

Understanding Elementary Teachers' Different Responses to Reform: The Case of Implementation of an Assessment Reform in South Korea

Jinyoung CHOI ^a*

^a Ewha Womans University, Republic of Korea

Received: 20 September 2016 / Revised: 21 November 2016 / Accepted: 17 January 2017

Abstract

This study explores how teachers implemented an assessment reform in South Korea, with an analysis of different aspects of the reform. Using a mixed method design, this study reveals that the relation between policy and practice depends upon the nature of the changes that reform policies propose. Teachers' implementation varies in terms of different aspects of the reform. Teachers struggle with implementing its fundamental aspects. This study suggests that it is important to take into account variation in implementation as the developmental progress of the reform and attend to scaffolding to make teachers further progress toward fundamental changes in practice.

Keywords: Educational policy and practice, Policy implementation, Assessment, Educational change.

Introduction

Policymakers have become very active in education. It is assumed that if policies compel local actors to put their reform ideas into practice, changes proposed on the policy documents will be manifested in changes in teachers' instructional practices and ultimately student performance will improve. However, results show that there is a chasm between policy makers' intentions and what actually happens in the classroom (Cuban, 2013).

Some researchers have raised questions about whether policies can really alter instructional practice itself. Cohen and Hill (2000) examined the relationship between instructional policy and practice and contended that in order for policies to affect practice, "successful instructional policies are themselves instructional in nature" (p.294). Their study showed that teachers' opportunities to learn about and from policies crucially influence the relationship between policy and practice. A number of other studies found similar results, showing that teachers mediate the relationship between policy and practice (Brian, Reid, & Boyes, 2006; Cohen & Hill, 2000; Joong, Xiong, Li, & Pan, 2009; Spillane & Jennings, 1997).

The question about whether policy can affect instructional practice itself is especially pertinent when the goal of reform policy is to change practice toward higher-order thinking

*  Address for correspondence: Jinyoung Choi, Ewha Womans University, Department of Elementary Education, Education Building A #411, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul, Republic of Korea. Phone: +82 2 3277 4557. E-mail: cjy@ewha.ac.kr

and more demanding content in the classroom. Most contemporary educational reforms call for teaching and learning that promote deep understanding. These reforms have emphasized that it is less important for students to recall and memorize facts and information, and more important for them to create new knowledge by analyzing, evaluating, and integrating information. For example, the mathematics common core state standards require students to demonstrate a deeper conceptual understanding of math (Common Core Math Standards Initiative, n.d.) and the next generation science standards also require students to develop a deep understanding of a smaller number of core ideas, practices, and crosscutting concepts (National Research Council, 2012). Such reforms ask teachers to rethink their beliefs about teaching and learning and teach in ways in which they have never taught before. Thus, policies requiring these kinds of changes are difficult to implement.

Spillane and Zeuli (1999) investigated math teachers' practices in response to national and state reform proposals and found that teachers were responsive to the more superficial aspects of the reforms, but failed to implement the more fundamental aspects. In this study, even though teachers believed that they were implementing the new curriculum, in practice they had different understandings of the new policies and, so, different responses to them. Thus, the relation between policy and practice depends on the nature of the changes that reform policies propose.

This inquiry arises in South Korea because assessment reform in South Korea has asked teachers to adopt new approaches to teaching and to design assessments that meet these new educational goals. Reform efforts to foster students' deep understanding require new kinds of assessments that support this vision of teaching and learning because assessment is an integral part of instruction. Previous studies show that assessment drives instruction and defines what content of the curriculum should be emphasized (Suurtamm & Koch, 2014). In particular, in countries like South Korea, where people are very interested in achievement and outcomes, what and how to assess strongly influence the success of educational reforms. If assessments are not properly aligned with the curriculum and practice in the classroom, it is impossible for educational reforms to succeed.

The need for alternative assessments has inspired South Korean policy as well. Even though the national curriculum emphasizes teaching and assessing higher-order thinking, teachers' instruction and assessments focus on lower-level cognitive demands (e.g., recall of factual knowledge) (Beck, 1998; Choi, 1998; Jung, 2001). In addition, the national government maintained use of its traditional high-stakes multiple-choice tests, which encouraged teaching to the tests. Parents paid attention only to the test scores that 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, widening the gap between the intended curriculum and the enacted curriculum.

In this context, the South Korean Ministry of Education began to focus on the creation of an assessment system that would better evaluate the kinds of higher-order thinking emphasized in the national curriculum. In 1998, the Ministry of Education mandated that all elementary, middle, and high schools use performance assessments. The mandate necessitated individualized, classroom-based performance assessment that was developed locally, rather than standardized large-scale performance assessments. Although the standards and content of performance assessment had to be based on the national curriculum, specific performance assessment tasks and rubrics were to be developed by individual teachers. Similarly, although assessment reform emphasizes performance assessment, traditional paper-and-pencil tests can also be used in classrooms. The assessment reform suggested that teachers change their assessment practices from near-exclusive reliance on traditional tests and incorporate performance assessment.

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 each grade level, teachers had to cover all information presented in the textbook and students had to read and memorize that information. Since the mandated performance assessment reform in South Korea requires teachers to assess higher-order thinking using performance assessment tasks (e.g., essays, reports, and presentations), the Ministry of Education expected that teachers' instructional practices would change such that they begin to focus more on students' ways of thinking (Jung, 2001).

However, the expected changes brought about by mandated assessment reform are unlikely to occur easily and quickly. Evidence shows that changing educational practices is not easy and becomes more difficult when the change involves transformation of the existing structure of schooling (Cuban, 1993; Tyack & Cuban, 1995). Cuban (1993) divides the reforms of the past century into "incremental" and "fundamental" changes. In Cuban's terms, South Korea's new forms of assessment are fundamental reforms because they require new paradigms. Shepard (2000, February) 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 adjusted to 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 assessment policy pushes them to implement assessment reform, and many have struggled to do so. Although the South Korean government has mandated that teachers implement performance assessment reform, teachers' practices may not fulfil the government's hopes. This problem is analogous to a problem that teachers face. Although teachers try to teach the content they want their students to learn, based on the curriculum, it may not be easy for all students to grasp what teachers expected them to learn, especially if learning requires complex thinking. Just as teachers should think about the problems and difficulties their students face, policy makers should start by understanding the problems and challenges teachers faces when they attempt to implement reform.

There are also indications that, regardless of policy context, teachers' responses to the reform mandates will vary depending on their capabilities and willingness to go along with the reform. Importantly, teachers' different views about the nature of knowledge, teaching, and learning strongly influence their responses to reform. Teachers who find that their beliefs are consistent with the reform support the changes proposed by the reform. On the other hand, teachers fail to implement the proposed reform if their old, rooted beliefs conflict with the ideals of the reform. Spillane's (2000) case study provides evidence for a strong link between teachers' beliefs and their enactment of instructional reform. His case shows that there are differences in the extent of change in instructional practices depending on teachers' different views of different subjects (language arts vs. mathematics).

There is also evidence that different teachers respond to a particular reform differently, implementing some aspects of the reform easily but disregarding other aspects of the reform (Spillane & Jennings, 1997; Spillane & Zeuli, 1999). Based on earlier research, we

hypothesized that teachers would implement the assessment reform differently, depending on the nature of the proposed changes. The superficial aspects of the reform would likely be adopted quickly and easily, but its fundamental aspects would not. Building upon earlier research, we explore the relationship between teachers' implementation of different aspects of the reform and their views of teaching, learning and, assessment. Cohen and Ball (2007) state that "Public education is currently the scene of a collision between rapidly rising expectations for school performance on the one hand, and modest capability for the use of innovations on the other (p.33)". Therefore, if the reform necessitates changes that teachers find demanding and challenging, teachers' capacity and will to implement the reform are vital factors as to whether the changes are put in place. If the nature of reform is complex and multidimensional, teachers' implementation should be understood as the developmental progress of the reform influenced by their capacity and will.

Thus, this study aims to investigate teachers' responses to the reform in order to understand the progress of the reform rather than just to see whether central policies can alter teachers' practices. Based on the results, this study attempts to provide policymakers, school administrators, and teacher educators with information about how to help teachers successfully implement assessment reform by examining what actually occurs when teachers attempt to implement the reform. Investigating teachers' practices, as translated from the policy, is important in 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 bring about more fundamental changes in their assessment practices.

Assessing progress of implementation of the reform

An important task in examining teachers' response to reform is to identify what counts as evidence of successful implementation. Conclusions about reform implementation will differ depending on the indicators used to measure implementation (Spillane & Zeuli, 1999). While the number of schools, classrooms, or teachers implementing the reform (breadth) would be one criteria to measure the success of the reform, it is more important to assess the depth of implementation (Coburn, 2003; Hargreaves & Fink, 2000). Coburn (2003) argues that reformers should pay attention to whether their reforms affect deep 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 both breadth and depth of implementation are satisfactory, large-scale reforms can have substantial impacts.

Some researchers documented reform efforts that have failed to bring about fundamental change. For example, Spillane and Jennings (1997) discovered that the state-level reform efforts to align policies were effective in changing surface-level elements of literacy instruction such as materials and grouping arrangements, but less effective in changing deeper elements of literacy instruction such as student tasks and classroom discourse. Spillane and Zeuli (1999) also reported similar findings on teachers' responses to the national and state mathematics reforms. All teachers in this study believed that they implemented the reforms as intended, but most teachers focused on superficial elements of the reforms. Thus, in order to assess teachers' implementation of a reform, we must pay attention to both superficial and deep elements that are put into practice.

Examining both aspects of reform helps we understand teachers' implementation patterns. Teachers' implementation of reform varies depending on the aspects of change intended by the reform. Saxe, Franke, Gearhart, Howard, and Crockett (1999) documented several patterns of change in assessment practices in terms of two aspects of change: format and function. One of the patterns they uncovered was that 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 did not do so in order to evaluate students' mathematical understanding and skills. Another pattern was that teachers "re-purposed" old assessment forms to evaluate new functions. In that study, a teacher used exercises to evaluate the percentage of "right" answers as well as written explanations of those answers. Understanding different implementation patterns is important in examining the progress of implementation. While large-scale reforms require teachers to do the same things, their practices do not change at the same rate; change depends on teachers' capacities and willingness. While some teachers may implement reform on the superficial level, making only the changes that are easy to make, some teachers may implement the reform on the more fundamental level, achieving change that goes beyond structural features.

To investigate how teachers respond to South Korea's assessment reform, I developed three aspects of assessment practices: (1) formats of assessment, (2) cognitive demands of assessment, and (3) purposes of assessment. The first two aspects are based a study by Saxe, Franke, Gearhart, Howard, and Crockett (1999). While formats of assessment, representing the superficial aspects of the reform, would be easier to implement, cognitive demands and purposes of assessment would be more difficult to implement, representing the deeper or fundamental aspects of the reform.

Formats of assessment refers to the particular types of tools teachers use to assess their students. The forms of assessment is typed into two: (1) traditional assessments, including multiple-choice, true/false, matching, and short answer tests, and (2) performance assessments, defined as the observation and rating of students' product that demonstrate their proficiency (Stiggins & Conklin, 1992). While this study focuses on teachers' use of performance assessment, teachers' use of other assessment formats is also examined for comparison purposes.

Performance assessments take various formats. Among these are: (1) extended tasks, such as projects, that require more time, (2) demonstrations that take the form of student presentations, (3) portfolios that collect student work and show development, and (4) teachers' observations that gauge student performance in the classroom. The South Korean assessment reform requires that teachers employ performance assessment tasks that allow students to demonstrate their understanding by constructing responses or by performing tasks, rather than by selecting "right" answers. The features of the assessment reform are well-connected to the national curriculum 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). The performance assessment reform suggests that teachers use a wide range of performance assessment formats.

Cognitive demands of assessment refers to what kind of intellectual work is assessed. For example, a teacher may use an assessment format that evaluates students' factual knowledge, but a teacher wants to assess students' reasoning by using another format. Wiggins and McTighe (1998) indicate that different types of assessments can evaluate different things. The essence of the performance assessment reform is to change what counts as important learning goals. The national curriculum has changed to place more emphasis on what students can do with their knowledge rather than on their ability to acquire a body of factual knowledge. Such a fundamental change in important learning goals prompted changes in ways in which students are assessed to align with the new goals of the national curriculum. The assessment reform suggests that while it is important to assess what students know, but it is also important to assess what students can do with what they know.

Since traditional tests such as paper-and-pencil tests and quizzes often have been used to assess lower-level cognitive demands, such as factual information recall, concepts, and

discrete skills, the performance assessment reform suggests that teachers should also use performance-based assessments to assess higher-level cognitive demands, such as analysis, synthesis and application of knowledge (Jung, 2001; Keon, 2000). The performance assessment reform emphasizes using assessment tasks that allow students to perform, demonstrate, and construct something rather than merely acquire knowledge produced by others. Also, the assessment reform suggests using tasks that ask students to apply real-world problems that they learn about. Since traditional tests have been criticized for their use of decontextualized questions, the assessment reform asks teachers to assess students using tasks that provide students with opportunities to apply what they learn to problems they likely will encounter outside school.

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 to assign grades and provide report cards to keep parents informed (assessment of learning). The second is to use assessment and its results to help teachers teach better and to help students learn more (assessment for learning) (Stiggins, 2005). Since in the traditional assessment often used by South Korean teachers, assessment results are used mainly to assign grades to students at the end of each semester for the purpose of assessment of learning, the performance assessment reform emphasizes the importance of the other purpose of assessment (assessment for learning). The assessment reform suggests that teachers use the results of performance assessment to give feedback to students as well as to design ways to improve learning. Teachers use information obtained from performance assessment to determine what their students know and can do, through directly observing students' demonstrations or products.

Building on this earlier research, I investigate teachers' implementation of the national large-scale assessment reform in terms of different aspects of the reform (superficial vs. deep). Based on a literature review, this study examines South Korean teachers' implementation of the assessment reform and how implementation relates to their capacity for and willingness to change. This study expanded upon previous studies that examined teachers' implementation of large-scale reforms. Teachers' capacity for and willingness to change are considered as both influences on their implementation and indicators of the progress of implementation. Previous studies on implementation have attempted to examine how teachers' capacity for and willingness to change influence implementation. While teachers' capacity for and willingness to change are important factors influencing implementation, I argue that teachers' capacity for and willingness to change should also be taken into account as evidence of successful implementation, along with their actual practices.

Method

Research Design

This is a mixed-method study that includes a broad survey, open-ended interviews, and document analyses. The survey provides information about teachers' assessment practices in general. The case studies provide a more holistic picture of teachers' assessment practices.

Participants

The questionnaire was distributed to elementary teachers who taught social studies in three regions of South Korea near the capital, Seoul. 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 making it difficult to separate their assessment practices for each subject. A total of 700 teachers completed the questionnaire. The majority of survey participants were working in public elementary schools (95.3%) and were female (84.1%). The majority of teachers (81.9%) had earned a baccalaureate degree, and 18.1 % of the

teachers had earned graduate degrees. While 35.0 % of teachers had four years or less of teaching experience, 33.7 % of teachers had taught for 15 years or longer.

Case study participants were selected based on responses to a pilot survey. A stratified sampling method was used to include teachers whose questionnaires suggested different patterns of reform implementation. Based on two dimensions of assessment practices, formats and cognitive demands, I searched for teachers who were in different stages of implementation of the reform: (1) teachers who implemented the reform successfully regarding both aspects of assessment, (2) teachers who implemented the reform successfully in only one aspect of assessment, and (3) teachers who failed 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 cognitive demands on average, the teacher was considered to have implemented the reform successfully in both aspects. If a teacher “never” or “once a semester” used performance assessment formats and gave “little” emphasis to higher-level cognitive demands, the teacher was considered as having “failed to implement in both aspects”.

Four teachers were selected for the case study. All teachers in the case study had more than 10 years of teaching experience. This sample includes two female teachers who had earned only baccalaureate degrees and two male teachers who had earned a master’s degree or both a master’s and a doctorate degree. One teacher taught in a private elementary school and the other three teachers taught in public elementary schools.

Data Collection and Analysis: Survey data for the quantitative study

Written survey responses from teachers were analysed quantitatively. Based on a literature review, a survey was developed to examine teachers’ implementation of the three assessment aspects. Assessment formats were categorized into two types. One type is the traditional format of assessment that includes paper-and-pencil tests consisting of selected responses to questions such as multiple-choice, true/false, matching, fill-in-the-blank and/or completion and short-constructed response tasks. Another type is the new formats of assessment, which includes extended-response tasks, portfolios, demonstrations, and observations. Teachers reported how frequently each type of format was used in their social studies classrooms. Next, cognitive demands were categorized into two types. One type is lower-level cognitive demands that ask students to recall factual knowledge, find answers in social studies textbooks, or find only one right answer/solution. The other is higher-level cognitive demands that ask students to consider multiple interpretations/perspectives, understand relations among central concepts, represent their responses in various ways, apply concepts to a real-world problem, make arguments with evidence, or explain what they solve. Third, purposes of assessment were categorized into two types. One type is the assessment of learning, in which teachers use assessment to assign grades and provide parents with report cards. Another type is the assessment for learning, in which teachers use assessment to improve their teaching methods and the amount that students learn.

The survey data were analysed in two steps. First, factor analysis were conducted to validate the implementation measure. The results of factor analysis showed that the factor structure for each aspect included two factors, as suggested above. Second, internal consistency reliability of the implementation measure was calculated. Cronbach’s alpha ranged from 0.65 to 0.85, indicating that the implementation measure was reliable. Finally, descriptive analyses and paired t tests were used to examine whether there are mean differences of implementation between the two types of each aspect of assessment.

Data Collection and Analysis: Data sources for the qualitative case study

Transcripts of interviews with teachers and other documents were collected for the qualitative case study. Interviews were conducted; each interview lasted approximately 90-120 minutes and was conducted after school in the teacher's classroom. A semi-structured interview protocol was used to ensure consistent coverage of important themes, but each interview was allowed to take its own direction once it started. Participants in the case study were asked to identify and provide documents relevant to their assessment practices. These documents include school curriculum frameworks, school and grade assessment plans, examples of assessment tasks, etc. The assessment tasks actually used in classrooms were the most important indicators of how teachers implemented performance assessment reform.

The case studies involved analyses of interview data and documents. The interview data were analysed via the following steps. First, I read interview 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 that emerged from the reading of the data. A second step was to code the content of transcripts in relationship to these themes. A qualitative computer software program 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 differences between them (across-case analysis).

Teachers' Reports on Their Implementation of the Reform

This section presents teachers' survey responses regarding three aspects of the assessment reform: formats, cognitive demands, and purposes of assessment.

Formats of Assessment

Teachers reported how frequently each type of assessment format was used in their social studies classroom. Table 1 presents teachers implementation in terms of assessment formats.

Table 1. Teachers' implementation in terms of assessment formats

Assessment format	Mean	sd	t	df
Traditional assessment	2.35	1.15	1.46	679
Performance assessment	2.43	.75		

On average, teachers used both types of assessment formats once a semester or once every other month. The mean use of new assessment formats was higher than that of traditional assessment formats, but the difference was not significant ($t= -1.46, p>.05$). This result indicates that teachers' assessment practices are shifting away from over-reliance on traditional formats.

Cognitive Demands of Assessment

Teachers reported how frequently they assessed different types of cognitive demands. Table 2 presents teachers' implementation in terms of cognitive demands.

Table 2. Teachers' implementation in terms of cognitive demands of assessment

Cognitive demands	Mean	sd	t	df
Lower-level	3.00	.93	10.97***	652
Higher-level	2.58	.75		

Note. *** = $p < .001$.

The most frequently-assessed cognitive demand was finding answers from social studies textbooks and the second was recalling factual knowledge. Lower-level demands were significantly more frequent ($t=-10.97, p<.05$). This result indicates that many teachers still focus on assessment of lower-level cognitive demands than on assessment of higher-level cognitive demands.

Purposes of Assessment

Teachers reported how much emphasis was given to the different purposes of assessment. Table 3 shows teachers' implementation in terms of assessment purposes.

Table 3. *Teachers' implementation in terms of assessment purposes*

Assessment purposes	Mean	sd	t	df
Assessment of learning	3.67	.93	7.84***	681
Assessment for learning	3.23	.80		

Note. *** = $p < .001$.

The strongest emphasis was on grading and the weakest emphasis was on improving teaching. When compared to the new purposes of assessment, the mean of traditional purposes was higher. A paired t-test found a significant mean difference between emphasis on the two different purposes ($t= 7.84, p< .05$). This result indicates that many teachers still focus more on assessment of learning rather than assessment for learning.

Four Cases of Teachers' Responses to the Reform

The survey responses suggest that teachers did not fully embrace the reform. Although the reform pushed teachers to use more new assessment formats, it was not successful in eliciting more fundamental changes because the reform did not make teachers change more critical aspects of assessment practices (cognitive demands and purposes of assessment). The survey responses show teachers' implementation of each aspect of assessment practices separately. The case studies allow us to see how teachers combine these different aspects of reform to design their own practices. In this section we describe four cases of implementation that represent different patterns with respect to: (1) formats of assessment, (2) cognitive demands of assessment, and (3) purposes of assessment.

Four implementation patterns were identified based on teachers' views and intentions concerning teaching and learning and teachers' practices in the three aspects of assessment. Figure 1 shows the different patterns of implementation of the reform. The vertical axis in Figure 1 shows teachers' actual practices in response to the reform according to implementation of the three aspects of assessment. The high end is Authentic Implementation and the low end is Symbolic Implementation (that is, the assessment addressed only the letter of the law). The horizontal axis in Figure 1 shows teachers' views on instruction and assessment. The right edge indicates more reform-oriented views, and the left edge indicates more traditional views.

Four teachers who serve as good examples were located in Figure 1 by examining both their views and their practices. Mr. Choi is highest in the figure and furthest to the right, exemplifying Profound Implementation. Mr. Kim, near the center of Figure 1, exemplified Transitional Implementation. Mr. Kim's views on instruction and assessment seemed close to the reform's intention, but he mixed traditional and reform-oriented assessment practices. The other two teachers, Ms. Lee and Ms. Park, seemed to be engaging in symbolic implementation. Ms. Lee intended to implement the principles of the reform, but could only integrate the formats, not the other aspects of assessment. Ms. Park held very traditional

views about instruction, and merely complied with implementation of reform as she has been ordered.

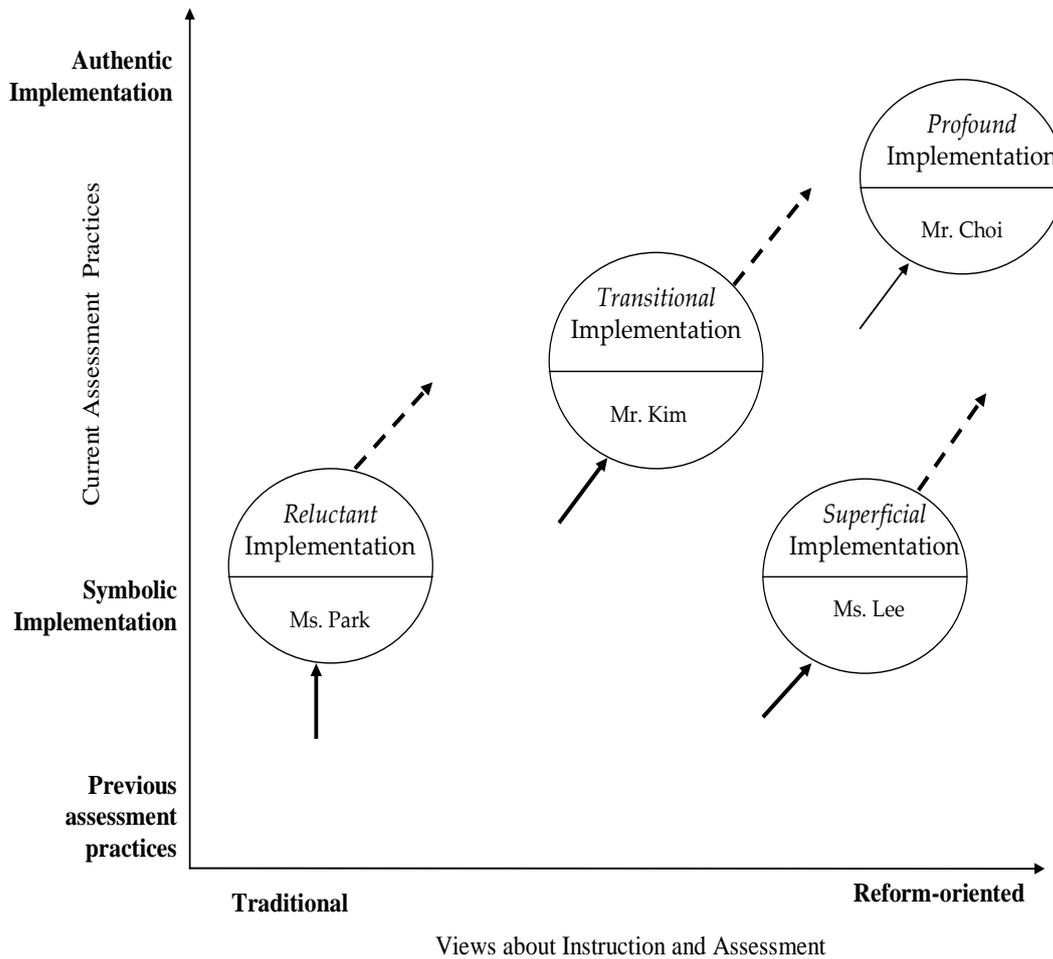


Figure 1. Variation in implementation of the assessment reform

Pattern 1: Profound Implementation

Mr. Choi used a variety of performance assessment formats such as essays, projects, portfolios, demonstrations, and observations. He most often used written reports as an assessment format. He provided students with topics for their projects, but students choose their own research questions. After researching their questions, students wrote reports. If the projects involved field trips, he 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 to comply with the sixth grade assessment plans developed by his fellow teachers, but he didn't like this type of assessment format because he thought the tests were not aligned with what he taught to his students. He used the textbook as a resource and his lessons were often reconstructed, but the tests mapped too closely to the information in the textbook. So, he did not think that these tests accurately illustrated what students learned in his 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. He gave more weight to scores from

performance assessment tasks when describing students' knowledge on their report cards. For example, if a student was good at writing reports, but did not obtain a good score on a traditional test, he did not give a bad grade in Knowledge when writing her report card.

Mr. Choi's students struggled with the new assessments. They were accustomed to memorizing historical facts represented in social studies textbooks. Since the students were unfamiliar with the new assessment formats, he gave them his assessment criteria before the students did assessment tasks. For example, before assigning writing journals, he told them he wanted to see them support their arguments with evidence and by suggesting alternative solutions to a given problem. Based on these criteria, he gave feedback to the students on the journals, and the students revised their journals or tried to reflect his comments when they wrote another journal entry. By articulating his expectations before assessing student work, he prepared students for his alternative assessments.

Mr. Choi assessed both content knowledge and higher-order thinking. He believed that knowledge and thinking skills such as inquiry and decision making should not be separate and that new formats of assessment would be more beneficial in illustrating how students solved problems using what they know. This view on assessment was reflected in the assessment tasks Mr. Choi used. 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 were 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 support his or her arguments with appropriate evidence? The first two criteria were meant to assess students' understanding of the content and the third was to assess students' reasoning skills.

Mr. Choi's assessment practices were connected to his teaching goals. He emphasized understanding of multiple historical interpretations and tried to get students to learn the importance of interpretation and to form their own perspectives. For example, in a historical unit on the Japanese colonization of Korea, each student was asked to compose an imaginary letter to a Japanese elementary school 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. These assessment tasks indicated that he wanted his students to learn something beyond factual knowledge, and, in order to assess more than mere recall of factual knowledge, he used new formats of assessment rather than traditional paper-and-pencil tests.

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

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 to students and parents. The most frequently used way was to provide direct comments on students' essays or reports as well as a one-page description of the results of assessment for parents. The one-page description detailed what the student did well and what the student needed to improve. To allow parents

to understand their child's position in 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.

Pattern 2: Transitional Implementation

Mr. Kim had views on instruction and assessment that were close to reform-oriented, but he mixed traditional and reform-oriented assessments. While he adopted performance assessment formats, he was more uneven in his commitment to practice the other two aspects of assessment (cognitive demands and purposes of assessment). He advocated a balanced approach between traditional and new formats of assessment. Discussion of a typical social studies class will demonstrate how he balanced the use of both traditional and new assessment formats.

Mr. Kim used project-based instruction to teach a history unit. Before starting the unit, he provided students with the topics or issues related to lessons in the unit. Students formed teams and team members met to complete their project after class or discussed the project via Instant Messenger. Every lesson started with one team's presentation. The team presented what they had learned during their project and had to answer the teacher's and other students' questions. 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 shows 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 other type of assessment format that Mr. Kim used was mid-term and final paper-and-pencil tests. Most questions were multiple-choice, while some required extended responses. In contrast to Mr. Choi, Mr. Kim believed that traditional assessment formats could assess basic, applied, and advanced questions. The basic questions consisted of multiple-choice items with one correct answer. Applied questions assessed inquiry skills, such as interpreting maps, and advanced questions assessed students' thinking and reasoning skills, such as comparing two arguments and taking and supporting one of the positions in writing. He mentioned that because he included these three types of questions his traditional paper-and-pencil tests assessed thinking skills.

Even if Mr. Kim could design 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 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 that was related to a particular unit. Before going on the field trip, Mr. Kim provided students with a booklet of questions for them to answer upon their return, based on their observations of the site (a palace) and the information they 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.

Mr. Kim's assessment practices mixed his purposes of assessment. First, the purpose of assessment that he considered the most important 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. 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. He 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 he scored the test reflected another of his assessment purposes: considering students' learning progress. Each score was calculated by subtracting the previous test score from the current test score. A progress score for each student was derived based on how many of his/her scores had improved. Mr. Kim expressed that assessing in this way was helpful in motivating all students to study hard to earn higher test scores.

Pattern 3: Superficial Implementation

Ms. Lee held reform-minded views of instruction and assessment, but could only implement the reform on a superficial level. Her assessment practices showed a shift from over-reliance on traditional paper-and-pencil tests to more employment of new formats of assessment. She did not use traditional paper-and-pencil tests because she did not want to test mere recall of knowledge. She thought that paper-and-pencil tests decreased students' interest in social studies learning by requiring recall of knowledge. The most often used new format was report-based inquiry projects that took at least one week. She chose the topics or problems for inquiry projects, but students could choose specifically what they wanted to investigate. She assessed students' reports and presentations. Another new format she used was tests with open-ended questions. These tests were given approximately ten minutes before the completion of a lesson. After students had given the presentation, she distributed a task to all students. The students answered the questions based on what they learned in the presentation or from their social studies textbooks.

But Ms. Lee 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, but these assessment tasks asked mainly for factual knowledge from the social studies textbooks. 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 find the answers in the textbook. That is, although she wanted to assess understanding of big ideas and not memorization, what she asked students to do was to find the correct answers in the social studies textbook. Similarly, some of her assessment tasks asked students to collect information via the Internet and summarize 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 how hard students worked and less on how they thought. If a student wrote a lot of information about the topic, she would say that the student knew the content well.

Ms. Lee also did not use 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. She mentioned that "I feel that I am doing assessment for assessment's sake." However, she exhibited dissatisfaction with her current use of assessment. She mentioned that, "I ask myself if the assessment [of learning] I am now doing is really necessary." She thought that in order to improve student learning she needed to provide parents with more detailed information about an individual student's progress.

Pattern 4: Reluctant Implementation

Ms. Park was similar to Ms. Lee in that she only implemented the superficial elements of the reform (formats of assessment), but she differed from Ms. Lee and Messrs. Kim, and Choi in that she was not dissatisfied with traditional assessments. Her traditional views were aligned

with her assessment practices (e.g., she relied heavily on traditional formats that assessed recall of knowledge).

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. Her heavy use of traditional paper-and-pencil tests appeared to be related to her beliefs about assessment. She believed that the purpose of assessment was to check whether students found correct answers to questions in social studies textbooks. She expressed difficulties when she used new formats of assessment because of the new curriculum. She thought that because assessment in the new curriculum required teachers to use open-ended questions that did not have right answers, the curriculum caused a problem. She said: "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."

Ms. Park's use of new formats of assessment symbolically complied with the assessment reform, since all elementary school teachers were 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." Even though she used the portfolio as a new assessment format, her assessments focused on finding and remembering correct factual answers. She asked students to collect everything they did including unit tests and file them in a scrapbook. Also, she told students, "The midterm and final tests will be based on the questions I gave in the unit tests. Try to look at your wrong answers and remember the right answers."

The major focus of assessment was on recalling facts. She believed that students should memorize many facts when they learned 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, she used many paper-and-pencil tests. Ms. Park's reason for using performance assessment formats was because the performance assessment tasks seemed interesting to students. For example, she asked students to design an advertisement for Kimchi (a traditional Korean food). She mentioned that students had fun doing this, but she did not have any assessment targets related to the unit.

She seemed to believe that students who received good scores on paper-and-pencil tests could apply their knowledge to a real-world problem. She explained how she assessed thinking skills based on paper-and-pencil tests. For example, she chose to use the information obtained from paper-and-pencil tests to assess whether students could draw a table of chronological events. If students answered relevant questions correctly, she assumed that they could draw a chronological table.

Ms. Park had two main purposes for assessment. One purpose was 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 acquired knowledge of the content of the course. She said that test scores from traditional paper-and-pencil tests were the main sources she consulted when writing report cards because tests scores showed whether students knew the right answers or drew the right conclusions.

Discussion and Conclusion

The case study show how teachers' responses to the reform varied. While all teachers implemented the performance assessment reform as they were ordered, their implementation did not progress at the same speed. 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. The case study shows 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, four patterns were identified. The questionnaire data, which is based on a larger sample size, support the finding that teachers struggled to implement the performance assessment reform on a fundamental level. While many teachers integrated superficial elements of the reform (assessment formats) without difficulty, they were less successful in implementing the deep aspects of the reform (cognitive demands and purposes of assessment).

Three noticeable findings about policy and practice emerged from my analyses of the case study and questionnaire data. First, teachers responded to the reform in different ways. More importantly, I found that, while teachers easily implemented superficial 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' responses to policy are mediated by their own knowledge and values, so their interpretation of reform ideals is 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 superficial level, we cannot say that these two teachers made 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 on a fundamental 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 the competence necessary for implementing the reform, which is not the case for teachers like Ms. Park. Since teachers like Ms. Park have no interest in implementing performance assessment reforms, it is more difficult for them to change their assessment practices. Thus, this study emphasizes the importance of having both an understanding of and a commitment to rationales for reform, and that both serve as indicators of the progress of the reform.

The examination of teachers' practices from both the cases and the questionnaire allowed us to understand that teachers do not implement reform at the same speed because they do not come from the same starting point. While some teachers can implement both superficial and deep aspects of the reform, some teachers can implement only the superficial aspects of the reform. We should not celebrate the success of the reform only by looking at the implementation of its superficial aspects. Also, we need not grieve the failure of the reform because teachers cannot implement the reform on a 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.

The findings from this study allow us to ask what teachers need to bring about fundamental changes in their assessment practices, given their current status. Merely mandating reform cannot increase or change teachers' capabilities and willingness; as McLaughlin (1987) noted, policy "cannot mandate what matters" (p.173). Instead of merely mandating reform, we need

to supplement the mandate with other instruments. What we are missing is a better understanding of how to change teachers' implementation of the reform from superficial compliance to substantial changes in practice. I will call this a "scaffolding instrument" that operates as support for teachers' learning and changes in practice. 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, teachers need scaffolding to move to the next step toward profound implementation. In Figure 1, black arrows could indicate the effect of the mandate and dotted arrows could indicate imaginary effects of scaffolding if teachers receive appropriate scaffolding based on their current points. Diagnosing how teachers implemented the reform is a way to create the right kind of scaffolding that will help teachers make progress in their implementation of the reform.

Looking toward future reform efforts, an important question is what kind of scaffolding we need to provide for teachers to be able to bring about 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 on a deeper level? This study suggests that providing educators with learning opportunities is the best scaffolding instrument for changing teachers' practices by increasing their capabilities and willingness.

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. According to McLaughlin and Mitra (2001), teachers need knowledge of a reform's first principle. Without knowledge about why teachers should implement the reform, they will not bring about fundamental changes. In this study, to successfully implement the assessment reform, teachers should understand why and how to implement it. However, many reforms in South Korea were mandated without giving sufficient learning opportunities for teachers to understand and enact the first principles that underlie the reform, and these reform efforts have been unsuccessful. Thus, to move from policy changes toward real changes in practice, teachers should have sufficient and high-quality learning opportunities. Learning opportunities can promote teachers' capabilities and willingness to implement reforms, as well as help teachers change their practices on a fundamental level. The case of Ms. Lee illustrates the importance of helping teachers learn about performance assessment. Even if her beliefs were in alignment with the principles of the reform, she lacked the capacity to implement performance assessment reform. It is necessary for her to learn about performance assessment. She must learn how to design and score performance assessment tasks to assess what she intended to assess (higher-order thinking). Ms. Lee also needs to learn what and how to teach in her social studies classes; since assessment cannot be separated from instruction, both instruction and assessment must be changed together.

For teachers like Ms. Park, simply providing more resources or more instruction would not enable them to fully implement the reform. Without understanding the rationale behind the performance assessment reform, teachers like Ms. Park will not make much effort to implement the reform. Teachers like Ms. Park may need to understand the ways in which they are expected to provide for students' learning in order to realize the importance of changes to their practices.

This study shows how difficult it is to change teachers' practices. Even if a policy mandates a reform, many teachers only implement the reform on a superficial level. To reduce the discrepancies between policy (intended reform) and practice (enacted reform), policymakers should consider adding scaffolding instruments to complement the mandate of the reform. This study suggests that the availability of high-quality learning opportunities is the best scaffolding instrument to support teachers. While teacher learning can play a critical role in connecting policy and practice, we should not expect that teachers' practices will change within a short period. We must wait until teachers learn new ideas and implement those ideas

through trial and error, providing teachers with the right kinds of support during implementation. Fullan (1993) said that, “problems are our friends.” Starting by understanding the problems teachers face, policymakers have to plan what they need to do to help teachers move toward fundamental change. When teachers feel comfortable facing problems, and searching for and applying possible solutions to the problems they have, they can make progress in changing their practices.



References

- Beck, S. (1998). *Theory and practice of performance assessment*. Seoul: Wonmisa.
- Brian, K., Reid, I., & Boyes, L. C. (2006). Teachers as mediators between educational policy and practice. *Educational Studies*, 32(4), 411-423, doi:10.1080/03055690600850396
- Choi, H. (1998). A study on the way of aligning performance assessment with school curriculum. *The Journal of Curriculum Studies*, 17(1), 45-67.
- Coburn, C. E. (2003). Rethinking scale: Moving beyond numbers to deep and lasting change. *Educational Researcher*, 32(6), 3-12.
- Cohen, D. K., & Ball, D. (2007). Educational innovation and the problem of scale. In B. Schneider & S-K. McDonald (Eds.). *Scale up in principle: Ideas in principle of scale, Volume 1* (pp.19-36). Lanham, MD: Rowman and Littlefield.
- Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: The Mathematics reform in California. *Teachers College Record*, 102(2), 294-343.
- Common Core Math Standards Initiative. (n.d.). *Common core state standards for mathematics*. Retrieved from http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf
- Cuban, L. (1993). *How Teachers Taught: Constancy and change in American classrooms 1890-1990*. New York, NY: Teachers College Press.
- Cuban, L. (2013). Why so many structural changes in schools and so little reform in teaching practice? *Journal of Educational Administration*, 51(2), 109-125. doi:10.1108/095782313113 04661
- Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. Levittown, PA: The Falmer Press.
- Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. New York, NY: Teachers College Press.
- Hargreaves, A., & Fink, D. (2000). The three dimensions of reform. *Educational Leadership*, 57(7), 30-34.
- Joong, P., Xiong, Y., Li, L., & Pan, C. J. (2009). Investigation into the perceptions of students, parents, and teachers in China's education reform in grades 7 and 8. *International Electronic Journal of Elementary Education*, 1(3), 141-254.
- Jung, M. (2001). *Performance-based assessment in social studies*. Seoul: Hakmunsa.
- Keon, D. (2000). Problems of performance assessment. *The Journal of Education Research*, 21(2), 83-102.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.

- McLaughlin, M. W., & Mitra, D. (2001). Theory-based change and change-based theory: Going deeper, going broader. *Journal of Educational Change, 2*(4), 301-323.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- 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.
- Saxe, G. B., Gearhart, M., Franke, M. L., Howard, S., & Crockett, M. (1999). Teachers' shifting assessment practices in the context of educational reform in mathematics. *Teaching and Teacher Education, 15*(1), 85-105.
- Shepard, L. A. (2000, February). *The role of classroom assessment in teaching and learning* (CSE Technical Report 517). Retrieved from <https://www.cse.ucla.edu/products/reports/TECH517.pdf>
- 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., & Jennings, N. E. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *Teachers College Record, 98*(3), 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.
- Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Suurtamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: experiences of mathematics teachers in Ontario, Canada. *Educational Assessment, Evaluation and Accountability, 26*(3), 263-287. doi:10.1007/s11092-014-9195-0
- The South Korean Ministry of Education. (1997). *The 7th elementary school curriculum commentary*. Seoul: Daehan Gyogwaseo.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.
- Wiggins, G. P., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.