assessment. Finding these different views among teachers, I reread those parts of the transcripts related to these views about instruction assessment and coded again. Table 29 shows the categories and examples of the interview language I used to code these views.

Table 29

Categories for Sorting Data about Teachers' Views about Instruction and Assessment

Category	Examples of interview language
Traditional	<p>Assessment for the new curriculum requires teachers to use open-ended questions that do not have right answers. All answers are right in this kind of assessment. This caused a problem. I said that there was no wrong answer when students answered the questions. After that, students took paper-and-pencil tests. I graded the tests based on the number of right answers. After receiving their test scores, students complained that their scores were unfair. I think that there are right answers and that students should know that the right answers are found in the social studies textbook."</p> <p>I use performance assessments because I was required to use them, but I think that quizzes or unit tests are better for checking what students know.</p>
Reform-oriented	<p>The main focus of instruction was not to tell facts to students. Knowing just names of different kingdoms or events that happened is not enough for learning history.</p> <p>Students think that the historical knowledge found in their social studies textbook is always true. I tried to get students to understand that historical knowledge in the social studies textbook is one of the interpretations created by historians. I taught that historical interpretation can be different based on different historical perspectives. Instead of memorizing the historical facts represented in the social studies textbook, students need to learn the importance of interpretation and to build their own perspectives.</p> <p>Teachers who teach the same grade usually give grades in knowledge based on students' scores on traditional tests. Whereas I think that traditional tests can easily measure whether students acquire knowledge of facts, these test scores alone are not enough to say what students know.</p> <p>I think that the good thing about using performance assessment is that it allows me to examine what my students understand by looking at their performance. I can assess what a student knows using new formats of assessment. Both Knowledge and Thinking Skills can be assessed with these types of formats. Traditional paper-and- pencil tests have a limitation. When using this type of test, it is difficult to assess both of these assessment areas."</p>

Based on both teachers' reported practices and views about the reform, the initially identified patterns were changed into four patterns. I expanded the third initial pattern (symbolic) to include the third and fourth final pattern (superficial/Ms. Lee and Reluctant/Ms. Park).

Appendix E

Correlation Tables

Table 30

Correlations between School Contexts and Teacher Implementation

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Implementation	1												
2. Parent support	.23**	1											
3. Student support	.22**	.48**	1										
4. Principal support	.04	.36**	.27**	1									
5. Colleague support	.18**	.49**	.40**	.36**	1								
6. Encouraging new ideas	.12**	.03	.08*	.05	.04	1							
7. Innovative teachers	.12**	.17**	.13**	.18**	.11**	.42**	1						
8. Discussion with new ideas	.22**	.15**	.13**	.16**	.08*	.46**	.50**	1					
9. Keep learning new ideas	.16**	.16**	.13**	.21**	.13**	.42**	.42**	.52**	1				
10. PD for new ideas	.18**	.17**	.14**	.08**	.06	.46**	.30**	.43**	.42**	1			
11. Time	.17**	.12**	.12**	-.08*	.11**	.15**	.13**	.14**	.04	.18**	1		
12. Resource	.21**	.14**	.12**	-.02	.14**	.14**	.14**	.15**	.07	.20**	.66**	1	
13. Class Size	.05	-.05	.03	.00	-.01	-.11*	-.09*	-.06	-.03	-.12**	-.05	-.06	1

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

Table 31

Correlations between Teacher Learning Opportunities and Teacher Implementation

	1	2	3	4	5	6	7	8	9	10
1. Implementation	1									
2. Discussing	.16**	1								
3. Working together	.20**	.52**	1							
4. Involvement	.13**	.15**	.18**	1						
5. Social studies assessment	.16**	.09*	.09*	.09*	1					
6. Social studies PA	.18**	.09*	.10*	.10**	.83**	1				
7. General curriculum	.15**	.09*	.16**	.12**	.65**	.65**	1			
8. General PA	.12**	.06	.07	.14**	.82**	.72**	.65**	1		
9. Social studies curriculum	.06	.09*	.14**	.18**	.45**	.45**	.56**	.43**	1	
10. Social studies teaching	.12**	.12**	.10*	.17**	.64**	.70**	.52**	.59**	.51**	1

* $p < .05$; ** $p < .01$

Table 32

Correlations between Teacher Capacity & Will Variables and Teacher Implementation

	1	2	3	4	5	6	7	8	9	10	11
1. Implementation	1										
2. Learning	.15**	1									
3. Curriculum	.15**	.29**	1								
4. Role of teacher	.12**	.29**	.32**	1							
5. Student activity	.19**	.20**	.21**	.27**	1						
6. What to assess	.28**	.05	.03	-.00	.03	1					
7. Assessment tasks	.32**	.05	.06	.04	.04	.64**	1				
8. Rubrics	.32**	.03	.06	.05	.02	.58**	.79**	1			
9. How to score	.25**	.03	.04	.04	.03	.60**	.71**	.75**	1		
10. Use of PA in social studies	.31**	.08*	.09*	.06	.06	.56**	.54**	.52**	.52**	1	
11. Teacher willingness	.31**	.09*	.06	.05	.05	.29**	.18**	.18**	.19**	.24**	1

* $p < .05$; ** $p < .01$

REFERENCES

REFERENCES

- Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis*, 12(3), 247-259.
- Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning professional: Handbook of policy and practice* (pp. 3-32). San Francisco, CA: Jossey-Bass Inc.
- Beck, S. (1998). *Theory and practice of performance assessment*. Seoul: Won-Mi Publisher.
- Borko, H., Elliott, R., & Uchiyama, K. (1999). *Professional development: A key to Kentucky's reform effort*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Borko, H., Mayfield, V., Marion, S. F., Flexer, R. J., & Cumbe, K. (1997). *Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Bulach, C., & Malone, B. (1994). The relationship of school climate to the implementation of school reform. *ERS SPECTRUM*, 12(4), 3-8.
- Byrne, B. M. (2000). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Chappuis, S., Stiggins, R., Arter, J., & Chappuis, J. (2004). *Assessment for learning: An action guide for school leaders*. Portland, OR: Assessment Training Institute.
- Cho, D. (1999). Performance assessment in social studies. *Journal of Citizenship Education*, 29, 161-191.
- Choi, H. (1998). A study on the way of aligning performance assessment with school curriculum. *The Journal of Curriculum Studies*, 17(1), 45-67.
- Cimbricz, S. (2002). State-mandated testing and teachers' beliefs and practice. *Education Policy Analysis Archives*, 10(2).
- Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis*, 23(2), 145-170.

- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.
- Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, 12(3), 311-329.
- Cohen, D. K., & Barnes, C. A. (1993). Pedagogy and policy. In D. K. Cohen, M. W. McLaughlin & J. E. Talbert (Eds.), *Teaching for understanding: Challenges for policy and practice*. San Francisco, CA: Jossey-Bass Publishers.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The mathematics reform in California. *Teachers College Record*, 102(2), 294-343.
- Cohen, D. K., & Hill, H. C. (2001). *Learning policy: When state education reform works*. New Haven, CT: Yale University Press.
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (second edition). Thousand Oaks, CA: Sage Publication Inc.
- Cronin Jones, L. L. (1991). Science teacher beliefs and their influence on curriculum implementation: Two case studies. *Journal of Research in Science Teaching*, 28(3), 235-250.
- Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms 1890-1990*. New York: Teachers College Press.
- Darling Hammond, L., & Aneess, J. (1996). Authentic assessment and school development. In J. B. Baron & D. P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities*: The National Society for the Study of Education.
- Darling-Hammond, L., Aneess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York: Teachers College Press.
- Davis, K. S. (2002). Change is hard: What science teachers are telling us about reform and teacher learning of innovative practices. *Science Education*, 87(1), 3-30.
- Department of Education. (1997). *Assessment of student performance*. Department of Education.
- Desimone, L., Porter, A. C., Birman, B. F., Garet, M. S., & Yoon, K. S. (2002). How do district management and implementation strategies relate to the quality of the professional development that districts provide to teachers?. *Teachers College Record*, 104(7), 1265-1312.

- 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.
- Eisner, E. W. (1999). The uses and limits of performance assessment. *Phi Delta Kappan*, May, 658-660.
- Elmore, R. F. (1995). Structural reform and educational practice. *Educational Researcher*, 24(9), 23-26.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), 95-113.
- Firestone, W. A., Monfils, L., & Camilli, G. (2001). *Pressure, support, and instructional change in the context of a state testing program*. Paper presented at the American Educational Research Association, Seattle, WA.
- Flexer, R. J., Cumbo, K., Borko, H., Mayfield, V., & Marion, S. F. (1995). *How "messaging about" with performance assessment in mathematics affects what happens in classrooms* (No. 396). Boulder: National Center for Research on Evaluation, Standards, and Student Testing.
- Flexer, R. J., & Gerstner, E. A. (1993). *Dilemmas and issues for teachers developing performance assessments in mathematics: A case study of the effects of alternative assessment in instruction, student learning, and accountability practices*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Francis, L. W. (1994). *Understanding the relationship between teachers' beliefs and the implementation of performance assessment: A exploratory case study*. Unpublished doctoral dissertation, University of South Carolina.
- Fuhrman, S., Clune, W. H., & Elmore, R. F. (1988). Research on education reform: Lessons on the implementation of policy. *Teachers College Record*, 90(2), 237-257.
- Fullan, M. (1991). *The meaning of educational change*. New York City: Teachers College Press.
- Fullan, M. (1992). Getting reform right: What works and what doesn't. *Phi Delta Kappan*, 73(10), 744-752.
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London: The Falmer Press.

- Fullan, M. (1994). Coordinating top-down and bottom-up strategies for educational reform. In R. F. Elmore & S. H. Fuhrman (Eds.), *The governance of curriculum*. Alexandria: The Association for Supervision and Curriculum Development.
- Gallucci, C. (2003). *Theorizing about responses to reform: The role of communities of practice in teacher learning*. Washington: Center for the Study of Teaching and Policy.
- Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L., Herman, R., & Yoon, K. S. (1999). *Designing effective professional development: Lessons from the Eisenhower program*. Washington, DC: U.S. Department of Education.
- Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective?: Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915-945.
- Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal*, 40(3), 731-767.
- Grant, S. G. (1998). *Reforming reading, writing, and mathematics: Teachers' responses and the prospects for systemic reform*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grant, S. G., Peterson, P. L., & Shojgreen-Downer, A. (1996). Learning to teach mathematics in the context of systemic reform. *American Educational Research Journal*, 33(2), 509-541.
- Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes & efficacy. *Education Policy Analysis Archives*, 13(10), 1-26.
- Jung, M. (2001). *Performance-based assessment in social studies*. Seoul: Hakmun Press.
- Kennedy, M. (1998). *Form and substance in inservice teacher education*: National Institute for Science Education, University of Wisconsin-Madison.
- Keon, D. (2000). Problems of performance assessment. *The Journal of Education Research*, 21(2), 83-102.
- Khatti, N., Kane, M. B., & Reeve, A. L. (1995). How performance assessments affect teaching and learning. *Educational Leadership*, 53(3).

- Khatti, N., & Sweet, D. (1996). Assessment reform: Promises and Challenges. In M. B. Kane & R. Mitchell (Eds.), *Implementing performance assessment: Promises, problems, and challenges* (pp. 1-21). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Kline, R. B. (1998). *Structural equation modeling*. New York: The Guilford Press.
- Knap, M. S. (2003). Professional development as a policy pathway. *Review of Research in Education*, 27, 109-157.
- Knapp, M. S. (2002). *Understanding how policy meets practice: Two takes on local response to a state reform initiative*. Seattle, WA: Center for the Study of Teaching and Policy.
- Korea Institute of Curriculum & Evaluation. (1999). *Study about implementation analysis of performance assessment policy and improvement solutions*. Seoul: Korea Institute of Curriculum & Evaluation.
- Little, J. W. (1993). Teachers' professional development in a climate of educational reform. *Educational Evaluation and Policy Analysis*, 15(2), 129-151.
- Liu, J. (2000). *The effect of performance-based assessment on eighth grade students' mathematics achievement*. University of Missouri-Columbia.
- Martinez, M. C. (2005). *Tracking the source of teaching practices*. Paper presented at the 26th Annual Ethnography in Education Research Forum, Penn State University. Philadelphia.
- Mayer, D. (1999). Measuring instructional practice: Can policy makers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.
- McLaughlin, M. W. (1991). Learning from experience: Lessons from policy implementation. In A. R. Odden (Ed.), *Education policy implementation* (pp. 185-195). Albany, NY: State University of New York Press.
- Miller, W. E. (1999). Temporal order and causal inference. *Political Analysis*, 8(2), 119-140.
- Mintrop, H. (1999). Changing core beliefs and practices through systemic reform: The case of Germany after the fall of socialism. *Educational Evaluation and Policy Analysis*, 21(3), 271-296.

- Mitman, A. L., & Lambert, V. (1993). Implementing instructional reform at the middle grades: Case studies of seventeen California schools. *The Elementary School Journal*, 93(5), 495-517.
- Newmann, F., & Associates. (1996). *Authentic achievement: Restructuring schools from intellectual quality*. San Francisco, CA: Jossey-Bass.
- Olsen, B., & Kirtman, L. (2002). Teacher as mediator of school reform: An examination of teacher practice in 36 California restructuring schools. *Teachers College Record*, 104(2), 301-324.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd edition). Newbury Park: Sage Publication.
- Pickett, J. (Ed.). (2001). *American heritage dictionary* (4th ed.). New York: New York: Houghton Mifflin Company.
- Prawat, R. S. (1992). Teachers' beliefs about teaching and learning: A constructivist perspective. *American Journal of Education*, 100(3), 354-395.
- Putnam, R. T., Heaton, R. M., Prawat, R. S., & Remillard, J. (1992). Teaching mathematics for understanding: Discussing case studies of four fifth-grade teachers. *The Elementary School Journal*, 9(2), 213-228.
- Ravitz, J. L., Becker, H. J., & Wong, Y. T. (2000). *Constructivist-compatible beliefs and practice among U.S. teachers*: Center for Research on Information Technology and Organizations.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and educational leadership* (pp. 37-75). Boston: Kluwer Academic Publishers.
- Saxe, G. B., Franke, M. L., Gearhart, M., Howard, S., & Crockett, M. (1997). *Teachers' Shifting Assessment Practices in the Context of Educational Reform in Mathematics*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing CRESST.
- Scheurman, G., & Newmann, F. M. (1998). Authentic intellectual work in social studies: Putting performance before pedagogy. *Social Education*, 62(1), 23-25.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, 54(2), 93-105.

- Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.
- Shepard, L. A., Flexer, R. J., Hierbert, E. H., Marion, S. F., Mayfield, V., & Weston, T. J. (1995). *Effects of introducing classroom performance assessment on student learning*.
- Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher' learning in the context of mathematics instructional reform. *The Elementary School Journal*, 100(4), 351-376.
- Spillane, J. P. (1998). State policy and the non-monolithic nature of the local school district: Organizational and professional considerations. *American Educational Research Journal*, 35(1), 33-63.
- Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: the mediating role of teachers' zones of enactment, learning opportunities. *Journal of Curriculum Studies*, 31(2), 143-175.
- Spillane, J. P. (2000). A fifth-grade teacher' reconstruction of mathematics and literacy teaching: Exploring interaction among identity, learning, and subject matter. *The Elementary School Journal*, 100(4), 307-330.
- Spillane, J. P. (2004). *Standard deviation: How schools misunderstand education policy*. Cambridge, MA: Harvard University Press.
- Spillane, J. P., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record*, 98, 449-481.
- Spillane, J. P., & Zeuli, J. S. (1999). Reform and teaching: Exploring patterns of practice in the context of national and state mathematics reforms. *Educational Evaluation and Policy Analysis*, 21(1), 1-27.
- Spillane, J. P. (2000). A fifth-grade teacher's reconstruction of mathematics and literacy teaching: Exploring interaction among identity, learning, and subject matter. *The Elementary School Journal*, 100(4), 307-330.
- Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). Upper Saddle River: Pearson Prentice Hall.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*: State University of New York Press.

- Sung, T. (2000). Barriers to implementation of performance-based assessment and recommendations. *The Journal of Educational Research*, 38(1), 153-184.
- Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy*, 14(3), 331-356.
- Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, 37(9), 963-980.
- The Korea Institute of Curriculum and Evaluation. (1999). *Guideline for implementation of performance assessment*. Seoul: the Korea Institute of Curriculum and Evaluation.
- The South Korean Ministry of Education. (1997). *The 7th curriculum*. Seoul: The Korean Textbook Publisher.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward Utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring student performance*. San Francisco, CA: Jossey-Bass Publishers.
- Wiggins, G. P. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. P., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Wilson, S. (1990). A conflict of interest: The case of Mark Black. *Educational Evaluation and Policy Analysis*, 12(3), 293-301.
- Wolfe, E. W., & Miller, T. R. (1997). Barriers to the implementation of portfolio assessment in secondary education. *Applied Measurement in Education*, 10(3), 235-251.
- Woodbury, S., & Gess-Newsome, J. (2002). Overcoming the paradox of change without difference: A model of change in the arena of fundamental school reform. *Educational Policy*, 16(5), 763-782.
- Yoo, K. (2001). *Research on improving teachers' performance assessment practices*. Paper presented at the The Korean Society for the Study of Elementary Education, Seoul, South Korea.