THE INFLUENCE OF STANDARDS ON K–12 TEACHING AND STUDENT LEARNING: A RESEARCH SYNTHESIS

Regional Educational Laboratory
Contract #ED-01-CO-0006
Deliverable #2005-10

Prepared by

Patricia A. Lauer
David Snow
Mya Martin-Glenn
Rebecca J. Van Buhler
Kirsten Stoutemyer
Ravay Snow-Renner

August 19, 2005

ACKNOWLEDGEMENTS

The authors wish to thank several individuals at McREL who contributed to this research synthesis. Thanks go to Zoe Barley who provided support and advice and to John Kendall and Lou Cicchinelli for their helpful review comments. We appreciate the efforts of Terry Young who obtained the research articles. We acknowledge Barbara Aiduk, Kirsten Miller, and Robyn Alsop for their editing and logistical support in preparing the document. Finally, we thank our external reviewers for their invaluable input. These were Robert Floden of Michigan State University and Sandra Wilson of Vanderbilt University.

This document has been funded at least in part with federal funds from the U.S. Department of Education under contract number ED-01-CO-0006. The content of this publication does not necessarily reflect the views or policies of the Department of Education nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. Government.

TABLE OF CONTENTS

PREFACE

The No Child Left Behind (NCLB) Act of 2001 focuses on the responsibility of schools and districts to help all children achieve at high levels. NCLB requires states to establish standards for student learning and to measure students' progress through assessments that are aligned with these standards. The assumption is that standards-based education results in improved teaching and student learning. Now that standards for K–12 education have been in existence for over a decade, is there evidence that this assumption is correct?

This report is the fourth research synthesis that Mid-continent Research for Education and Learning (McREL), the Regional Educational Laboratory for the Central Region states, has conducted in its laboratory leadership area of standards-based educational practice. In 2001, McREL completed and published a synthesis of research on standards-based classrooms (Apthorp, Dean, Florian, Lauer, Reichardt, & Snow-Renner). That report used narrative reviews to examine research on standards-based instruction in literacy and mathematics and on the practices and policies needed for professional development and school organizations in a standards-based education system. In 2002, McREL conducted a research synthesis on the effectiveness of strategies designed to assist low-achieving or at-risk students during the school day so that all students can ultimately achieve standards (Barley et al.). The 2002 synthesis provided reviews of research on six classroom strategies: general instruction, cognitively oriented instruction, grouping structures, tutoring, peer tutoring, and computer-assisted instruction. In 2003, McREL synthesized the research on strategies to assist low-achieving students in reading and mathematics outside the school day, such as after school and summer school (Lauer, Akimbo, Wilkerson, Apthorp, Snow, & Martin-Glenn).

The 2003 synthesis used meta-analytic techniques to analyze the research and reached the overall conclusion that out-of-school time can have small but positive impacts on the achievement of low-achieving or at-risk students. In 2004–2005, McREL's synthesis addresses the broad research problem: What is the influence of standards on K–12 teaching and student learning?

The goals for the current research synthesis are:

1. To identify a comprehensive collection of studies on standards-based education, gathered through a structured search process and screened for relevance and research quality

2. To describe and analyze these studies through a systematic review that examines teacher and student outcomes in relation to the variables of standards-based curriculum, standards-based instructional guidelines, and standards-based assessment

3. To describe the implications of study findings for researchers, education administrators, and policymakers

This synthesis is organized into six chapters. Chapter 1 describes the research problem and conceptual framework, provides background information on standards-based education, and discusses prior research. Chapter 2 describes the methods used to search the literature, code the studies, and synthesize results. Chapters 3, 4, and 5 review research on the influences of standards-based curriculum, standards-based instructional guidelines, and standards-based assessment, respectively. Chapter 6 summarizes the findings across the studies of the three variables and provides general conclusions. Appendices include additional details on the literature searches, the instrument used to code the studies, and tables of the reviewed studies. The tables describe in detail the characteristics, methods, and findings of the studies and serve as an annotated bibliography.

The authors of this document worked as a team to conduct the synthesis and produce the report, making individual contributions based on their areas of expertise. Patricia Lauer authored Chapters 1, 2, and 6 and led the synthesis team. Rebecca Van Buhler, Mya Martin-Glenn, and Kirsten Stoutemyer wrote Chapter 3. Ravay Snow-Renner wrote Chapter 4, and David Snow wrote Chapter 5. Mya Martin-Glenn also directed the search for and documentation of synthesis research studies, and Ravay Snow-Renner also assisted with editing.

The primary audience for this report is education researchers and state education administrators who have a general understanding of research-based evidence. The secondary audience includes policymakers and district and school administrators who have some background in research. Although this document is not written for a practitioner audience, it does describe implications of research findings for practice and policy.

EXECUTIVE SUMMARY

The No Child Left Behind (NCLB) Act of 2001 focuses on the responsibility of schools and districts to help all children achieve at high levels. NCLB requires states to establish standards for student learning and to measure students' progress through assessments that are aligned with these standards. The assumption is that standards-based education can result in improved teaching and student learning. Now that standards policies for K–12 education have been in existence for over a decade, is there evidence that this assumption is correct?

Many research and evaluation studies have addressed different aspects of K–12 standards and standards-related policies. However, only a few systematic reviews have examined evidence that the implementation of standards improves education outcomes. In general, there has been little coherence in the research and evaluation of the influence of standards. Some researchers contend that evaluations of standards-based education should consider the entire education system. Other researchers support designs that focus on the classroom, maintaining that if system changes do not reach the classroom in the form of changes in practice then the benefits of standards-based education will not be realized.

While acknowledging that system variables are important, this research synthesis focuses on the three variables most closely related to teaching and student learning. It examines research studies that address one or more of the following questions:

1. What is the influence of standards-based curriculum on teacher instruction and student achievement?

2. What is the influence of standards-based instructional guidelines on teacher instruction and student achievement?

3. What is the influence of standards-based assessment on teacher instruction and student achievement?

After conducting systematic searches of the education research literature related to standards-based education, researchers read 697 study reports published in the U.S. since January 1995. To be included in the review, studies had to involve K–12 students; concern a core subject area; assess or document teacher instruction or student achievement; have explicit connections to national, state, or local standards; and address standards-based curriculum, instructional guidelines, or assessment. Quantitative studies had to describe instrument development, report a response rate for surveys, describe the

processes used for data collection and analysis, and include sufficient evidence of the results. Qualitative studies had to describe the processes used for data collection, describe the methods used for data analysis, include sufficient evidence of the results, and use a process to validate the results, such as a search for disconfirming evidence.

One hundred and thirteen studies met these inclusion criteria. Each study was coded with regard to methodology, variables addressed, population and sample, results, and implications for education practice and policy. Included were 48 studies of standards-based curriculum, 36 studies of standards-based instructional guidelines, and 29 studies of standards-based assessment. Of the total, 71 studies examined influences on teacher instruction, 56 examined influences on student achievement, and some studies measured both. The majority of the studies on curriculum and instructional guidelines addressed national standards, while almost all the studies of assessment concerned state standards (as represented on mandated state assessments). Across the three variables, both elementary and secondary education levels were represented.

Because the great majority of the reviewed studies used non-experimental methods without control or comparison groups, study results were analyzed through a systematic narrative review, that is, findings were summarized and discussed based on study variables, outcomes measured, and research method. Only two studies used an experimental research design, 17 studies employed quasi-experimental designs, and an additional four studies using mixed methods included a quasi-experimental design. There were 34 studies that were quantitative but non-experimental, 33 qualitative studies, and 27 studies that used mixed methods in which both quantitative and qualitative data were collected.

The reviewed studies of standards-based curricula found predominantly positive influences on student achievement, including that of at-risk students. The majority of these studies used quantitative quasi-experimental designs, so there is reason to be confident in the overall finding. Possible modifiers of this effect are time and the type of measure. Several studies indicated that student achievement improved more with longer exposures to the standards-based curriculum. Some studies suggested that a disparity between the achievement measure and the curriculum goals influenced the student achievement outcome. However, other studies provided evidence that students can demonstrate the knowledge gained from a standards-based curriculum on traditional standardized tests. The results were more mixed for the influence of standards-based curricula on teacher instruction, although in most of the studies, teachers were changing their instruction to reflect the standards. A possible modifier of this effect was the extent to which systemic supports were in place and aligned with the curriculum. Another possible modifier was teachers' perceived lack of time for preparation of the instruction required for a standards-based curriculum.

The reviewed studies of standards-based instructional guidelines found evidence for a weak positive influence of standards-based, reform-oriented, instructional practices on student achievement. The main moderator of this finding was the nature of the achievement measure. For example, in studies of mathematics instruction, achievement tests that emphasized higher-order skills were more likely to show a positive relationship with instruction recommended by the National Council of Teachers of Mathematics than tests that focused primarily on important basic skills such as computation. Another possible moderator identified in some studies was differential student access, suggesting that at-risk students may experience more traditional forms of instruction than other students. Regarding the influences of standards-based instructional guidelines on teacher instruction, some studies found that teacher instruction changed to be more consistent with the overall intent of standards policies, while other studies did not document such changes. An important modifier identified was the measure of instructional practices. On surveys, teachers reported being knowledgeable about standards and using standards-based practices in their classrooms, but observations of their instruction indicated otherwise.

All but three of the studies of standards-based assessment addressed influences on teacher instruction, and the six studies that examined student achievement were inconsistent in their findings. Based on descriptive evidence from teacher surveys, teacher interviews, and classroom observations, standards-based assessments have strong influences on teacher instruction. Across the studies there was evidence that teachers changed what they were teaching in the classroom to align with state assessments. In some studies, secondary teachers were broadening their content coverage to include more topics. Other studies reported that elementary and secondary teachers were narrowing their content coverage to the exclusion of non-tested subjects. In addition, teachers reported using test-preparation sessions to simulate standards-based assessments.

The results of this synthesis lead to several conclusions and implications for practice, policy, and research related to standards-based education:

- Standards-based curricula and standards-based instructional guidelines can have positive influences on student achievement. To facilitate gains in student learning, educators should implement standards-based curricula as intended, and administrators should sustain teachers' and students' exposure to such curricula.

- Standards-based curricula and standards-based instructional guidelines can influence teachers to adopt reform-oriented instructional practices. To influence instruction, educators need to translate content standards into curricula that are aligned with instructional materials and student assessments, and teachers need opportunities for professional development.

- Standards-based state assessments influence both the content and pedagogy of classroom instruction. Depending on the assessment, teachers may broaden or narrow their content coverage, use more or fewer reform-oriented instructional practices, and emphasize more or less test preparation in their instruction. State education administrators should judge the quality of standards-based assessments in terms of their potential to induce teachers to make favorable curricular and pedagogical decisions.

- At-risk students may experience less access to reform-oriented instruction than more advantaged students. At-risk students can benefit from reform-oriented instruction, but administrators and policymakers need to find ways to make instruction equitable among diverse groups of students.

- Results from studies of standards-based education depend on how the outcomes are measured. Teachers tend to overestimate on surveys their use of standards-based instructional practices compared to classroom observations of their instruction. Disparity between a student achievement measure and the goals of a standards-based curriculum can negatively influence student scores. Researchers and evaluators should consider the influence and limitations of the measures on outcomes, and so should those who are using the research.

- The breadth and quality of research on standards-based education needs to improve. Now that America is entrenched in standards-based reform, the research should address not only the question of "does this work?" but also "how can we make it work it better?" More studies are needed that use rigorous methodology with respect to the question being asked. More studies of language arts and social studies are needed to inform the discourse about effective standards-based instruction in these subject areas.

The overall conclusion is that standards-based policies influence teaching and student learning in K–12 classrooms, but the nature of these influences depends on how standards-based policies are perceived and implemented by teachers. According to Resnick and Zurawsky (2005), if standards-based reforms are to achieve their promise of helping all students meet high standards, then more attention and resources are needed for the instructional support system in schools, including curriculum, instruction, professional development, and interventions for struggling students. The results of this synthesis support this observation and suggest that the next step in improving standards-based education is to help teachers in their efforts to implement standards policies in their classrooms.

CHAPTER 1: INTRODUCTION

The language of standards, assessment, and accountability currently dominates the education scene. According to Education Week's *Quality Counts* report for 2005, 48 states and the District of Columbia have established content standards in the four core subject areas of mathematics, language arts, science, and social studies. And all 50 states administered student assessments linked to state or local standards (Education Week, 2005). Therefore, most educators would agree that standards are a defining characteristic of American education and are likely to be so for the next several years. For this reason, it is important and timely to examine the influences of standards-based education on teaching and student learning. This chapter describes the background, conceptual framework, and research context of this synthesis.

BACKGROUND

In their compendium of K–12 standards, Kendall and Marzano (2000) outline the history of the standards movement. They cite the 1983 publication of *A Nation at Risk* (National Commission on Excellence in Education) as the catalyst for events that culminated in a national call for content standards (National Education Goals Panel, 1991). Such standards were to reflect high expectations for all students to learn, a reaction against the minimum-competency testing movement of the 1970s and 1980s (National Council on Education Standards and Testing, 1992).

Resnick and Zurawsky (2005) describe four tenets of standards-based education that emerged in the 1990s from national-level discussions among educators, business leaders, and legislators: (1) a public process to establish standards for what students should know and be able to do at different grade levels, (2) standards-based assessments to inform students about their learning and teachers about their instruction, (3) standards-based instructional programs and teacher professional development provided by schools and districts, and (4) accountability systems to determine whether students are achieving the standards. Professional organizations responded to the national agenda by establishing model standards in the different subject areas. The first content standards were published for mathematics in 1989 (National Council of Teachers of Mathematics, or NCTM) followed by other content areas in the mid-1990s (e.g., science, language arts, and social studies).

The Elementary and Secondary Education Act (ESEA) of 1994 established federal and state roles in standards policies. Also known as the Improving America's Schools Act,

this law required that states adopt formal standards in reading and mathematics and administer assessments based on these standards as a condition for receiving funds for Title 1 students. A state's standards for Title 1 students were to be the same as those established for other students in the state. The 1994 law effectively required those states that had not adopted standards to do so (Resnick & Zurawsky, 2005). NCLB, the 2001 version of ESEA, continued the federal focus on standards-based education and increased the emphasis on improving the achievement of historically underperforming groups of students. Among its provisions, NCLB requires that schools receiving Title 1 funds make adequate yearly progress toward achieving high standards as indicated by student performance on standards-based tests and that achievement data be disaggregated for subgroups of students.

Standards are mandated by law, but why are standards important? Ravitch (1995) explains that standards are a logical necessity because people cannot accomplish what they do not know they are trying to accomplish. "Content standards — what children are expected to learn — are necessary for educational improvement," states Ravitch, "because they are the starting point for education" (p. 25). She observes that federal involvement in the standards movement was due to the failure of educators to agree about what students should learn, leaving these decisions to interest groups and textbook publishers.

How can standards improve education? As described by Ravitch (1995), supporters of standards claim that standards can improve student achievement, equalize student opportunities, coordinate different parts of the education system, provide accurate information to parents and students about expectations, and indicate to stakeholders the degree of student progress in relation to expectations. Ravitch, like other supporters of standards, asserts that

> *standards can improve achievement by clearly defining what is to be taught and what kind of performance is expected*. They define what teachers and schools should be trying to accomplish. They can raise the quality of education by establishing clear expectations about what students must learn if they are to succeed. If the goals of teaching and learning are spelled out, students understand that their teachers are trying to help them meet externally defined standards and parents know what is expected of their children in school. (pp. 25–26)

According to Kendall and Marzano (2000), standards are important not only because they clarify and explicitly state expectations for student learning, but also because they provide a common set of expectations for all those involved in the education process. Wheelock (1995) views standards as proxies for teachers' high expectations for student learning, a concept that is difficult to measure and observe without explicit standards.

Another view of standards concerns their role in raising intellectual rigor. Sandoltz, Ogawa, and Scribner (2004) comment, "Academic standards are intended to create more intellectually demanding content and pedagogy; thereby improving the quality of education for all students" (p. 1178). For example, standards emphasize the development of students' thinking skills and teaching for understanding (Wheelock, 1995; McClure, 2005). Equity is a recurrent theme in proponents' discussions of standards. A common set of standards for all children in all schools was expected to equalize education opportunities and, as a result, close the achievement gap between advantaged and disadvantaged children (Sandoltz et al., 2004; Buttram & Waters, 1997; McClure, 2005).

Thus, standards can provide a common set of clear expectations for all students with the assumption that their implementation will result in an improved system of education and higher student achievement.

CONCEPTUAL FRAMEWORK

Two definitions of standards-based education dominate the literature. One definition describes a policy approach and does not address instruction specifically. The other definition primarily emphasizes sweeping reforms in teaching and learning, consistent with the recommendations of the National Council of Teachers of Mathematics (NCTM) and other reform groups.

In the definition of standards as a policy approach, standards-based education entails education based on (1) goals for student learning that incorporate broad descriptions of knowledge and skills that students should acquire for a given content area, (2) specific descriptions of student performance that indicate mastery of a given content area, and (3) assessment that provides feedback about student performance relative to learning and performance goals (McLaughlin & Shepard, 1995; McREL, 2000). This definition makes no distinction based on variations in the emphasis of different standards. Standards that are "reform-oriented" (e.g., focusing on higher-order thinking), those that are more "traditional" (e.g., emphasizing the acquisition of important basic skills), and those that fall somewhere between — all can serve as the basis for a standards-based system using this definition.

The definition of standards as a basis for instructional change incorporates constructivist ideas about learning, including student-centered pedagogy, active learning, and cooperative grouping structures, rather than more traditional, teacher-centered classroom discourse (Snow-Renner, 2001; Thompson, 2001). Standards documents by the NCTM (1989, 1991, 2000) and the National Research Council (1996) called for a shift in the way content was taught. In mathematics, for example, this was a shift from emphasizing computation and memorization of facts and procedures to emphasizing the development

of conceptual knowledge. This definition of standards-based education is clearer in its implications for pedagogical change than the more general policy definition.

Although these two definitions of standards-based education appear in the literature, they are not consistently identified in research studies. To address this issue, the synthesis authors have linked the studies examined in this synthesis to the specific standards they address whenever possible.

Regardless of whether standards are considered in terms of policies about content or in terms of instructional change, many educators contend that systemic reform is required for successful implementation of standards-based education. Clune's (2001) analysis, for example, emphasizes the role of system variables in achieving standards-based education. He describes the central thesis of standards-based reform as follows:

> *Standards-based reform* (SR), through its purposeful activities, leads to *Standards-based policy* (SP), which leads to a *rigorous, implemented standards-based curriculum* (SC) for all students, leading to measured *high student achievement* (SA) in the curriculum taught. (p. 15)

According to Clune, the impact of standards occurs through the dynamic interaction of reform activities, policies related to the content and implementation of standards, and a curriculum that includes content and pedagogy that align with standards. Thus, Clune's analysis emphasizes the role of systemic reform in standards-based education.

Massell, Kirst, and Hoppe (1997) define standards-based systemic reform as (1) the establishment of student learning standards, (2) the alignment of policies to standards (e.g., policies on testing, accountability, teacher professional development, and teacher certification), and (3) the restructuring of governance so schools and districts are responsible for developing instruction that meets the standards. Though this definition overlaps somewhat with the policy definition of standards-based education (both reference academic standards), Massell et al.'s description emphasizes the point that there are education system variables that support standards implementation, including policies, alignment, and governance.

The National Research Council (NRC) provides a more detailed view of standards-based education and standards-based reform in its 2002 publication *Investigating the Influence of Standards: A Framework for Research in Mathematics, Science, and Technology Education.* In the NRC framework, standards-based reforms influence the education system through three main paths or channels: (1) curriculum (state and district policies, instructional materials, texts); (2) teacher development (preparation, certification, professional development); and (3) assessment and accountability (classroom, state, and district assessments; accountability systems; college entrance). Together, these channels

affect teachers and teaching, which in turn influence student learning. In addition, there are contextual forces that occur outside the education system that directly or indirectly influence teaching and learning. Contextual influences may emanate from the business community, public opinion, and politicians. The NRC framework suggests two main questions to guide inquiry about the influence of standards on the education system: "1) How has the system responded to the introduction of nationally developed standards?, [and] 2) What are the consequences for student learning?" (p. 34). Importantly, the NRC points out that no one type of study can be used to answer these questions:

> Rather, various types of studies, each guided by its own appropriate methodologies, will be needed to establish the scale and scope of influences, identify routes by which standards actually exert influence, and ascertain the direction and educational consequences of those influences (p. 10).

The NRC proposes that researchers place studies on standards within the NRC framework so that the channels of influence (e.g., state policies about curriculum, teacher professional development) and connections, or lack thereof, to other parts of the education system (e.g., assessment, classroom teaching) can be identified.

How should the influences of standards-based education be studied, given standards-based reform's breadth and complexity? Researchers have different views about designs and frameworks for evaluating the influences of standards-based reform. Some contend that evaluations of standards-based education should consider the entire education system (Puma, Raphael, Olson, & Hannaway, 2000; Dutro, 2002; Chatterji, 2002). These researchers believe that because a systems perspective is needed to change education outcomes, a systems perspective is needed to study education reform. For example, Puma et al. (2000) propose a multi-level design that would monitor and evaluate policies and activities at the state, district, school, and classroom levels. The study would culminate in a randomized experiment in which schools that are weakly aligned with systemic reforms would be assigned to a treatment or a comparison group. Treatment schools would receive technical assistance and resources designed to bring them to a high level of systemic reform. Student achievement would be measured before and after the intervention. Puma et al.'s proposal represents an ideal way to study standards-based education, but it is also costly and time-consuming. Dutro also supports the use of systemic approaches to study standards-based education, but she notes, "We may never be able to directly answer the question 'What impact are state content standards having on student learning?'" (p. 6).

Other researchers acknowledge the influences of system components but emphasize the need to link standards-based education to instruction and student achievement (Cohen, 1995; Nave, Meich, & Mosteller, 2000). In an early discussion of systemic reform, Cohen

stressed that in a successful standards-based system, standards-based instructional guidelines need to influence teachers and their classroom instruction. With regard to student learning, Nave et al. observe, "Standards-based reforms might improve student achievement, but little research definitely linking the two is available" (p. 132). They urged researchers to examine more than just the implementation of standards and to evaluate reformers' theories of action by which standards are expected to improve student achievement.

Conley's (1993) model of education examines both the education system and its components, similar to the NRC (2002) framework described previously. However, the NRC framework for inquiry on standards makes no claims about the relative importance or centrality of any one channel of influence or part of the education system. In contrast, Conley's analysis of school restructuring focuses on 12 dimensions grouped into three types of variables that play different roles in reform: (1) supporting variables (governance, teacher leadership, personnel, and working relationships); (2) enabling variables (learning environment, technology, school-community relationship, and time); and (3) central variables (learner outcomes, curriculum, instruction, and assessment/evaluation). According to Conley, enabling and supporting variables may be required or needed, but the central variables are the reason for public education, and change at this level is fundamental to reform. Conley also comments that change in central variables is more difficult to achieve than is change in the enabling and supporting variables. "When developing 'restructuring strategies,'" writes Conley, "most educators appear to prefer to look first at change in almost anything other than these variables" (p. 107).

As Conley (1993) indicates, ultimately it is in classrooms that children will receive the instruction and assistance that they need to achieve standards. In Marzano's (2000) analysis of the influences of school, teacher, and student variables on variation in student achievement, variables related to teaching (e.g., instructional practices) accounted for 13.3 percent of the variation, while school-level variables (e.g., school climate) accounted for 6.6 percent. Student variables (e.g., socio-economic status) accounted for 80 percent of the variation in student achievement. These results suggest that variables that are closer to the classroom (central variables) have the potential to influence student outcomes to a greater degree than do those that occur at the school level (enabling and supporting variables). Furthermore, because of the influence of student variables, it is important for teachers to have access to effective instructional practices that can help children from all backgrounds achieve.

Standards-based systemic reform and standards-based education are complex constructs, and each cannot be accomplished without the other. Teaching and learning cannot change unless they are supported by changes in the education system, such as changes in policies and governance (Massell et al., 1997). Yet if these policy changes do not reach the

classroom in the form of changes in practice, then the benefits of standards-based education cannot be fully realized (Cohen, 1995).

While acknowledging that system variables and contextual influences are important (NRC, 2002), McREL's research synthesis focuses on the three variables most closely related to teaching and student learning: standards-based curriculum, standards-based instructional guidelines, and standards-based assessment. McREL researchers review studies that examine the influences of each of these three input variables on teaching and/or student learning, using the definitions and methods described in Chapter 2. Figure 1 illustrates the conceptual framework for this synthesis. Standards-based curriculum, instructional guidelines, and assessment are assumed to influence instruction, which in turn, influences student achievement.

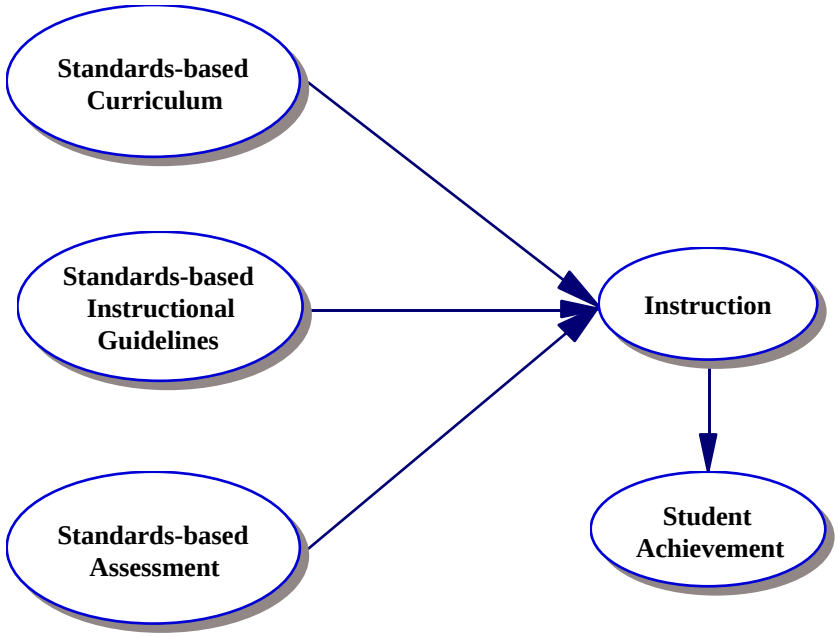

Figure 1. The Influence of Standards on Teaching and Learning

To summarize, this synthesis addresses the following research problem: What is the influence of standards on K–12 teaching and student learning? McREL attempts to answer this by synthesizing research studies that address one or more of the following questions:

1. What is the influence of standards-based curriculum on teacher instruction and student achievement?

2. What is the influence of standards-based instructional guidelines on teacher instruction and student achievement?

3. What is the influence of standards-based assessment on teacher instruction and student achievement?

RESEARCH CONTEXT

Many research and evaluation studies have addressed different aspects of K–12 standards. However, only a few systematic reviews have examined evidence that the implementation of standards improves education outcomes. In this section the synthesis authors discuss these reviews and other bodies of research that informed the synthesis. These include reviews of the literature on standards in particular content areas (mathematics, literacy, and science), a review of the types of research designs used to study the influences of standards, research generated through study of the National Science Foundation's (NSF's) Systemic Initiatives (SIs), research on how standards apply to students with special learning needs, policy analyses evaluating the nature of different states' standards and accountability systems, and research on educator and public attitudes about standards-based education.

Prior Research Reviews

In mathematics, Ross, McDougall, and Hogaboam-Gray (2002) conducted a narrative review of studies conducted between 1993 and 2000 that investigated the effects of mathematics education reforms on student outcomes and the difficulties in implementing such reforms. They concluded that compared to students in traditional classrooms, students in classrooms that have implemented mathematics education reforms have higher achievement on reform measures such as problem solving and are no worse on traditional measures such as computation. Ross et al. also observed better attitudes toward mathematics among the students in reform classrooms. The researchers documented many barriers to implementing mathematics reforms, including the challenges of delivering instruction of the type that teachers did not receive as students and teachers' lack of subject-matter knowledge.

In the area of literacy, Valencia and Wixson (2000) reviewed policy-related research on standards and assessments. They found mixed evidence for positive influences of literacy standards on teachers' beliefs and practices and noted that effects were mediated by many factors. These included the political, economic, and social conditions of the schools and districts; the support teachers received from administrators; and the stakes associated with the standards and assessment policies. The researchers state,

> It is equally clear that policy by itself is not sufficient to promote desired change; simply implementing new assessments or creating new standards does not insure [sic] improved teaching or learning. What is less clear,

> however, is just what it would take to promote change in the desired direction or to insure [sic] improved teaching or learning. (p. 930)

Valencia and Wixson comment that few studies they reviewed included measures of student achievement and that future research on standards-based education in literacy needs to address this shortcoming.

A report published by the National Academy of Sciences (Hollweg & Hill, 2003) describes a workshop and research reviews on the influences of the National Science Education Standards (NSES) on science curriculum, professional development, assessment and accountability, teachers and teaching, and student achievement. The reviews were conducted by researchers of science education and varied in comprehensiveness. Hill (2003) summarized the researchers' assessments of the literature related to standards-based science. There is evidence that the NSES have influenced science curriculum, but the majority of instructional materials used by teachers are not yet aligned with science standards. There has been some influence of the NSES on professional development, but the evidence is weak, and there is less evidence for influences on state-level policies related to professional development and teacher preparation. Results related to the NSES influences on assessment and accountability were inconclusive due to the lack of research, but it was noted that assessments aligned with the NSES should be different from traditional assessments. There is more evidence that the NSES have influenced teachers' beliefs and attitudes than their actual classroom practices. Teachers agree with the science standards but many teachers, particularly in lower grades, lack the necessary training to implement them in the classroom. There is weak support for a link between the NSES and improved student achievement but no evidence that the NSES have decreased the achievement gap. However, there was no evidence that the science standards had negative impacts on student achievement. All the reviewers mentioned the need for additional research studies.

Chatterji (2002) reviewed methods of inquiry about standards-based reform and suggested that, to date, there has been little coherence in the research and evaluation of the influence of standards. Chatterji's review focuses on the *designs* used by research and evaluation studies to examine systemic reform. Our synthesis adds to this knowledge by describing and analyzing the *results* from extant research on the influence of standards on teaching and student learning.

Research on National Science Foundation Systemic Initiatives

The NSF SIs have stimulated a body of research related to standards-based education in mathematics and science (e.g., Zucker, Shields, Adelman, Corcoran, & Goertz, 1998; David & Shields, 2001). In 1990, NSF established a program to support systemic reform

in mathematics and science education, beginning with the Statewide Systemic Initiatives (SSIs), followed by the Urban Systemic Initiatives (USIs) and the Rural Systemic Initiatives (RSIs). States, cities, and rural entities submitted proposals to NSF. Funded proposals were managed through cooperative agreements with NSF that included both internal and external evaluations (Kahle & Kelly, 2001). The NSF vision for these programs was one of systemic reform that includes six central elements or "drivers": (1) high standards-based instruction for all students supported by curriculum, professional development, and assessment; (2) aligned policies, practices, and accountability mechanisms; (3) coordinated resources; (4) involvement of the community of stakeholders; (5) increased student achievement in mathematics, science, and technology; (6) a reduction in the achievement gap between disadvantaged students and their peers (Zucker et al.). Kahle and Kelly observe that in general, most SIs began with a focus on reformed teaching practice through teacher professional development, and they addressed policy changes in the later years of their programs. McREL's examination of the research revealed great variation in methods and approaches in the evaluations of the SIs. Several of the studies are included in this research synthesis; however, in some of the studies, it was not possible to make direct connections between our input variables of curriculum, instructional guidelines, and assessment and the outcomes of teaching and student learning.

Research on Learning Disabled Students

Research on the influence of standards-based education on students with learning disabilities is an important area of inquiry but is in the early stages of development as an area of research. McREL found valuable discourse on this topic (e.g., Fraser, 1996) but few empirical studies that fit the conceptual framework for the synthesis. For example, based on case studies in four states, Raber, Roach, and Fraser (1998) describe how standards-based reforms at the state level interact with the efforts of local school districts to serve students with learning disabilities. They recommend more involvement of special educators in the development of state policies related to standards. The researchers also call for a discussion of accountability measures for special education students and question whether state standards and curriculum frameworks are appropriate for diverse students, including those with learning disabilities. Cooney (2001) reports on the efforts of nine school districts to respond to standards-based education policies in relation to special education in secondary schools. Based on teacher interviews and observations, Cooney suggests that standards-based reforms lead to standardization and inflexibility that hinder efforts to meet the needs of special education students. Woodward and Montague (2002) discuss the tenets of mathematics reform and describe studies that show the challenges that learning-disabled students face when taught in ways consistent with NCTM standards, such as instruction on problem solving. Other studies have addressed the issue of test accommodations for learning disabled students participating in state assessments (Thompson, Blount, & Thurlow, 2002). This brief discussion suggests there

are many issues related to standards-based education and learning disabled students. It is a complex area of concern, and although most of the issues are beyond the scope of this synthesis, it is a topic that deserves attention by education researchers as well as administrators and policymakers.

Policy Analyses

Another area of study that relates to the current synthesis concerns the "state of state standards." In recent years, several organizations have conducted descriptive analyses of the rigor of state standards and the related system components of curriculum, assessments, and accountability. Many of the studies in this synthesis examine teacher and student outcomes in relation to state standards, so it is useful to consider the quality of these standards in the approximate years that the research was conducted. However, it should be noted that judgments of state standards are not uniform; the entities evaluating those standards use differing criteria for quality, and consequently the same state standards often are rated differently by different raters.

The Council of Basic Education (CBE) examined state standards for mathematics and language in 1998 (Joftus & Berman). In their analysis, rigorous standards "address essential concepts and skills" and "require student understanding and application of these essential concepts and skills at a level of sophistication or complexity that is appropriate and challenging to students at a particular grade level" (pp. 12–13). At the time of the CBE study, 43 states had mathematics standards ready for review, and 42 states had language arts standards ready for review. Ratings for states' mathematics standards were 37 percent "very rigorous," 56 percent "rigorous," and 7 percent "low in rigor." Ratings for states' language arts standards were 17 percent "very rigorous," 50 percent "rigorous," and 33 percent "low in rigor." CBE found state mathematics standards to be more rigorous than state language arts standards. Mathematic standards addressed most major concepts and skills, but most language arts standards addressed mainly basic skills and excluded higher-order skills such as literature study.

In 2001, the American Federation of Teachers (AFT) reported on state efforts to implement standards-based education. In their analysis, 58 percent of the states had "clear and specific" (p. 5) standards in the core subject areas of mathematics, language arts, science, and social studies at the elementary, middle, and high school levels. Seventy-six percent of the states had begun to align their tests with their standards, and 18 percent had aligned tests in the four core subjects at the different education levels. The AFT was particularly concerned about states' lack of fully developed curriculum models because they view curriculum as the road map for guiding teachers to help students meet standards. According to the AFT, state curriculum models should include learning continua for grade level progression, instructional resources, instructional strategies, performance indicators, and lesson plans. They found that 82 percent of the states had

less than half of the curriculum components fully developed across the core subject areas, with more state curriculum support for language arts than the other subject areas.

Quality Counts, published by *Education Week*, also reported on state standards in 2001, noting that 47 states had established standards in the core subject areas, although not necessarily at all education levels. All 50 states administered student assessments, with a majority using both multiple-choice and short answer formats, but only seven states used essays questions for subjects other than language arts. For the "grades" that *Quality Counts* assigned to states for their standards and accountability mechanisms, 46 percent of the states received As or Bs, 14 percent received Cs, and 40 percent received Ds or Fs. Grades were determined by the clarity and specificity of standards in the four core subject areas, the types of test items on the state assessments, the use of criterion-referenced tests, and number of accountability mechanisms, including school report cards and ratings, rewards, assistance, and sanctions. The *Quality Counts* report for 2005 indicates that 48 states and the District of Columbia have established standards in the four core subject areas. (The exceptions are Rhode Island, which had standards in three subject areas, and Iowa, which had no standards at the state level.) In the "grades" assigned to states in 2005 for their standards and accountability, 66 percent of the states received As or Bs, 14 percent received Cs, and 20 percent received Ds or Fs.

Studies of Attitudes about standards-based Education

One final area of research that relates to the current synthesis is research on educator and public attitudes towards standards-based education. Public Agenda summarized surveys of public opinions toward public education (Johnson & Duffet, 2003). In general, teachers, parents, and employers indicate strong support for high academic standards, although they demonstrate less support for standardized tests. The vast majority of teachers believe that standards can help students improve performance. There are other studies of teachers' attitudes toward standards, but McREL found few attempts to link teacher attitudes to teacher instruction and/or student achievement. With NCLB in place, teachers must address standards whether they support them or not. However, an interesting question (although one that is beyond the scope of the current synthesis) is whether teacher and principal attitudes towards standards-based reforms influence instruction and student achievement.

Summary

Prior research related to the current synthesis includes subject-specific literature reviews, a review of methods of inquiry used to investigate standards-based reforms, research on NSF-supported reform initiatives, a developing body of research on issues related to standards-based education for disabled students, recurring descriptive analyses of the

condition of state standards and accountability systems, and survey studies of educator and public attitudes toward standards.

Based on McREL's consideration of this prior research, the current research synthesis contributes to the knowledge base about standards-based education in the following ways:

- This synthesis examines research on curriculum, instructional guidelines, and assessment, the input variables in a standards-based education system that are most closely connected to the classroom.
- This synthesis examines research that links standards-based variables to teacher instruction and to student achievement, the two outcomes that are the primary concerns in a standards-based education system.
- Studies of the four core subject areas are included in this review: language arts, mathematics, science, and social studies.
- Studies in this review are coded for research methods, and synthesis results are described with respect to these methods.

CHAPTER 2: METHODOLOGY

McREL researchers used systematic methods to identify and review research on the influences of standards on K–12 teaching and student learning. McREL staff consulted other methodologies that were published on synthesis methodology and/or conducted research syntheses (Cooper, 1998; Shanahan, 2000; Wilson, Floden, & Ferrini-Mundy; 2001) for design guidance and review.

LITERATURE SEARCHES

McREL researchers conducted several literature searches for studies related to standards-based education and the goals of this synthesis. The original date parameters for the synthesis were 1991 through February 2005. The synthesis authors later changed these to 1995 through 2005 to reflect the years in which standards-based reforms were more widely implemented and would be expected to begin to demonstrate impacts.

ERIC Searches

Two searches of the Education Resources Information Center (ERIC) database were conducted. Each search had a unique design that involved sub-searches. The primary searches were conducted in April 2004, prior to the U.S. Department of Education's reorganization and improvement of the ERIC database and search capabilities. The secondary searches were conducted in November 2004, using the updated ERIC website that became available September 1, 2004.

The primary ERIC searches were designed as a comprehensive search of the education research literature related to standards-based education and reform. Using FirstSearch, four separate ERIC searches were conducted using each of the following search terms for the years 1991-2004: *accountability, alignment, assessment, curriculum, governance, instruction, policy, professional development, inservice, reform, teacher, teacher education, preservice, teacher preparation.* Each search term was combined with secondary search terms of *standards and study, standards and evaluation, reform and study,* and *reform and evaluation.* (Every search also included the phrases *not college, not foreign.*) The total number of citations (after removing duplicates) was 3,278. Four researchers divided the set of 3,278 citations into four groups and read and screened the abstracts for initial inclusion in the synthesis. The abstract screening resulted in 798

studies for possible inclusion and 2,480 excluded studies. The primary reason for exclusion (1,464 studies) was that the abstract did not indicate any empirical data (articles were historical essays, position papers, or guidelines rather than research). Additional reasons for exclusion were that studies were not related to K–12 standards (n=575), studies did not concern K–12 education (n=401), or studies were conducted in foreign school systems (n=11). To assess the reliability of abstract screening, an independent researcher coded a random sample of 10 percent of the abstracts that were excluded by each of the four coders. The degree of agreement about the decision to exclude was 99.6 percent. Two researchers then read each of the 798 included abstracts to identify which full reports should be read, based on whether studies addressed standards-based education. The researchers reviewed their results and resolved discrepancies after discussion amongst themselves and with an independent researcher. There were 326 studies identified for reading from the primary ERIC searches. Based on the results initially, the researchers determined that the synthesis should focus on the variables of standards-based curriculum, instruction, assessment, and professional development. However, later, the synthesis authors decided not to include professional development studies in the current synthesis based on the observation that this body of research needs a separate review.

The secondary ERIC searches were designed to double-check the primary ERIC search using the updated ERIC database. Each secondary search was limited to studies conducted in the years 1991–2004, with descriptors of *academic standards*, *state standards*, or *national standards* and report types of *evaluative* or *research.* (Each search also included the phrases *not college, not foreign*.) Four separate searches were conducted using the following keywords: (1) *mathematics*; (2) *science*; (3) *language arts* or *reading* or *writing*; and (4) *social studies, history*, *geography*, *civics,* or *economics*. A fifth search was conducted using the keywords of *teachers* and *not teacher education, mathematics, science, reading, writing, history, geography, foreign.* The total number of citations (including duplicates) was 1,184. A researcher read each abstract to identify those that addressed the standards-based variables within the synthesis framework and for which the report was not previously ordered. The secondary ERIC searches resulted in the identification of 172 studies to read.

Psychological Abstracts Search

Using PsycINFO, four separate keyword searches of *Psychological Abstracts* were conducted using each of the following search terms for the years 1991–2004: *curriculum, instruction, assessment, professional development, inservice, teacher, reform*. Each search term was combined with the secondary search terms of *standards and study, standards and evaluation, reform and study, reform and evaluation.* Each search was limited to *empirical studies*; *childhood* or *adolescence,* and *journals, dissertations, books or chapters*. The total number of citations (after removing duplicates) was 1,031. A

researcher read each abstract to identify those for which the report should be ordered, based on whether the studies addressed standards-based education. Most of the studies addressed topics unrelated to education, but 71 studies were identified for possible inclusion. Three additional keyword searches were conducted using the following search terms: *standards, education and reform, education and alignment.* The total number of citations from these additional searches (after removing duplicates) was 59. Two researchers then read each of the 130 study abstracts to identify which full reports should be read, based on whether studies addressed the standards-based variables in the synthesis framework. The researchers reviewed their results and resolved discrepancies after discussion amongst themselves and with an independent researcher. There were 42 studies identified for review from the searches of *Psychological Abstracts*.

Dissertations Search

Using UMI ProQuest Digital Dissertations, a search was conducted of dissertation abstracts for the years 1991–2003 using the search terms of *state standards* or *national standards* and limiting the search to subject areas of *elementary education* or *secondary education*. The total number of citations was 360. A single researcher read each abstract to identify those for which the report should be obtained, based on whether studies addressed standards-based variables. There were 48 dissertations identified for possible ordering. A second researcher read each of the 48 abstracts and also the first 24 pages that were provided for each of 42 of the dissertations. No relevant dissertations were identified from the search of dissertation abstracts. (Eleven dissertations had previously been identified from the ERIC and PsycINFO searches.)

Supplementary Searches

Websites for 20 organizations were searched for research reports on standards-based education. Appendix A lists these organizations. The search of organization websites resulted in the identification of 11 reports for ordering. (Most relevant reports on websites had been previously located through the ERIC searches.)

As of March 1, 2005, the ERIC database did not include citations for 2004. Therefore, the tables of contents of 25 education research journals were searched for relevant articles published from January 2003 through February 2005. Appendix A lists the searched journals. Twenty-six reports were identified for ordering through the search of tables of contents.

Researchers also searched the reference sections of ordered studies for additional citations related to the synthesis research questions. “Reference chasing” resulted in the identification of 120 additional reports for ordering.

Summary of Literature Searches

In summary, McREL researchers read 3,278 abstracts from the primary ERIC searches, 1,184 abstracts from the secondary ERIC searches (the latter included duplicate abstracts), 42 dissertation abstracts, and 130 PscyhINFO abstracts. The researchers obtained a total of 540 study reports from these four sources, and ordered an additional 156 study reports from searches of websites, tables of content, and the reference sections of previously reviewed reports. In total, the synthesis authors read 697 study reports.

CRITERIA FOR INCLUSION OF STUDIES

The criteria for including studies in the synthesis reflect the research problem, the conceptual framework for the synthesis, and the goals of the synthesis. After the first screening of the 697 studies, the synthesis authors modified the criteria to exclude 16 studies that met the inclusion criteria but that were reported before 1995. McREL researchers also decided to exclude the 60 studies that met the inclusion criteria and that focused primarily on standards-based teacher professional development. These studies deserve fuller consideration in a separate research synthesis, which McREL hopes to conduct in the future. The final criteria for including studies in the current synthesis were the following:

1. *Studies must include empirical data.*

2. *A research or evaluation study must be published or reported in or after January 1, 1995.*

3. *Studies must be conducted in the United States or its territories.*

4. *Studies must involve K–12 students.*

5. *Studies must involve one or more of the four core subject areas of reading/language arts, mathematics, science, or social studies.*

6. *Studies must include assessment or documentation of teacher instruction or student achievement outcomes.* For the purposes of the synthesis, teacher instruction is defined as indications of classroom teaching practice in core subject areas. Teacher instruction includes content coverage (or the standards that the teacher addresses in class, also called the teaching curriculum) and pedagogy (or the specific teaching instructional strategies used by teachers, separate from the content they cover). Instruction is measured through teacher perceptions about their classroom practice (reported on surveys or in interviews), classroom observations, principal perceptions of teacher

instruction (reported on surveys or in interviews), principal evaluations of teacher instruction, and teacher reports on their instruction (e.g., in logs or journals). Student achievement is defined as indications of student academic performance in core subject areas. In quantitative studies, student performance is measured by a direct assessment (e.g., classroom assessments, standardized tests, grades in subject areas) of students' academic achievement. In qualitative studies, student learning in core subject areas may be documented through observations or teacher reports.

7. *Studies must have explicit connections to national, state, or local content or performance standards or to standards-based reform.* The synthesis uses Kendall's (2001) definitions of content and performance standards. "A content standard is a summary description regarding what it is that students should know and/or be able to do within a particular discipline....A performance standard describes levels of student performance in respect to the knowledge or skill described in a single benchmark or a set of closely related benchmarks....A benchmark is a clear, specific description of knowledge or skill that students should acquire by a particular point in their schooling" (p.2). Standards-based reforms were defined as those with, at a minimum, core features of a standards-based education system — "content and performance standards for each school discipline, along with assessments aligned to the standards" (Briars & Resnick, 2000, p.1). As addressed in Chapter 1, content standards in such a system may address a range of knowledge, from basic to higher-order skills, and synthesis authors indicate this when it is noted in a particular study. Further, some studies explicitly addressed standards as the basis for instructional change consistent with "reform-oriented" practice (e.g., constructivist practices consistent with the recommendations of the NCTM). In such cases, the synthesis authors describe the particular standards addressed in the study.

8. *Included studies had to address the stated research questions.* Studies that were beyond the scope of this synthesis were excluded. For example, studies that focused solely on the influences of stakes or consequences attached to tests were excluded because this issue primarily concerns accountability policies, which the synthesis does not address. To be included, studies had to address the influence of one or more of the following standards-based variables: curriculum, instructional guidelines, and assessment. Standards-based curriculum was defined as a course of study that reflects or is aligned with

content and performance standards (McLaughlin & Shepard, 1995). This includes specific standards-based curricula in core subject areas (e.g., Everyday Mathematics) and standards-based policy documents, such as those developed by the National Council of Teachers of Mathematics (NCTM, 1989). In particular instances, such standards-based curricula imply an alignment with "reform-oriented" practices, and, where appropriate, that is noted. McREL researchers defined standards-based instructional guidelines as guidelines for standards-based teaching in the content areas.[1] An example is the Professional Standards for Teaching Mathematics. "These standards present a vision of what teaching should entail to support changes in the curriculum set out in the Curriculum and Evaluation Standards" (NCTM, 1991, p. vii). As described in Chapter 4, the studies addressing such guidelines generally reflect the "reform-oriented" definition of standards-based education, and where appropriate the synthesis references the particular guidelines and/or practices. McREL researchers defined standards-based assessment as classroom, district, state, or national assessments that reflect or are aligned with content and performance standards (Clune, 2001). As explained in Chapter 5, this definition does not carry implications of reform orientation with it; rather it implies the policy focus on the content of particular standards.

9. *Studies of reform models had to disaggregate the influence of standards on outcomes from the influences of other aspects of the reform.* Based on this criterion, the synthesis excludes studies of the Comprehensive School Reform (CSR) models identified in Borman, Hewes, Overman, and Brown (2003) and studies of other reform initiatives (e.g., Philadelphia's Children Achieving), if the study did not disaggregate outcomes for standards-based influences.

10. *Quantitative studies had to meet all of the following criteria for rigor: (a) describe instrument development; (b) report a response rate for surveys; (c) describe in detail the processes used for data collection and analysis; (d) include sufficient evidence of the results such as such as tables of means and statistics.*

[1] McREL researchers changed the name of the instruction variable in their conceptual framework to "standards-based instructional guidelines" to distinguish between instruction as input to standards-based education and instruction as an outcome of standards-based education.

11. *Qualitative studies had to meet all of the following criteria for rigor: (a) describe in detail the processes and methods used for data collection; (b) describe in detail the processes used for data analysis; (c) include sufficient evidence of the results, such as detailed descriptions of events or observations and/or samples of responses from interviewees; (d) use a process to validate the results, such as a search for disconfirming evidence.*

12. *Mixed methods studies had to meet the applicable criteria established for quantitative and qualitative studies.*

Both published and unpublished studies, including evaluation reports, conference presentations, and dissertations were included in the items reviewed. This approach was an effort to avoid the null hypothesis problem whereby studies that do not find effects from an intervention or policy are excluded from the synthesis because they are not published (Cooper, 1998).

Of the 697 studies read, 120 studies met the inclusion criteria. Because seven of these were multiple volumes of the same studies, the actual number of different studies reviewed was 113. A total of 501 studies were excluded upon review. (An additional 76 studies reported prior to 1995 or that concerned primarily professional development were excluded previously.) A second researcher verified the decision to exclude a study, and disagreements were resolved through conference.

The authors conducted two article screenings to determine whether studies should be included, which resulted in excluding 377 and 124 studies respectively. The first screening determined whether studies met inclusion criteria 1–8. The most frequent reasons for excluding studies at this stage were the following: studies did not make explicit connections to national, state, or local content or performance standards or to standards-based reform (n=182); studies did not assess or document teacher instruction or student achievement (n=101); and studies did not address the synthesis research questions (n=50). The second screening determined whether studies met inclusion criteria 9–12. The most frequent reasons for excluding studies at this stage were the following: studies of CSR models or reform initiatives did not disaggregate outcomes in the study for standards-based influences (n=54); and studies did not meet the criteria of research rigor for quantitative, qualitative, or mixed methods designs (n=50). (It is probable that many studies already excluded in the first screening would not have met the criteria for research rigor.)

CODING STUDIES

The instrument used to code studies is contained in Appendix B and has four parts. Part 1 is a summary that indicates the study citation, the source of the report (e.g., journal, dissertation, etc.), and a checklist of the criteria for inclusion. Part 2 describes methodology, which includes the research approach (quantitative experimental, quantitative quasi-experimental, quantitative non-experimental, and qualitative), the variables addressed (curriculum, instructional guidelines, and assessment), the outcomes and measures, the subject areas, the level of standards, and systemic influences included in the studies (e.g., governance, leadership). If the study addressed two or three of the standards-based variables, McREL designated as primary the variable that was the main focus of the study. For example, if the main focus of a study was curriculum, then that study was assigned to the curriculum chapter for review. If a secondary emphasis was assessment, that study would still be reviewed in the curriculum chapter, but briefly discussed in the assessment chapter as well. Part 3 of the coding instrument examines the study population and study sample, including at-risk indicators and the number and characteristics of study participants. Part 4 is a detailed description of results for the both the "primary" and "secondary" variables addressed by the study. The synthesis authors also noted implications for education practices and policies. To facilitate consistent coding, a coding manual was developed that includes definitions of terms and explanations and examples for each entry on the coding instrument.

Coding procedures were designed to reach a common understanding of each study and to check for the reliability of coding results among the authors. Coding procedures incorporated Stock's (1994) recommendations for reducing coding errors.

Each of the synthesis authors participated in coder training, which involved a review of the coding form and coding manual, explanations for items in each section, and examples of information from studies to be extracted and judged. The synthesis authors confirmed that they had a common understanding of terms used for coding and that the instrument included sufficient information for adequate description of study characteristics. Following initial training, each author independently coded the same two studies, one quantitative and one qualitative. The authors then compared completed forms and identified and resolved discrepancies. Based on the resolutions, clarifications were made to the coding instrument and manual. The authors independently coded two additional studies, which resulted in improved coding consistency.

Coding procedures and decisions were double-checked at several points during study coding and analysis of findings. The majority of studies that passed the first screening were screened for the second time by a second researcher. A different researcher coded the majority of studies that passed the second screening. As mentioned previously, throughout study selection, two researchers examined every study excluded. Double-

checking of coding results occurred during preparation of chapter tables and reporting of findings, during internal review of the chapter drafts, and during chapter revisions. All discrepancies were resolved among the synthesis authors.

SYNTHESIZING STUDY FINDINGS

Researchers have discussed different types of syntheses in the literature and provided guidelines for their conduct (Cooper, 1998; Shanahan, 2000). There is disagreement, however, about which synthesis methods are most appropriate (Wayne & Youngs, 2003). The decision to conduct a meta-analysis as part of the synthesis methodology typically is based on whether there are sufficient quantitative studies, whether the studies report the necessary data for a meta-analysis, and whether it makes sense to combine the studies in a single analysis. After reading and coding the studies identified for this synthesis, McREL researchers determined that a meta-analysis was neither feasible nor appropriate.

Of the 113 studies that met the inclusion criteria, two studies used experimental research designs with random assignment, and 17 studies used quasi-experimental research designs, 12 of which addressed standards-based curriculum. Thirty-four studies used some type of quantitative non-experimental research method such as a survey with correlational analyses. There were 33 studies that used a qualitative approach, and 23 studies used mixed methods, with most combining a qualitative method (e.g., interviews) with a quantitative non-experimental method (e.g., surveys). As this description indicates, the great majority of the studies that address the synthesis research questions used non-experimental methods without control or comparisons groups. As a result, a meta-analysis would require the exclusion of most of the studies, and therefore would not serve the goals of the synthesis.

Instead, McREL conducted a systematic review, similar to the review of teacher preparation research reported by Wilson et al. (2001) and commissioned by the U.S. Department of Education. Wilson et al. reviewed empirical studies that addressed five questions concerning teacher preparation. The authors established criteria for rigor using standards associated with the methodology that each study employed. They synthesized and interpreted the studies by describing the findings for each question and by identifying weaknesses and gaps in the available research.

Similar to Wilson et al's (2001) review, the method for the current synthesis reflects the goals of the synthesis, the research problem, and the nature of the studies that met the inclusion criteria. McREL researchers established procedures for reviewing and interpreting study findings and systematically applied these. Researchers coded each study's characteristics and followed a protocol for cross-study analysis. The protocol required researchers to

- identify studies that address the same standards-based variable (i.e., curriculum, instructional guidelines, assessment)

- construct tables that identify and describe the characteristics of each study, including research method, outcomes and measures, and the main findings

- summarize and discuss findings in relation to the research method (i.e., quantitative experimental, quantitative quasi-experimental, quantitative non-experimental, qualitative, and mixed)

- draw conclusions based on the amount, type, and quality of research studies that address the influence of the standards-based variable on teacher instruction and student achievement

- describe implications for practice and policy based on support from the available research

- identify the relationships between standards-based variables and teacher and student outcomes that need additional research.

CHAPTER 3: THE INFLUENCE OF STANDARDS-BASED CURRICULUM ON TEACHING AND STUDENT LEARNING

Education reform initiatives over the last several decades have grappled with the problem of how to develop a school environment in which every child succeeds, and a great deal of time and faith have been put into creating a standards-based education system. Clear and comprehensive high expectations for students have been captured in national, state, and district content and performance standards (Gaddy, Dean, & Kendall, 2002). Content and performance standards stipulate the knowledge and skills that students should learn and master, but standards also provide the foundation for other factors in a standards-based education system, including resources, professional development, assessment, and curriculum (Lachat, 1999). From standards, educators are able to develop curriculum that reflects what students should be taught in the classroom.

A standards-based curriculum is one that reflects or is aligned with national, state, and district content and performance standards (McLaughlin & Shepard, 1995; NCREL, 2000). Student learning in the context of a standards-based curriculum means that the curriculum is based on the same expectations for all students (Lachat, 1999). Although, as mentioned previously, standards vary considerably in their relative orientation between important basic skills and higher-order thinking, standards-based curricula are student-centered, integrated around real world tasks and require students to engage in reasoning, problem solving, and communication — particularly if they are based on national standards like the NCTM or the NSES (Lee, 1998; Lachat, 1999; Schoenfeld, 2002; Weiss, Pasley, Smith, Banilower, & Heck, 2003). Some research suggests that a curriculum aligned with standards is associated with improved student achievement (Isaacs, Carroll, & Bell, 2001). Furthermore, standards-based reform works best when curriculum, assessment, professional development, and instruction are aligned with standards as part of a coherent system (Schoenfeld, 2002).

Teachers are ultimately responsible for implementing a curriculum in the classroom; what students learn is influenced by how and what they are taught (NSES, 2005). Therefore, it is important to understand how standards are implemented at the classroom level, and how this translates into teacher instruction (content coverage and pedagogy) and student achievement. Standards-based education is more successful when teachers

have the capacity to understand what standards mean in terms of the curriculum and their pedagogy and when they are able to engage their students in activities that promote learning (Manouchehri & Goodman, 1998; McREL, 2000). Examining the research literature on standards-based curricula can lead to a better understanding of how to implement these curricula, how they can effectively guide teacher instruction, and how they can lead to favorable learning experiences for students.

This chapter examines research and evaluation studies of two types of curriculum influences on teacher instruction and student achievement; a specific standards-based curriculum in which standards are translated into specific content with suggested pedagogical methods (Adams, Brower, Hill, & Marshall, 2000), or broader standards documents that provide clarity about what to teach in a content area, without suggested pedagogy (Firestone, Camilli, Yurecko, Monfils, & Mayrowetz, 2000). For instance, specific standards-based curricula might include a package such as *Everyday Mathematics*. On the other hand, studies of broader standards documents might examine the influence of the Kentucky State Standards.

In this chapter, studies of the influence of standards-based curricula on teacher instruction and student learning are organized by content area as follows: (1) mathematics and mathematics/ science, (2) science only, and (3) in language arts, social studies, and other content areas. An additional section addresses studies reviewed in other chapters of this synthesis in which curriculum was not the primary focus of the study but which was examined as a study variable. Unless otherwise noted, studies reviewed in this chapter that address the influence of standards-based curriculum on teacher instruction focus on changes in teacher pedagogy. Content coverage is assumed with adoption of the curriculum.

OVERVIEW OF STUDIES

Using the criteria described in Chapter 2, 48 studies were identified for inclusion in this chapter. Detailed descriptions of each study, including methods, sample, measures and relevant findings, can be found in Appendix C.

The studies came from a variety of sources: 21 journal articles, 6 conference presentations, 17 technical reports, 3 dissertations, and 1 book chapter. Of the 48 included studies, 26 examined mathematics and mathematics/science curricula, 11 focused on science curricula, four examined language arts, two investigated history, and five looked at more than one content area or at standards policies in general. The studies included standards influences at varying levels. Twenty-one of the studies addressed national standards, 11 studies were related to state standards, and 16 addressed more than one level of standards, that is, some combination of national, state, and district standards.

The studies were conducted across the span of grade levels: 14 studies of elementary grades (K–5), 10 studies of middle school grades (6–8), and 9 studies of high school grades (9–2). Fifteen studies addressed more than one of these grade spans.

Thirty studies measured student learning and achievement. The majority of studies measured student achievement with a state assessment (n=10), a standardized test (n=10) or a researcher/curriculum developed test (n=8). Qualitative measures of student learning included classroom observations, teacher and administrator interviews, and teacher and student self-reports. Twenty-nine studies measured teacher instruction, mostly pedagogy. The majority of studies examined teacher instruction through teacher and administrator surveys (n=14), teacher and administrator interviews (n=14), and classroom observations (n=14). Other measures included teacher logs and document review. Many studies used multiple measures to assess student achievement and teacher instruction.

Studies varied in the type of methodology employed. None of the studies used an experimental research design with random assignment of students to different curricula, but 12 studies used a quasi-experimental design in which two or more groups of students who were learning different curricula were compared. Eight studies used a quantitative non-experimental methods such a survey design 12 employed qualitative designs, and 16 used mixed methods that combined quantitative and qualitative methods Among the mixed methods studies, four included a quasi-experimental design, and the remaining 12 studies used a non-experimental design. Seventeen of the studies included at-risk students. At-risk identification was based on low socioeconomic status (SES) in 10 studies. Additional indicators of at-risk used in these studies were low-performing (n=6), minority status (n=9), special education (n=2), and English language learners (n=3). Across the studies, 17 referenced systemic influences. The majority of systemic factors addressed were resources, accountability policies, administrator leadership, educator expectations, and professional development.

STUDY FINDINGS

In this chapter, the synthesis authors report on studies of standards-based curricula organized by content area, and examine influences of standards-based curricula on teacher instruction and student achievement. The first two content areas, 1) mathematics and combined mathematics/science, and 2) science, account for the vast majority of studies and their description has been organized by research design. The third general category is organized around the content area of the study and includes studies in history, mathematics/literacy, and literacy. More attention is given to studies in each area that had strong research designs allowing for some analysis of association (e.g., quantitative quasi-experimental), studies with careful descriptions of implementation, detailed quantitative and qualitative analyses, and those studies that had noteworthy features. An abbreviated study table with salient details of included studies is provided in each

subsection. The chapter concludes with implications for practice and policy and a summary of important findings across all the included studies of standards-based curricula.

Standards-Based Curricula in Mathematics and Mathematics/Science

Anderson (1995a, 1995b, 1996) identified several common themes among standards-based science and mathematics reforms, particularly those based on the recommendations of the NCTM and the NRC. First, both include an emphasis on students learning to think in more complex ways, understanding that there can be multiple solutions to problems, and being able to address uncertainties in science. Second, students in mathematics and science reforms are described as active participants in their learning, rather than passive learners. Third, information is presented in a way that supports higher levels of understanding, concepts are studied at a deeper level, and information is not presented in isolation but integrated across subjects. Traditional courses are more focused on mastery of skills, while standards-based courses that reflect a reform orientation focus more on concepts and problem solving. For example, traditional mathematics courses focus on arithmetic, algorithms for whole numbers, common fractions, decimals, algebra in high school, and more pen and pencil computation. On the other hand, reform-oriented mathematics courses concentrate more on geometry, measurement, descriptive statistics, and the use of calculators and computers (Schoen, Fey, Hirsh, & Coxford, 1999).

Table 3.1 lists the 26 studies that addressed mathematics and mathematics/science standards and/or curricula. Nineteen of these studies examined specific mathematics curricula, and one study examined a specific mathematics/science curriculum. Nine specific curricula were identified across these 20 studies (e.g., *Everyday Mathematics, Connected Mathematics, PALMS*). The remaining six studies in this section examined the influences of national, state, and/or district standards on mathematics and mathematics/science instruction and/or student achievement. Overall, the studies covered the range of K–12 grade levels. Of the 17 studies that measured student achievement, the majority used some type of achievement test. Of the 15 studies that measured teacher instruction, teacher surveys, teacher interviews, and classroom observations were the predominant measures. The research designs of the studies were quasi-experimental (n=9), quantitative non-experimental (n=4), mixed-methods (n= 8), and qualitative (n=5)

Table 3.1. Studies of Standards-Based Curricula in Mathematics and Mathematics/Science

Author(s) & Year	Method	Standards Addressed	Subject Area(s)	Outcome(s)	Measures
Adams, Brower, Hill, & Marshall (2000)	Quantitative non-experimental	National: NCTM, NSES State: TX	Math Science	Teacher instruction	Teacher surveys
Anderson	Qualitative	National: NSES,	Math	Teacher	Classroom

Volume 2 (1995a) Volume 3 (1995b) Volume 1 (1996)		NCTM, AAAS, NSTA	Science	instruction Student achievement	observations Teacher, principal, student, & parent interviews Document reviews
Baxter, Woodward, & Olsen (2001)	Qualitative	National: NCTM (1989)	Math	Student achievement Teacher instruction	Classroom observations Teacher reports Teacher interviews
Ben-Chaim, Fey, Fitzgerald, Benedetto, & Miller (1998)	Mixed methods (quasi-experimental)	National: NCTM	Math	Student achievement	Researcher-developed tests Student interviews
Briars & Resnick (2000)	Quantitative quasi-experimental	National: NCTM District: Pittsburgh	Math	Student achievement Teacher instruction	Standardized tests
Cain (2002)	Mixed methods (non-experimental)	National: NCTM	Math	Student achievement Teacher instruction	State assessment Standardized test Teacher & student surveys Teacher interviews Site visits
Carroll (1997)	Quantitative non-experimental	National: NCTM	Math	Student achievement	State assessment
Firestone, Camilli, Yurecko, Monfils, & Mayrowetz (2000)	Quantitative non-experimental	State: NJ	Math Science	Teacher instruction	Teacher surveys Review of lessons
Fuller (2001)	Mixed methods (non-experimental)	National: NSES State: MA	Math Science	Teacher instruction	Teacher surveys
Fuson, Carroll, & Drueck (2000)	Quantitative quasi-experimental	National: NCTM	Math	Student achievement	Standardized tests Researcher-developed tests
Grant & Kline (2001)	Qualitative	National: NCTM	Math	Teacher instruction	Teacher interviews Classroom observations
Hannafin (2002)	Quantitative quasi-experimental	State: MA	Math	Student achievement	State assessment
Haug (1998)	Qualitative	National: NCTM State: CO	Math	Teacher instruction	Teacher & administrator interviews Classroom observations Document reviews

Huntley, Rasmussen, Villarubi, Sangtong, & Fey (2000)	Mixed methods (quasi-experimental)	National: NCTM	Math	Student achievement	Researcher-developed tests Teacher interviews
Johns (2004)	Qualitative	National: NCTM State: MD	Math	Teacher instruction	Teacher interviews Classroom observations Document reviews
Kerr (1999)	Mixed methods (non-experimental)	National: NCTM State: IL District: Chicago	Math	Student achievement Teacher instruction	State assessment Standardized tests Student, parent, & teacher surveys
Lee (1998)	Quantitative non-experimental	National: NCTM State: CA, MN	Math	Teacher instruction	Teacher surveys Document reviews
Michlin, Seppanen, & Sheldon (2001)	Mixed methods (non-experimental)	National: NCTM State: MN	Math	Teacher instruction	Teacher surveys Teacher interviews Classroom observations
Reys, Reys, Lapan, Holliday, & Wasman (2003)	Quantitative quasi-experimental	National: NCTM State: MO	Math	Student achievement	State assessment
Ridgway, Zawojewski, Hoover, & Lambdin (2003)	Quantitative quasi-experimental	National: NCTM	Math	Student achievement	Standardized tests
Riordan & Noyce (2001)	Quantitative quasi-experimental	National: NCTM State: MA	Math	Student achievement	State assessment
Schoen, Cebulla, Finn, & Fi (2003)	Mixed methods (non-experimental)	National: NCTM	Math	Student achievement Teacher instruction	Standardized tests Teacher surveys
Schoen & Hirsh (2003)	Quantitative quasi-experimental	National: NCTM	Math	Student achievement	Standardized tests Researcher-developed performance tests
Weiss, Pasley, Smith, Banilower, & Heck (2003)	Mixed methods (non-experimental)	National: General State: General District: General	Math Science	Teacher instruction	Classroom observations Teacher interviews Teacher surveys
Woodward & Baxter (1997)	Quantitative quasi-experimental	National: NCTM (1989)	Math	Student achievement	Standardized tests

Woodward, Monroe, & Baxter (2001)	Quantitative quasi-experimental	National: NCTM (1989)	Math	Student achievement	Researcher developed tests

AAAS = American Association for the Advancement of Science
NCTM = National Council of Teachers of Mathematics
NSTA = National Science Teachers Association
NSES = National Science Education Standards

Quantitative quasi-experimental studies. Most of the nine studies that used a quantitative quasi-experimental design made attempts to match groups on various variables (e.g., demographics). These studies analyzed influences at the elementary, middle, and high school levels. All nine quantitative quasi-experimental studies examined the influence of mathematics curricula on student achievement, and three of the studies focused on low-achieving students.

Schoen and Hirsh (2003) compared student achievement scores of students in the high school *Core-Plus Mathematics Project* (CPMP), an NCTM-oriented program, to students in traditional mathematics courses. The groups were matched using pretest data. Students were given the Ability to Do Quantitative Thinking (ITED-Q) assessment, a subtest of the Iowa Tests of Educational Development (ITED). and CPMP performance assessments. Data showed that after students completed the first of three CPMP courses, they scored significantly higher than the pre-algebra comparison group and algebra comparison group on the overall ITED-Q test. Results from the CPMP performance assessments showed that CPMP students, after course one, did significantly better on the contextual algebra subtests than comparison students. Comparison students did significantly better on the procedural algebra subtest. After students completed the second CPMP course, CPMP students did significantly better on the coordinated geometry and contextual algebra subtests than the comparison students, and there was no significant difference between groups on the procedural algebra subtest.

At the middle school level, Reys, Reys, Lapan, Holliday, and Wasman (2003) studied two NCTM-influenced mathematics curricula, *Math Thematics* and *Connected Mathematics*. Three districts were matched with three comparison districts on middle school grade organization, prior mathematics achievement, locale, and FRL. Student achievement was measured using the state mathematics assessment. Students had been using a standards-based curriculum for two years. Students in all three districts using a standards-based curriculum scored significantly higher on the data analysis, probability and statistics, and algebra sections of the state assessment than students in the comparison districts. In two of the districts, the students using a standards-based curriculum also scored higher than comparison students in three additional areas, including number sense, geometric and spatial sense, and discrete mathematics. Contrary to critics who argue that students in standards-based mathematics courses are not learning algebra (as cited in Reys et al., 2003), the results from this study suggested otherwise.

Another study of *Connected Mathematics* (CMP) also documented positive changes in student achievement (Ridgway, Zawojewski, Hoover, & Lambdin, 2003). In this study, volunteer samples of CMP students and non-CMP students were matched on ability, locale, and ethnicity. Scores on the Iowa Test of Basic Skills (ITBS) and the Balanced Assessment (BA) for the CMP students and non-CMP students were compared. The ITBS measures more traditional areas of mathematics, such as numbers and operations, while the BA measures a broader range of concepts including mathematical reasoning, problem solving, and connections. CMP students had significantly higher scores than non-CMP students on the BA but not on the ITBS. When scores were analyzed for students who had been exposed to the curriculum for three consecutive years, CMP students were significantly ahead of non-CMP students by the end of year three on both the ITBS and BA.

Two relevant conclusions stem from the Ridgway et al. (2003) study. First, the longer the students were exposed to the curriculum, the more achievement improved on both the traditional measure and the higher-order measure. This suggests that students' gains in mathematical knowledge from the curriculum are likely to be sustained over time. Second, the initial differences in student achievement outcomes for the two assessments call attention to the importance of the achievement measure. Depending on the content and possibly the format of the achievement measure, it may be better aligned with a traditional approach to content, or it may reflect a reform orientation. So the test may interact with the reform/traditional orientation of particular standards and/or a particular curriculum in determining whether it is successful or not in raising achievement.

Additional studies have found that the greater the fidelity of implementation, the stronger the influence on student achievement (Riordan & Noyce, 2001; Briars & Resnick, 2000). Briars and Resnick compared fourth-grade mathematics achievement scores on standardized tests of weak implementer schools and strong implementer schools using *Everyday Mathematics*, an NCTM-oriented curriculum. Weak implementer schools were defined as schools where the majority of teachers were not using *Everyday Mathematics* in its entirety, and strong implementer schools were defined as schools where teachers were using all of the *Everyday Mathematics* components and provided more student-centered pedagogy. There were significant differences in student mathematics scores between strong and weak implementer schools, favoring strong implementer schools. Briars and Resnick also found that fourth-grade student performance increased more when students were taught using a cohesive standards-based approach including the *Everyday Mathematics* curriculum, assessment, and professional development. In addition, the achievement gap between white and African American students was closed in the strong implementer schools.

Woodward and Baxter (1997) examined the achievement of low-performing students using *Everyday Mathematics*. Mathematics scores from the ITBS and scores from the

Informal Mathematics Assessment (IMA) were used to compare the achievement of third-grade students using *Everyday Mathematics* with students in comparison classes. The groups were comparable on SES and demographics. At-risk students were identified in both groups based on IEP (individual education plan) status or low scores on the ITBS. Overall, students in the intervention group scored significantly higher on the concepts subtest of the ITBS than the comparison students. But when students were categorized by ability level, there was no significant difference on the ITBS between low-performing students in the intervention group and low-performing students in the comparison group, although average- and high-performing students in the intervention group both scored significantly higher on various subscales than their counterparts in the comparison group. Low-performing students in the intervention group showed a modest improvement on their problem-solving abilities at the end of the school year, while low performing students in the comparison group showed no improvement. This study suggests that a standards-based curriculum can have positive influences, although to a lesser extent, on the student achievement of low-performing students compared to average- and high-performing students. The authors suggested that small, homogenous groups offer the best setting for low-performing students when using reform-oriented pedagogy, such as student discourse and student-centered activities.

Hannifin (2002) conducted a study of *PLATO*, a high school remedial mathematics program. Pretest and posttest scores were compared using state assessment data for students in *PLATO* and a convenience sample of students who were not in the program. While both groups showed a significant increase in test scores, and the PLATO group posted significantly larger increases than the comparison group, the comparison group's average was still significantly higher than the average for *PLATO* students at the end of the study. It should be noted that the groups were not matched on any type of variable, and the analysis did not control for pretest scores

Mixed methods studies. Eight studies used a mixed-methods approach that included both qualitative and quantitative methods. Six studies used non-experimental quantitative methods (e.g., surveys) as well as qualitative methods (e.g., interviews and classroom observations). Two studies used a quasi-experimental design in which qualitative and quantitative data were collected from a treatment and control group. Six studies addressed teacher instruction and five addressed student learning. Six of these studies focused on mathematics and two studies examined mathematics/science.

The combined findings of two studies (Cain, 2002; Michlin, Seppanen, & Sheldon, 2001) corroborate the positive relationship between the time invested in implementing a curriculum and student achievement (Ridgeway et al., 2003). Similarly, it takes time for teachers to become comfortable with curriculum materials and change their instruction, although the idea of standards-based education may emerge earlier in their knowledge and awareness. When teachers were asked about their perceptions of standards-based

curricula, their responses were generally positive (Kerr, 1999; Fuller, 2001; Cain, 2002). For example, a study of *Partners Advancing Learning of Math and Science* (PALMS) — an NCTM/NSES-oriented curriculum, found that teachers perceived significant improvement in their classroom management, classroom culture, reflective teaching, and resource utilization after using PALMS for six years. On the other hand, some teachers have reported that a standards-based curriculum is sometimes challenging to teach and can increase the time required for preparation, teaching and remediation. They sometimes report feeling stress due to the curriculum's fast pace (Fuller, 2001, Cain, 2002).

Other studies have provided evidence of an association between NCTM-oriented standards-based curricula and changes in teacher instruction. Teachers implementing one such standards-based curriculum reported using more reform-oriented instructional practices, such as having students explain their answers and addressing alternative solutions in their teaching (Michlin, Seppanen, & Sheldon, 2001). Teachers also have reported that they feel their own content knowledge has improved due to teaching another standards-based curriculum (Cain, 2002).

Weiss, Pasley, Smith, Banilower, and Heck (2003) conducted classroom observations, teacher interviews, and a follow-up teacher survey with a nationally representative sample of K–12 mathematics and science teachers to examine what mathematics and science instruction looks like in the classroom and to identify the factors that influence teachers' instruction. Teachers indicated that the content of what they taught in a lesson was influenced by state or district curricula (74%) but not national standards (1%). Factors that influenced teachers' choice of the instructional strategies used in a lesson were teacher knowledge, beliefs, and experience (90%); class textbooks/programs (71%), student characteristics (52%), district-provided professional development (31%), collegiality (18%), principal input (8%), and to a lesser extent, state/district standards/frameworks (5%) and accountability tests (7%). No teachers reported that their instruction was influenced by national standards, although it is possible that these standards were reflected in local standards frameworks. With the number of external influences that have bearing on teachers' instruction, the researchers stressed that teachers need a coherent set of messages and clear goals to guide their instructional choices. Further, classroom observations revealed high-quality lessons (as assessed on a variety of reform-oriented indicators) had learning environments that engaged student thinking and helped students make sense of mathematics and science content. Interestingly, rural districts and those with a high percentage of minority students had lower-quality lessons.

Schoen, Cebulla, Finn, and Fi (2003) identified characteristics of high school teachers' instruction that were associated with student achievement in schools using *Core-Plus Mathematics Project* (CPMP), a curriculum associated with the NCTM standards. The Ability to Do Quantitative Thinking, a subtest of the ITED, was administered as a pretest

and posttest. Teachers completed an implementation survey and a concerns survey and classroom observations were conducted. Regression techniques were used and several teacher instruction variables were shown to be significantly and positively related to growth in student achievement. These variables included professional development on how to teach the CPMP course, more group and pair work among students, less teacher presentation and fewer whole-group discussions, less time spent on non-academic activities, the use of a variety of assessment strategies, the use of instructional and assessment materials provided by the curriculum, and teacher collaboration. Student demographics did not explain any of the variance in student achievement nor did the length of the class period.

Another study of the *Core-Plus Mathematics Program* assessed high school students' algebraic understanding and skills and their problem-solving abilities (Huntley, Rasmussen, Villarubi, & Fey, 2000). CPMP students and control students were matched using standardized test scores from the previous year. Assessments were developed to measure contextual problems, symbolic manipulations with no contextual information, and open-ended contextual problems. Interview data with teachers indicated that there was great variability in the implementation of the traditional and the standards-based curricula. Even so, CPMP students performed better on tasks that included contextual information and tasks for which students could use graphing calculators. Control students performed better on traditional symbol manipulation tasks when there was no contextual information and no calculator use.

With respect to student achievement, student performance on rate, density, ratio, and scaling problems were assessed by comparing students using *Connected Mathematics*, an NCTM-oriented curriculum, and students using a traditional curriculum (Ben-Chaim, Fey, Fitzgerald, Bendetto, & Miller, 1998). A treatment group and control group consisting of seventh-grade students were identified and assessed on proportionality problems. Twenty-five percent of the students were also interviewed about their reasoning and thinking regarding the mathematics problems. CMP students outperformed non-CMP students on all of the problems. CMP students were also more proficient in their writing and used a wider variety of problem-solving strategies than non-CMP students. Students' increased use of effective problem-solving strategies was an important finding in this mixed-methods study. Findings from two additional studies using student survey data indicated that students perceived their standards-based mathematics curriculum as positive and helpful in learning problem-solving strategies (Cain, 2002; Kerr, 1999).

Importantly, these studies highlight the influence of standards-based mathematics curricula (aligned with the NCTM standards) on students' problem-solving abilities, specifically their abilities to use contextual information in solving problems. However, the findings from Huntley et al., (2000) and Schoen and Hirsh (2003) suggest that

students in traditional mathematics courses perform better on out-of-context problems and paper and pencil manipulations than do students in NCTM-oriented curricula. As Huntley et al. (2000) indicated, students learn what is emphasized in their mathematics curriculum. Consequently, educators need to decide what mathematics content is the most important for students to learn and ensure that their standards, curricula and achievement measures are aligned to reflect that decision.

Quantitative non-experimental studies. Four studies used quantitative non-experimental designs. Two studies focused on mathematics, and two studies focused on a combination of mathematics and science. Three of these studies examined teacher instruction and one study included student achievement.

Carroll (1997) studied schools that were using *Everyday Mathematics* in their third-grade mathematics classes. School scores for *Everyday Mathematics* schools were compared to the state average and the county average on the state assessment. *Everyday Mathematics* schools scored higher on the state assessment compared to the state mean and the mean score for the county. Schools that had implemented *Everyday Mathematics* since kindergarten had a higher mean score than schools that had recently implemented it. Furthermore, the majority of students in *Everyday Mathematics* schools met or exceeded state goals. In addition, when school achievement was assessed by school poverty levels, *Everyday Mathematics* schools with the highest poverty levels still scored above the state and county averages. Again, this study reinforced the finding that a standards-based curriculum is more successful when it is implemented over a longer period of time. Furthermore, while the type of achievement measure was an issue raised in an earlier study, the author of this study maintained that students using the standards-based mathematics curriculum were able to transfer this knowledge to more traditional achievement measures, such as the state assessment. Even so, students using the standards-based curriculum still made more gains in areas of mathematics that are not covered extensively in traditional courses (e.g., geometry) compared to students not using the curriculum.

Firestone, Camilli, Yurecko, Monfils, and Mayrowetz (2000) investigated how the introduction of New Jersey state standards impacted fourth grade mathematics and science teachers' instruction three years after the standards' introduction. A statewide sample of teachers completed telephone surveys and a mailed questionnaire, and provided examples of classroom activities. Researchers identified topics in both mathematics and science that would have been part of the curriculum before the state standards were implemented and topics that were likely added in response to the state standards. Results showed that 15 percent of teachers reported teaching more mathematics and 14 percent taught more science since the state standards were implemented. Teachers reported that they spent more time on traditional mathematics content but also reported an increase in the amount of time spent on new topics in

mathematics. Science teachers reported more equivalent amounts of time spent on teaching traditional science topics and new science topics. Teachers in poorer school districts reported teaching to the test more than teachers in wealthier districts.

Two studies concentrated on state and national standards. In both studies, teacher surveys were used to measure knowledge of state policies, standards, and instruction. Lee (1998) conducted a secondary data analysis using a combination of national surveys. Findings from the study indicated content-driven state policies on mathematics education were not associated with teachers' reported use of mathematics pedagogy consistent with NCTM. Teachers used more standards-based pedagogy when student assessments and textbooks were linked to a standards-based curriculum. Principals perceived that there was an association between teachers' instructional changes and state-level curricula policies; however, this relationship was not statistically significant. The study found a positive influence of new textbooks and a negative influence of district/local tests on the use of instructional practices that were consistent with NCTM. Instruction was likely to be most influenced at the district level, less at the state level, and least influenced by national policies.

In the study by Adams, Brower, Hill, and Marshall (2000), a stratified random sample of mathematics and science teachers in grades four through eight from four Texas regions was identified and surveyed. Overall, survey results indicated that teachers rated pedagogical strategies such as hands-on activities, problem-solving activities, and technology integration, all aspects of an orientation toward the NCTM reforms, as important in creating an ideal classroom. But when asked what strategies they used during the week, teachers reported using more traditional strategies such as seatwork. (The extent to which the Texas standards aligned with reform tenets is unclear; they may very well have emphasized basic skills instruction.) Mathematics teachers applied what they knew about reform-oriented instruction in their classrooms more than the science teachers, with more frequent use of problem solving, while science teachers reported more traditional practices like seatwork. Both mathematics and science teachers reported technology use as an effective instructional strategy but it was one of the least used strategies by teachers. Other seldom-used activities (but recommended by the NCTM) were authentic tasks and inquiry and experimental methods. An important theme from this study is that many teachers, even though they supported the philosophy of standards-based curricula and/or standards, still struggled in changing their instruction to reflect more reform-oriented methods of instruction.

Qualitative studies. Five studies used qualitative research designs. These studies provide detailed information that adds to the findings from previously discussed studies. Of these studies, four focused on mathematics and one on mathematics/science. All five studies examined the impact of standards-based curricula on teacher instruction and two

studies also examined influences on student learning. One study focused on how a standards-based curriculum influenced the learning of low-performing students.

Johns (2004) assessed teachers' understanding of *Everyday Mathematics* and how they implemented it in the classroom. Teachers were divided into three groups based on their comfort and expertise level in mathematics. Classroom observations and teacher interviews were conducted. Teachers who were the least comfortable with mathematics content used the curriculum to enhance their knowledge of mathematics and their instruction. Teachers who were confident in mathematics and those that had specialized training in mathematics were more likely to deviate from the curriculum when there was dissonance between the reform-oriented mathematics curriculum and the way they had been taught mathematics. Even though less knowledgeable teachers deviated less from the curriculum, they still struggled with content knowledge and pedagogical knowledge. This made it difficult for them to understand the curriculum, and, as a result, some of the less knowledgeable teachers would omit lessons and were unable to make important connections in the content. The manner in which the curriculum was delivered in the classroom was ultimately mediated through teachers' beliefs and prior knowledge.

In some instances, standards influenced instructional change consistent with NCTM's reforms. For instance, Haug (1998) conducted case studies of two school districts in Colorado to examine how teachers' instruction was influenced by state education policies in mathematics. While there were many differences across teachers, teachers reported using more problem-solving activities, integrated writing in mathematics, and in the elementary grades, integrated mathematics across subjects. Similarly, in another study, teachers' increased used of manipulatives and pair work was associated with an NCTM-oriented curriculum in mathematics (Baxter, Woodward, & Olsen, 2001). Anderson (1995a, 1995b, 1996) used case studies to determine the influences of mathematics and science reform-oriented curricula in middle-school and high-school classrooms. The teacher's role in the classroom changed from being a director of knowledge to being a coach and a facilitator, and instruction became more student-centered. Students were engaged in the subject matter for a longer period of time, learned content in terms of real world contexts, and were given opportunities to develop and use their thinking skills. While there was much variability in teachers' practice, findings across these qualitative studies emphasize that teacher instruction is changing consistent with the recommendations of reformers.

Baxter, Woodward, and Olsen (2001) focused on how low-achieving students learn in the context of standards-based curriculum and instruction. Five third-grade teachers using *Everyday Mathematics* were observed and interviewed. Low-performing students were identified in each class based on low performance on the mathematics portion of the ITBS or based on IEPs. Results revealed that a typical mathematics lesson included whole group work and pair work. During whole-group discussions, low-performing

students were passive, unengaged, and often off task. During pair work, low-performing students were more engaged but at a lower level and most likely participated in a non-mathematical capacity. A promising practice by one teacher was the use of "ad hoc" groups to focus on a skill or problem with which students were struggling. This usually involved 8 to 11 lower-achieving students. A noteworthy finding from this study was that low- performing students in the "ad hoc" groups tended to be highly involved compared to their low level of participation when they worked with the whole group or in pairs.

Summary of findings from studies of standards-based curricula in mathematics and mathematics/science. Across the areas of mathematics and mathematics/science, the majority of studies focused on curricula aligned with the national content standards (NCTM and NSES). Overall, these studies reported positive findings related to improved student achievement and learning and teacher instruction related to the reforms. All 17 studies that assessed student achievement and learning found a positive relation between use of the curricula and achievement. The majority of studies were quantitative quasi-experimental designs, which allowed comparisons between students who were using standards-based curricula (generally reform-oriented) and students who were using more traditional curricula. In these studies, students using the standards-based curriculum scored higher on achievement tests than students in comparison groups, although four studies did show mixed results. While several studies mentioned possible issues regarding the type of achievement measure used (e.g., whether it targets traditional or reform-oriented content), findings from other studies suggest that knowledge acquired from a reform-oriented standards-based curriculum does transfer to performance on traditional measures. Findings also suggested that student learning increases when a curriculum is implemented over consecutive years. Additionally, findings showed that standards-based curricula are influential for a range of students, including minority students, low-performing students, and low SES students.

Standards-based curricula can change teacher instruction, although eight of the 15 studies presented mixed results on this outcome. Positive findings indicated that standards-based curricula can motivate and help teachers to change their pedagogy so it more closely reflects the recommendations of the NCTM. Teachers using these curricula were more likely to have students explain their answers, allow for multiple solutions to problems, incorporate more problem solving activities into their classrooms, use more pair work and spend less class time on presentation and whole group work than teachers using traditional curricula. On the other hand, a notable finding is that many teachers expressed knowledge about NCTM-oriented practices but struggled with using them in their classroom. Some teachers also said that they experienced more stress and increased preparation time when implementing standards and standards-based curricula, compared with traditional curricula. Finally, implementation of a standards-based curriculum alone did not influence changes in teachers' instruction unless student assessments and

textbooks were aligned with that curriculum — indicating that systemic support is important for changing pedagogical practice.

Studies of Standards-Based Curricula in Science

This section describes the eleven studies that met the inclusion criteria and addressed solely science standards and/or curricula. The studies are listed in Table 3.2. Thompson, Zeuli, and Borman (1997) traced the beginning of the national science reform movement to the publication of the AAAS *Science for All Americans* (1989) followed by the AAAS's *Benchmarks for Science Literacy* (1993), the NSTA *The Content Core* (Pearsall, 1993) and the NSES (1996).

As mentioned previously, the national standards-based science and mathematics reforms share several themes (Anderson 1995a, 1995b, 1996). These include teaching students to think in complex ways, which allow for multiple solutions to problems and uncertainties, making students active participants in their learning (e.g., student-centered learning and inquiry-based learning); and presenting lessons in ways that support deeper levels of understanding embedded in experience (e.g., project-based learning or integrating information across subjects).

Of the 11 studies about science, seven concerned national standards, one concerned state standards, and three addressed a combination of two or more levels — national, state or district standards. Nine of the studies in this section examined the influence of exposure to standards-based curricula and/or materials. Six of these studies addressed the influence of a specific curriculum in science (e.g., *Foundations of Science, JASON Project, Big Things*, and *Scope Sequence and Coordination*), and three addressed the influence of curriculum units or materials developed to align with science standards at the district, state, or national level. In contrast, the remaining two studies examined more general influences of standards; one study examined the influence of a NSF SSI on instruction, and one study examined the influence of state standards on instruction. None of the studies were conducted exclusively at the elementary level, five studies focused on middle school, three focused on high school (9–12), and three included more than one grade level span. Of the seven studies that measured student achievement, none used a state assessment as the achievement measure, one used a standardized test, three used either a test developed by the researcher or curriculum developer, one used student work, and two used a combination of measures. Of the seven studies that measured teacher instruction, two used teacher surveys, one used classroom observations, and four used a combination of two or more measures.

Table 3.2. Studies of Standards-Based Curricula in Science

Author(s) & Year	Method	Standards Addressed	Subject Area(s)	Outcome(s)	Measures
Ba, Martin, & Diaz (2001)	Mixed (quasi-experimental	National: NSES	Science	Student achievement Teacher instruction	Researcher developed pre/post tests Videotaped student presentations Classroom observations Teacher & principal interviews Teacher & student surveys
Ba, Admon, & Anderson (2002)	Quantitative, non-experimental	National: NSES	Science	Teacher instruction	Surveys
Chance & Anderson (2003)	Mixed (non–experimental)	State: NV	Science	Teacher instruction	Surveys
Goldenberg, Ba, Heinze, & Hess (2003)	Mixed (non–experimental)	National: NSES	Science	Student achievement Teacher instruction	Interviews Observations, Teacher surveys Teacher logs Student surveys Student content activity
Lawrenz, Huffman, & Lavoie (2001)	Mixed (quasi–experimental)	National: NSES	Science	Student achievement Teacher instruction	Student & teacher surveys & interviews Classroom observations Researcher developed tests Laboratory experiments
Marx, Blumenfeld, Krajcik, Fishman, Soloway, Geier, & Tal (2004)	Quantitative, non–experimental	National: AAAS, NSES	Science	Student achievement	Curriculum developed tests
Parker & Gerber (2000)	Mixed (non – experimental)	National: NSES State: GA	Science	Student achievement	Researcher developed test Teacher logs Student surveys
Rivet & Krajcik (2004)	Quantitative, non-experimental	National: AAAS, NSES State: MI District: Detroit	Science	Student achievement	Researcher developed pre-post tests
Schneider,	Qualitative	National: AAAS,	Science	Teacher	Classroom

Krajcik, & Blumenfeld (2005)		NSES		instruction	observations
Schneider, Krajcik, Marx, Soloway (2002)	Quantitative, quasi-experimental	National: AAAS, NSES	Science	Student achievement	Standardized tests
Thompson, Zeuli, & Borman (1997)	Qualitative	National: AAAS, NSES State: MI	Science	Teacher instruction	Classroom observations Interviews

AAAS = American Association for the Advancement of Science
NSES = National Science Education Standards

Quantitative quasi-experimental study . McREL researchers identified only one quantitative quasi-experimental study that used a national standardized test as a measure of student achievement in science. Schneider, Krajcik, Marx, and Soloway (2002) examined the influence of a three-year integrated project-based science curriculum, *Foundations of Science Project*, on high-school student achievement on the National Assessment of Educational Progress (NAEP) science test. The study compared student achievement scores of 42 primarily white and middle-to-upper-SES 10^{th} and 11^{th} grade students who participated in the project with three national NAEP comparison samples (including the overall national sample, a sample matched by being not eligible for FRL and a white sample). Results indicated that students exposed to *Foundations* scored significantly higher than all three comparison samples. Item analyses showed that the *Foundations* students scored significantly higher on about half of all the items compared to the three comparison samples. Compared to the overall national NAEP sample, *Foundation* students scored significantly higher on a greater percentage of constructed response items than on the multiple choice items, and significantly higher on most of the scientific investigation items and on more than half of the conceptual understanding items. Overall, this study found a positive relationship between an NSES-oriented, standards-based curriculum and student achievement. The researchers concluded that students in integrated project-based science courses are not disadvantaged on standardized achievement tests.

Mixed methods studies. Five studies used a mixed-methods approach that employed both qualitative and quantitative methods. Two of these studies incorporated a quasi-experimental design and the remaining three studies included non-experimental quantitative methods.

In the first-year report of their three-year evaluation of the *JASON Project*, a multimedia science curriculum emphasizing project-based learning, technology and inquiry, Ba, Martin, and Diaz (2001) incorporated a quasi-experimental design to study student

achievement. The researchers used comparison classes (matched for age and general ability level) in two of the eight treatment schools to compare student learning on a researcher-developed inquiry-based measure. They found a positive effect of the curriculum on overall student gains, with more *JASON* students showing gains (69%) than comparison students (20%). Across all eight *JASON* schools, 66 percent of *JASON* students made overall gains, with more making gains in the process (67%) than the content (46%) portion of the test. *JASON* teachers reported incorporating more hands-on student learning activities, technology, and reform-based assessment since implementing *JASON* and reported that the *JASON* curriculum made project-based learning easier to implement.

Lawrenz, Huffman, and Lavoie (2001) conducted a six year study of *Scope Sequence and Coordination,* an inquiry-based ninth- and tenth-grade science curriculum aligned with the NSES. In this study, measures of non-treated students the year prior to implementation were used as the comparison for the next five years of the study. Although 13 schools originally participated, five representative schools were chosen to be case study sites and were followed over time. Only three of the five sites were still using the reform in year six, and one of these reported making major modifications, suggesting uneven levels of implementation across the schools. Despite this, observations and interviews showed an increase in the amount of time spent on student-centered teaching in the five schools: from 10–25 percent to 40–90 percent. As expected, the two sites that had continued using the original reform showed more gains in time spent on student-centered activities. All students completed an annual took a researcher-developed test and a random sample completed laboratory tests. The achievement results were mixed and inconclusive. Although the researchers found trends towards improved student scores with longer exposure to the curriculum, they concluded that there were no consistent effects on achievement, perhaps due to problems with implementation in several schools.

The remaining three mixed-methods studies used qualitative and quantitative non-experimental methods. Two of these studies examined student achievement and found positive influences of standards-based curricula. Of the two studies that also examined teacher instruction, one had positive results and the other had mixed results. In the first study, which examined student achievement, Parker and Gerber (2000) found evidence of student learning on a researcher-developed pre-post test after ten two-hour sessions of NSES-based science program. Pre-post test gains were significant for all 11 African American fifth- and sixth-grade students in the sample.

In the third year evaluation report on the *JASON Project* (which corroborates the first year report discussed previously in this section), Goldenberg, Ba, Heinze, & Hess, (2003) also found positive influences of a standards-based curriculum on student achievement and teacher instruction. Using content activity to measure student learning, the researchers found that students demonstrated good content knowledge, a sophisticated

science vocabulary, and asked relevant questions. Teacher logs and student and teacher interviews and surveys confirmed that the curriculum increased students' acquisition of science knowledge. In terms of teacher instruction, teachers reported incorporating more hands-on activities into their classrooms and participating in increased collaboration with other teachers as a result of the *JASON project*.

In contrast, Chance and Anderson (2003), in their survey study of 195 secondary science teachers, found minimal influence of standards on instruction. Only 42 percent of respondents reported that Nevada's science standards impacted their instruction. It is important to note that this study, in contrast to most of the others in this section, examined the influence of state standards in general and was not examining the influence of specific curricula, units, or materials aligned with national science standards. It is possible that the variation in standards may account for some of the variance in these findings.

Quantitative Non-Experimental Studies. Three science studies met the inclusion criteria and used solely quantitative non-experimental designs. Two of these studies looked at the influence of standards-based curricula on student learning, and one study examined teacher instruction. It is notable that all three studies had large sample sizes.

Rivet and Krajcik (2004) found a positive influence of an eight-week NSES-based science project called *Big Things* on at-risk students' learning as measured with a researcher developed pre-post test. A sample of more than 2,500 6th grade students in 15 schools in lower-SES and largely African American neighborhoods participated. Students showed statistically significant learning gains overall from pretest to posttest in all four years of the project (effect sizes ranged from 1.36 to 1.61), and on each of the four learning goals of the project. Marx et al. (2004) also found significant pre-post test gains for at-risk students. In this study, 8,000 sixth through eighth- grade students in Detroit were exposed to four eight-to-ten week AAAS/NSES-based science curriculum units over the course of three years, and were given a pre-post unit assessment comprising a process test and a content test. There were statistically significant gains for seven of the eight content and process tests. Effect sizes were stronger for content scores than process scores and increased over the three years.

The second year evaluation study of the *JASON Project* also found a positive influence of the project on instruction and corroborated the findings of the first and third year reports referenced previously (Ba, Admon, & Anderson, 2002). In this survey study, 1,886 *JASON* teachers, reported that *JASON* helped them meet national/local standards (88%) and that they used the *JASON* tools, techniques, and group research activities (over 70%).

Qualitative studies. Two science studies used qualitative research designs to investigate the impact of standards and/or standards-based curricula on teacher

instruction. Both found mixed results. Schneider, Krakcik, and Blumenfeld (2005) examined the implementation of an eight-week project-based science unit, linked to the national science standards, in four Midwestern urban schools by four eighth-grade teachers. Their classrooms were observed and videotaped and rated for seven instructional categories. Half of the teachers received high ratings consistent with the intent of the curriculum, and the other half received medium to low ratings. The researchers suggested professional development and instructional guidance may have been insufficient to support implementation. Thompson, Zeuli, and Borman (1997) studied the influences of the Michigan SSI, which highlighted reform-oriented science practice, on a sample of 32 elementary and middle school teachers in Michigan. Although many of the teachers they interviewed reported that their teaching reflected reform practices, classroom observation data suggested otherwise. The researchers reported that only two of these teachers truly used the reform ideas in their instruction, a finding paralleled by other researchers in mathematics (Spillane & Zeuli, 1999) and described more fully in Chapter 4. These findings indicate that the changes in pedagogy envisioned in the national mathematics and science standards reforms may require considerable time and investment in teacher learning opportunities.

Summary of findings from studies of standards-based curricula in science. The science studies reviewed for the synthesis indicated a positive relationship between science standards and standards-based science curricula on student achievement and student learning. Six of the seven studies that included student learning as an outcome variable showed positive effects of the science standards and/or curricula on student test scores or student work, while the seventh found no consistent findings. Two of the three studies incorporated a quasi-experimental comparison group for their student achievement measure and found positive associations with student achievement. The quasi-experimental study that used a national standardized test (NAEP) as a measure of student achievement found positive results when the treatment group was compared to matched national NAEP samples (Schneider et al., 2002). A study with variable implementation of the curriculum found no significant student achievement results; nevertheless, researchers found trends towards improved student scores with longer exposure to the curriculum (Lawrenz, Hoffman, & Lavoie, 2001). Few studies in this section examined an interaction between participation in a standards-based curriculum and item format on achievement tests, although Schneider et al. (2005) found an interesting trend, with participating students scoring significantly higher on constructed-response items than on multiple-choice items, and higher on scientific- investigation and conceptual-understanding items compared to practical-reasoning items.

The reviewed studies indicated mixed effects of science standards and standards-based science curricula on teacher instruction. Four of the seven studies that included teacher instruction as an outcome variable showed a positive relation between the science standards and/or curricula and pedagogy consistent with national science standards

reforms, while the remaining three had mixed results for this outcome. The three studies reporting mixed findings included the only study of the impact of general state standards, which, as already noted, may vary in nature from the national science education standards, and the only study of the impact of a statewide systemic initiative on teachers from across the state (Thompson et al., 1997; Chance & Anderson, 2003). This suggests that the impact of standards on instruction is dependent on the nature of the standards and may be greater when mediated by a specific NSES-oriented curriculum implemented in individual schools. In the remaining study with mixed findings, the researchers suggested that professional development and instructional guidelines were insufficient for full implementation at the study site (Schneider et al., 2005). These studies suggest that any impacts of state standards and statewide initiatives on teacher instruction are mediated by the development of curricula, materials, and instructional guidelines aligned with those particular standards and supported through professional development.

Studies of Standards-Based Curricula in Other Content Areas

The studies of standards-based curricula studies reviewed thus far have focused on mathematics and/or science, which is not surprising considering the strong and early influences of the NCTM and NSES standards. Still, a total of eleven studies that met the inclusion criteria addressed other content areas, and these studies are listed in Table 3.3. Four studies concerned language arts, and of these, only one focused on a specific curriculum (*Signatures).* Two studies addressed history standards, and one study did not focus on a content area but rather on how standards in general were being taught to special-needs students. The remaining four studies examined the influences of national, state, and district/local standards on mathematics/language arts. Only one of the studies concerned national standards, eight studies were related to state standards, and two addressed more than one type of standard. Four studies were conducted at the elementary level (K–5), none focused solely on middle school grades (6–8), two focused on high school grades (9–12), and five addressed more than one of these grade spans. Of the six studies that measured student achievement, four used the state assessment as the achievement measure, one used a standardized test, two used district-level tests, one used a curriculum-developed assessment, and one used a researcher-developed test. Of the seven studies that measured teacher instruction, three used teacher surveys, five used classroom observations, six used teacher interviews, and one used teacher logs. The research designs of the studies were quasi-experimental (n=2), quantitative non-experimental (n=1), mixed-methods (n= 3), and qualitative (n=5).

Table 3.3. Studies of Standards-Based Curricula in Other Content Areas

Author(s) & Year	Method	Standards Addressed	Subject Area(s)	Outcome(s)	Measures
Caron (2002)	Mixed methods (non-experimental)	State: IN	LA	Teacher instruction	Teacher surveys Teacher interviews
Goertz, Floden, & O'Day (1995a, 1995b, 1996)	Mixed methods (non-experimental)	State: CA, MI, VT	LA Math	Teacher instruction	Teacher surveys Teacher interviews
Grisham & Brink (2003)	Qualitative	State: WA	LA	Teacher instruction	Teacher interviews Classroom observations
Indiana University (1998)	Quantitative non-experimental	National: National Council of Teachers of English State: unspecified	LA	Student achievement	Standardized test Curriculum developed assessment
Jerome & Gilman (2003)	Quantitative quasi-experimental	State: IN	LA	Student achievement	State assessment
McLaughlin (2000)	Qualitative	State: Unspecified	General	Teacher instruction	Classroom observations Teacher interviews
Sandholtz, Ogawa, & Scribner (2004)	Qualitative	State: CA District	LA Math	Student achievement Teacher instruction	State assessment District test Classroom observations Teacher interviews Teacher surveys
Smith (2003)	Qualitative	State: VA	SS	Teacher instruction	Teacher interviews Classroom observations Social studies department meeting observations
Speas (2003)	Mixed methods (non-experimental)	State: NC	LA Math	Student achievement	State assessment WCPSS Effectiveness Index (District measure) Curriculum-developed assessment
Speas (2004)	Quantitative quasi-experimental	State: NC	LA Math	Student achievement	State assessment

VanSledright (2002)	Qualitative	National: National Center for History in the Schools	SS	Student achievement Teacher instruction	Researcher developed assessment Lesson plans examined Lessons videotaped Classroom observations Teacher journal

Studies of standards-based history curricula. McREL researchers found two qualitative studies that looked specifically at history standards. VanSledright (2002) focused his study on a class of fifth grade students that he taught for four months. He adjusted his instruction to follow guidelines from the National Center for History in the Schools and examined the influences of these changes on student learning. To evaluate student learning, he selected a representative sample of students in the class to complete a pre-post performance task. The performance task consisted of the students reading historical documents and viewing at historical images. Their responses were analyzed based on a coding scheme that rated the complexity of the student's historical thinking ability. On the initial performance task, most of the students' responses were rated at the low level, but by the end of the study, students used more of the higher-level strategies, showing a greater ability to think critically and express their interpretations of events more clearly. Actual gains between performance tasks varied among students. Those who started out as low-level readers did not make as much progress as the high–level readers, indicating the importance of literacy skills for success across the content areas. VanSledright stressed the importance of individualization of instruction for students at different levels. He also noted that pedagogically speaking, this teaching approach required more preparation time and actual classroom time, which is something that many elementary teachers do not have.

Smith (2003) also examined changes in teacher instruction related to history standards. The study focused on the impact of the Virginia Standards of Learning (S.O.L) history curriculum on five high school teachers in one suburban high school. When interviewed, teachers reported no change in their instruction in response to the standards-based history curriculum. However, based on multiple classroom observations over the course of a school year, the researcher found changes in instruction from student-driven discussions at the beginning of the year to a teacher lecture format later in the year. Teachers resorted to "cramming" as the test date drew closer in an attempt to ensure that all S.O.L materials were covered. Teachers also complained of a narrowing of the curriculum because the S.O.L history curriculum materials lacked the viewpoints of women and minorities.

Studies of standards-based mathematics/language arts curricula. There were four studies that examined both mathematics and language arts standards. Goertz, Floden,

and O'Day (1995a; 1995b;1996) conducted a mixed-methods study using teacher interviews and questionnaires that included questions from three nationally-representative teacher surveys, which drew on a purposive sample of teachers from 12 reforming schools across three states. The researchers explored the impact of state mathematics and language arts standards on teacher instruction. They found that in mathematics, instruction shifted at the elementary level from "whole-number computation, estimation, and number facts" (1996, p. 87) to an emphasis on problem-solving and mathematics discussion, a trend consistent with the state standards. In the middle-school classrooms, teachers reported a general decrease in the amount of time that students were engaged in whole-number computation and an overall increase in algebra work. In language arts, teachers in all three states adjusted their instruction to align with topics emphasized by the state standards; however, they indicated a desire to mix content from the past with the new standards-based content in order to round out their instruction.

In a qualitative case study by Sandholtz, Ogawa, and Scrubner (2004), researchers compared students' state assessment data with the district test and found a gap between the expectations expressed in district and state standards. The district's curriculum standards were lower than the state standards, sometimes by as much as one grade level. The district standards had three different levels of standards (minimum, essential, and accelerated) based on the academic level of the student. If a classroom consisted of students at a low academic level, the teachers focused their instruction on meeting the minimum standard level, thus setting low expectations for student learning. The teachers reported using more drill, practice, and routine teaching strategies with less time devoted to hands-on learning activities, in-depth instruction or supplemental work. Ninety percent of teachers reported that they focused on basic skills. Student scores also were higher on the district assessment than on the state assessment, which the authors said reflected the district's lower and more narrowly focused standards. This is an example of a standards-based curriculum that is driving pedagogical practice in a direction consistent with the nature of those particular standards, but which is inconsistent with the broader notion of standards expressing higher expectations for all students.

Overall student achievement improved in a district in the Wake County Public School System (WCPSS) in North Carolina when it implemented an instructional initiative called *Project Achieve* during the 2001–2002 school year. Speas (2003) conducted a mixed-method, formative evaluation on the implementation of the program in six elementary schools and two middle schools. Speas examined student state test scores in 2001 and 2002 in reading and mathematics and the WCPSS Effectiveness Index (an index created by using prior student achievement, special education status, and two measures of SES). All six elementary schools met the state's ABC High Growth standards for 2002; only three of the schools had met the ABC standards in 2001. Between 80 and 91 percent of the students were at or above grade level, compared with 73–83 percent in 2001. In general, the overall WCPSS Effectiveness Index scores were

higher in 2002 than in 2001. Middle school students also improved; both schools met the ABC standards in 2002, compared to only one in 2001.

A second evaluation study reported by Speas (2004) used a quantitative quasi-experimental design and included more schools. Three years of student state test scores in reading and mathematics for the *Project Achieve* schools were compared with non-*Project Achieve* schools in the same district. All of the schools met the state's ABC Standards, an increase from the previous two years. The *Project Achieve* schools, as a group, showed higher growth than non-*Project Achieve* schools in reading (grades 4–5) and in mathematics (grades 3-5). There was more growth in reading scores in the *Project Achieve* middle schools than the district average in grades sixth, seventh and eighth. Overall, the *Project Achieve* schools had a higher percentage of low-achieving students, special education students, and students who were eligible for free or reduced price lunch than the district average, so gains in these schools were interpreted as encouraging.

Studies of language arts curricula. Several studies looked at specific elements of language arts standards such as reading, literacy, and English. Caron's (2002) mixed-methods study used surveys and interviews of high school English teachers to investigate changes in instructional practices in response to Indiana state standards. Most teachers indicated a slight increase in their learner-centered practices and less use of lecture formats, textbook-based worksheets, and restricted response tests in their classrooms. Time spent on student reading did not increase, but different types of reading materials were introduced. At least a third of the teachers indicated that due to the standards, they were assigning more short essays, three-plus page reports, letter writing, and creative writing. Students gave more oral reports and participated in more interactive classroom discussions. However, a third of the teachers said they were unable to individualize instruction as much as they had before the implementation of standards. Almost half of the teachers responded that their reliance on the curriculum guides had increased and that this led to less creativity on their part. Caron also found that teachers with less than seven years of experience appeared more willing to change their instruction to be standards-based than those who had been teaching longer than seven years. Interview data indicated that there were teachers who reported little change in their pedagogy because they felt that they were already employing the practices that were called for by the standards.

In a longitudinal, qualitative case study conducted by Grisham and Brink (2003), teachers also stated that they felt that the standards were only reinforcing their current pedagogical practices and that they did not have to make many changes. Three elementary teachers were interviewed during the 1997–1998 school year and participated in follow-up interviews in 2000. Based on multiple classroom observations, the researchers indicated that teachers' actual instructional practices lagged behind their knowledge of standards-based reform, consistent with findings in science (Thompson, Zeuli & Borman,1997) and mathematics (Spillane & Zeuli, 1999). As with other studies in this section (Caron, 2000;

Sandholtz, et al., 2004; McLaughlin, 2000), one of the main concerns about standards was that they would stifle the creativity of teachers by forcing them to adhere to strict curriculum guidelines.

Jerome and Gilman (2003) used a quantitative quasi-experimental design to examine differences in student achievement after a writing improvement curriculum based on the Indiana state standards was implemented school-wide. In a sample of test scores from two third-grade classes across two school years, the researchers found that the students who did not participate in the writing improvement actually had higher scores on the Indiana Statewide Testing for Educational Progress writing test than the students who participated in the school-wide writing improvement program for two years. The researchers acknowledged that they did not control for between–group variables such as teacher influence or previous academic ability.

Researchers at the Center for Innovation in Assessment at Indiana University (1998) examined the impact of the *Signatures* program, a standards-based reading curriculum developed by Harcourt Brace. In a quantitative non-experimental evaluation study, they analyzed pre- and post-tests on the *Signatures* assessment and the reading section of the Stanford 9 Achievement Test for both first- and fourth-grade students in 18 classrooms across six states. Results showed that students in both grades performed significantly higher on the curriculum assessment and the Stanford 9 test after using the curriculum for one year.

Studies of general standards-based curricula. One qualitative case study examined the influence of standards-based curriculum in general on the instruction received by special education students, which is important given the equity emphasis that high standards should apply to all students. Over the course of four years, McLaughlin (2000) looked at four districts in four states that were in different stages of standards-based curriculum implementation. After conducting classroom observations and interviews with principals, teachers, and affiliated special education teachers in 28 elementary and middle schools, McLaughlin found varying results in teachers' instructional practices related to special education students across the four districts. For example, observations in the urban district suggested that lecture was the primary approach to instruction, and there were very few small-group interactions. In the rural district, most of the responsibility for adapting the standards-based materials and instruction for special education students fell on the special education teachers, even though the students were often integrated into the general classroom. Observations in the suburban district found that teachers used standards to guide instruction, including using small-group work.

Across the districts, there was a wide range of accommodation and modifications for special education students to help them master skills and concepts. The use of portfolios

worked especially well for the special education students because it allowed for multiple ways to demonstrate their achievement. In the affluent district a great effort was made to provide access to the curriculum and individualized instruction for the special education students. At the end of the fourth year of the study, the researcher perceived a change "from the presence and participation of students with disabilities in general education classrooms to an expectation that the students must learn what was being taught because they would be tested on that curriculum and their scores would matter to the school" (McLaughlin, 2000, p. 25). In general, teachers in all four districts were concerned with the amount of instructional time needed to incorporate standards in the classroom at all learning levels to ensure that all students met expectations. When some students did not fully understand the lesson, the teachers expressed their frustration with having to continue to the next lesson and keep up the instructional pace as dictated by the curriculum. Another commonality among teachers in all four districts was that they felt more comfortable implementing standards in their classrooms when they had been involved with writing and designing standards in their district.

Summary of findings from studies of standards-based curricula in other content areas. Overall, the 11 studies in this section found mostly positive influences of standards-based curricula on teacher instruction and student achievement. In the six studies that specifically examined student achievement, four reported increases in student performance on some type of achievement measure, and two reported negative findings. In one of the studies with negative findings, alignment with a standards-based curriculum meant alignment with district standards, which communicated lower expectations for student achievement than the state standards. Therefore, students had higher scores on district assessments than state assessments. In the other study that showed a negative relationship between use of a standards-based writing curriculum and student achievement, the researchers did not control for prior achievement between groups. This may have skewed the results.

Standards-based curricula affect teacher instruction — depending on the nature of the standards and curricula used. some reform-oriented influences on teacher instruction included an increased focus on student-centered instruction and improved alignment of standards, instruction, and curricular goals. Differences were found in instruction between locales (e.g., an urban locale used more lecture compared to a suburban locale). Another interesting finding was that veteran teachers were less willing to change their instruction to reflect standards than less-experienced teachers.

In three other studies, however, results were mixed. In these studies teachers used less student-centered instruction and focused more on basic skills and other traditional instructional strategies. However, in two of these studies, the standards referenced may have introduced a source of variance; Smith's (2003) study examined the role of Virginia state standards and Sandholtz, et al (2004) examined a district that explicitly referenced

standards that had low expectations for student learning. In the third study, teachers expressed the idea that they were already teaching in the ways that the standards implied, and that further change was not required (Grisham & Brink, 2003).

Studies with Secondary Focus on Standards-based Curricula

Several studies have been reviewed in other chapters of this synthesis but warrant discussion here because they addressed standards policies and standards-based curricula as a secondary focus. Nine studies from Chapter 4 on standards-based instructional guidance inform the synthesis findings. Several studies (Adams, 1999; Saxe, Gearhart, & Selzar, 2000; McCaffrey et al., 2001) validate the initial findings here that students using a standards-based curriculum can perform better on achievement tests than their counterparts not participating in such a curriculum. This is particularly the case on measures of problem solving. However, another study (Watson, 1996) implied that participation in an NCTM-oriented algebra course did not significantly improve student achievement and, in some cases, was associated with lower achievement than that of students in a traditional algebra course. There were some questions, however, about the equivalence of treatment and comparison groups in this study. Another study by Shymansky, Yore, and Anderson (2000) found no significant relationship between teachers' use of NSES-related strategies, as rated by the science supervisor, and student achievement in science. In both studies, a possible explanation of the negative results was the mismatch between the achievement measure and the curriculum being assessed.

Several studies confirmed the finding that teachers were knowledgeable in terms of the language they use in talking about standards but struggled with actually changing their classroom instruction from a traditional approach to one that is more reform-oriented (Spillane & Zeuli, 1999; Spillane & Jennings, 1997; Smith, 2000). Grant and Kline (2000) supported the finding that implementing a standards-based curriculum takes time and presents many challenges to teachers, including how to encourage student ideas and how to employ more student-centered instructional strategies.

Three additional studies from Chapter 5 on standards-based assessment also contribute to the findings reported here. A qualitative study by Clarke et al. (2003) supports the finding that state standards can influence classroom instruction. In this study, most educators reported an increased emphasis on problem solving and writing due to the influence of state standards. A small minority expressed concerns that the material and pace were developmentally inappropriate, the curriculum was being narrowed, and that their flexibility in teaching had decreased. In addition, Fernandez (2004) found that teachers perceived state standards as informative for lesson preparation. An important conclusion from another study is that what teachers teach is influenced by a number of factors; Janson (2002) found that, in Ohio teachers had not aligned their classroom assessments with the state assessments, partially because of their dependence on publisher–developed

tests, which were different from those of the state. This indicates that alignment of these different influences — curriculum, state expectations, classroom tests, and state proficiency tests — is crucial.

Summary of Findings of Studies of Standards-based Curricula

This review of the research about standards-based curricula and standards policies has examined their influences on student achievement and teacher instruction across mathematics, science, language arts, and social studies.

Generally, the majority of studies addressing student achievement found positive relationships between standards-based curricula and/or standards policies and improved student work and student achievement on tests. Use of standards-based curricula can improve student learning. Further, many studies found trends that suggest student achievement improves substantially when students are exposed to a standards-based curriculum for an extended period of time. While a few studies found a negative or no relationship between standards-based curricula and student achievement, this can be attributed to variability in implementation and lack of control variables in the research design (e.g., control for previous student achievement). Other possibilities are that the nature of the standards on which the curriculum is based may influence the relationship between participation and achievement, as well as a disparity between the achievement measure and the curriculum goals. Even so, a number of studies show that students are able to demonstrate the knowledge gained from exposure to a standards-based curriculum, even one emphasizing reform-oriented practices, on traditional standardized achievement measures. Another important finding across studies was that a standards-based curriculum can have positive influences on student achievement for a range of students (e.g., low-performing, low SES, minority).

In general, standards-based curricula and policies affect teacher instruction as well, although the results were less consistent than with achievement. Overall, teachers were changing their pedagogy to reflect more reform-oriented approaches, although this finding is attenuated due to the variation with which different standards reflect a reform or traditional orientation. Instructional activities reported by teachers included using a variety of strategies such as student-centered activities, small group work, addressing multiple answers to problems, and less lecture. In terms of content, a few teachers reported a perceived narrowing of the curriculum and lack of time to cover required content adequately. In addition, some teachers reported increased stress related to standards and indicated the need for more time to prepare to teach a standards-based curriculum. Many studies indicated that teachers are knowledgeable and aware of standards-based reform, but their actual pedagogy did not align with or lagged behind the tenets of standards-based instruction.

IMPLICATIONS FOR PRACTICE AND POLICY

Findings from the studies on curricula reviewed in this chapter demonstrate several critical issues that need to be addressed by educators when implementing a standards-based curriculum and/or standards policies. Research suggests that fidelity of a curriculum's implementation is an important factor in determining the effects of such a curriculum on achievement or teacher instruction. For successful implementation, teachers need to be knowledgeable about the content expressed in the curriculum's standards and they need to be able to apply that knowledge to their classroom instruction, both in terms of the content they cover and the pedagogical strategies they use. Professional development and instructional guidelines are important factors in preparing teachers in standards-based content and instructional strategies. Another important element is time. It has been reiterated in the research that teachers need ongoing support and that substantive changes in teacher instruction, at least those implied in reform-oriented standards documents, take time.

With regard to student achievement and learning, those studies of standards-based curricula and standards policies that link to reform-oriented standards show promising results. For the most part, students exposed to these curricula and learning environments are performing at higher levels than students in more traditional education environments. In the context of standards-based reform, low-achieving students show some improvement, but need extra help and support in order to be active learners. For standards to really affect all students, teachers and administrators will need to learn how to make necessary accommodations and how to individualize instruction in ways that engage all students in standards-based learning environments.

Administrators and policy makers play a crucial role in the success of standards-based education. The research suggests that standards (defined as agreements about the content that students should know and the levels at which they should know it) are most effective when translated into curricula and curriculum frameworks that are aligned with instructional materials and student assessments. All of these factors need to be addressed for system success and appropriate resources need to be supplied to support the necessary changes. In addition, equity issues need to be addressed by policymakers. The research reviewed here suggests that pedagogy in poor districts is more traditional and textbook-oriented compared to wealthier districts, which emphasize more reform-oriented instruction. Finally, it is advantageous to include teachers when designing and writing standards and curricula. This type of involvement creates vested interests and increased knowledge for teachers so they can more effectively incorporate standards in their instruction.

CONCLUSIONS

Overall, there is evidence that standards-based curricula and standards policies contribute to the ultimate goal of standards-based education, which is to hold all students to high standards. Standards-based curricula and standards policies provide opportunities to improve teacher instruction and student achievement, as long as the standards themselves reflect ambitious expectations for teaching and learning.

Although the overall findings of these studies are promising, further research is needed in several areas. More research on the influence of standards-based curricula on special populations, such as special education and English language learners, is needed. Also needed are further detailed studies of curriculum implementation. Variations in how teachers implement a standards-based curriculum can have serious implications for what content students learn and how well they learn it. This plays into concern about equity issues, which warrant more extensive study in order to see how access to ambitious curricula is distributed for different groups of students in a standards-based system. Finally, longitudinal studies would contribute to the knowledge base by clarifying the extent to which student knowledge acquired from a particular curriculum is sustained over time. Longitudinal studies also could provide insights about how to decrease stresses on teachers who are trying to change their instruction to align with standards.

CHAPTER 4:
THE INFLUENCE OF STANDARDS-BASED INSTRUCTIONAL GUIDELINES ON TEACHING AND STUDENT LEARNING

As noted in Chapter 1, standards-based education can take the form of either a policy approach based on the specification of content that students should learn or an approach to learning content that draws on the early reform work of mathematics and science groups. The latter include the NCTM and the NRC which emphasize the acquisition of higher-order learning, which is a distinction beyond that of content. The first definition can have a broad array of implications for teacher pedagogy, ranging from teacher lecture to small group work. The second definition implies a clearer push toward what is often called "reform-oriented" instructional practice — or more student-centered, active learning practices. Most of the studies addressed in this chapter use the second definition, although the particular standards referenced are described study by study.

Advocates of standards-based systemic reform in the early 1990s aimed to make a number of adjustments to the entire education system. They centered that system around a series of ambitious new content standards that described what students — all students — should know and be able to do. Then, around that centerpiece, related assessments, ambitious curricula, and revised teacher education programs were to be aligned to support students to learn those standards (Smith & O'Day, 1990). Rather than exposing students to the low-level drilling of procedures and rote learning encouraged in the minimum-competency testing reforms, the standards movement shifted the instructional focus to a "thinking curriculum," emphasizing higher-order, more complex learning skills, and strategies by which to strengthen those skills (Resnick & Resnick, 1992).

Such attention to changes in instruction was highly ambitious, as instruction, one of the central aspects of schooling, has traditionally been one of the most difficult aspects to reform (Conley, 1993). Historically, schools have engaged in a number of surface changes (such as site-based management, collaborative decision making, block scheduling) that do not substantively alter the nature of how teachers and students interact together around learning (Cuban, 1989). However, a number of sweeping documents from professional associations, such as the NCTM in mathematics and the

NRC in science, began to provide large-scale guidelines for changing instruction in ways that encourage inquiry learning, real-world problem solving, and student development as more active learners. Further, groups like the NSF began to fund implementation projects around these types of reforms, broadening dissemination nationally.

This chapter focuses on the recent research about the ways in which teacher instruction, primarily pedagogical practice, has changed relative to the recommendations of standards-based reforms and the effects of such reform-oriented instruction on student achievement. Due to the influence of the NCTM curriculum and professional standards, most of this research addresses mathematics instruction, although some studies also examine reading, writing, and science instruction relative to standards.

For the purposes of this synthesis, standards-based instructional guidelines are defined as any broad guidelines for standards-based teaching in the content areas, such as the *Professional Standards for Teaching Mathematic*s, which "present a vision of what teaching should entail to support changes in the curriculum set out in the *Curriculum and Evaluation Standards"* (NCTM, 1991, p. vii). Similar guidelines include the NSES, which, in addition to laying out the content of what students know and should be able to do in science, highlight an inquiry approach based on authentic questions about science.

OVERVIEW OF STUDIES

A total of 36 studies met the inclusion criteria and addressed the influence of standards-based instructional guidelines on student achievement and/or teacher instruction. Detailed descriptions of each study, including study methods, sample, and measures, as well as a summary of relevant findings, can be found in
Appendix D.

The vast majority of studies (n=33) focused on national standards, with only three studies focusing at all on district/local standards and four examining state standards. Of the total, seven studies addressed standards at multiple policy levels. Mathematics was the content area most frequently examined by the studies (n =31), and science was a distant second (n=6). Three studies examined instruction in language arts, and eight studies examined more than one subject, most frequently mathematics and science or mathematics and reading. Additionally, studies were fairly even in terms of whether they focused on elementary students (n=22) or middle school students (n=19). Eleven studies examined standards-based instructional guidelines at the high school level, and 12 studies addressed multiple levels of schooling.

A total of 20 studies focused on the relation between instructional guidelines and achievement, with 11 of those studies using standardized tests, eight studies using a

researcher-developed assessment, three studies using a state test, and one study using grades as an achievement measure. Sixteen studies examined the relation between instructional guidelines and teacher instruction, with nine studies using teacher surveys to measure instruction, eight using teacher interviews, six using classroom observations, four using instructional document review, and one study each using principal surveys, student surveys, and student interviews. Multiple measures of instruction were used for a number of the studies that concerned teacher instruction.

The sources of studies included 17 journal articles, 10 conference presentations, 6 technical reports, and 3 dissertations. Two studies had experimental designs, with subjects randomly assigned to treatment groups, and three used quasi-experimental designs, in which comparison groups were used. Most studies (n=20) were quantitative and non-experimental in design, and several of these were secondary analyses of large-scale education data sets. Seven qualitative studies were included, and four studies used mixed methods.

Eleven studies considered at-risk students, either including sub-analyses for at-risk students or drawing from at-risk samples. Of the studies that focused on at-risk students, five defined at risk in terms of low socio-economic status (SES), three in terms of ethnicity (minority status), three as English-language learners (ELL), and one looked at students with disabilities. One study examined both ELL and ethnicity.

Systemic influences were examined in only four studies that passed synthesis selection criteria. They noted the importance of teacher learning, accountability policies, and school/community relationships in instructional guidelines.

STUDY FINDINGS

McREL researchers organized the studies synthesized in this chapter primarily by the outcome variable (whether studies examined the relation between instructional guidelines and student achievement or the relation between instructional guidelines and pedagogy). Within each of these broad categories, researchers grouped studies by research method. More attention is given to studies in each area that had research designs with strong controls for threats to the validity of their findings, studies that allowed for some analysis of association between variables, studies with careful descriptions of implementation, detailed quantitative and qualitative analyses, and those studies that had noteworthy features.

Many of the studies that examined student achievement were not designed to attribute differences in achievement to a specific instructional guideline or treatment. While several experimental and quasi-experimental studies measured particular treatments

related to standards and could therefore attribute achievement differences to differences in the treatment, many studies lacked such comparison groups. These studies are described further below and focus on the direct link between reform-oriented pedagogy (consistent with recommendations from particular standards-based instructional guidelines) and achievement.

Studies of Influences on Student Achievement

The 20 studies that addressed the influence of standards-based instructional guidelines on student achievement are listed in Table 4.1. Sources of guidelines varied in this group, with 10 studies focusing on the effects of using or training teachers in the use of NCTM-based mathematics curricula or units (including *Seeing Fractions*, the *Integrated Mathematics Project, QUASAR – or Quantitative Understanding: Amplifying Student Achievement and Reasoning, Project JASPER,* or the *Southern California Regional Algebra Project, or SCRAP).* One study focused on the effects on student achievement of a standards-based science curriculum (*PALS)* aligned with the NSES. Two of the studies were quantitative and used true experimental designs and three were quantitative, using quasi-experimental designs. One study used mixed methods.

The remaining 14 studies were quantitative, non-experimental studies, including two large-scale studies of the Mosaic Project, using instruction and achievement data from a number of NSF-funded Systemic Initiative/Local Systemic Change (SI/LSC) sites. However, it is problematic to attribute achievement trends described in these studies to any sort of instructional guiding influence other than the broad influence of the NCTM *Professional Standards* (1991). Six of these non-experimental studies, all focusing on mathematics, focused more on actual teacher instruction relative to reform than they did the guidelines for teachers to teach in those ways. Therefore, these studies provide some valuable information about the relationship between specific teaching practices, consistent with the NCTM reforms, and student achievement.

Table 4.1. Studies of Standards-based Instructional Guidelines and Achievement

Author (s) & Year	Method	Standards Addressed	Subject Area(s)	Measures
Adams (1999)	Quantitative quasi-experimental	National: NCTM District	Math	Standardized tests
Brenner, Mayer, Moseley, Brar, Duran, Reed & Webb (1997)	Quantitative quasi-experimental	National: NCTM	Math	Researcher-developed assessment
Butty (2001)	Quantitative non-experimental	National: NCTM	Math	Standardized tests
Erickson & Niess (1996)	Quantitative non-experimental	National: NCTM	Math	Researcher-developed assessment

Ginsburg-Block & Fantuzzo (1998)	Quantitative experimental	National: NCTM	Math	Researcher-developed assessment
Hamilton, McCaffrey, Stecher, Klein, Robyn & Bugliari (2003)	Quantitative non-experimental	National: NCTM, NSES	Math Science	Standardized tests, state assessments, researcher-developed assessments
Hickey, Moore, & Pellegrino (2001)	Quantitative quasi-experimental	National: NCTM	Math	Standardized tests
Klein, Hamilton, McCaffrey, Stecher, Robyn, & Burroughs (2000)	Quantitative non-experimental	National: NCTM, NSES	Math Science	Standardized tests, state assessments, researcher-developed assessments
Lane, Silver & Wang (1995)	Quantitative non-experimental	National: NCTM	Math	Researcher-developed assessment,
Lubienski (2002)	Quantitative non-experimental	National: NCTM	Math	Standardized tests
Mayer (1997)	Quantitative non-experimental	National: NCTM	Math	Standardized tests
Mayer (1998)	Quantitative non-experimental	National: NCTM District	Math	Researcher-developed assessment
McCaffrey, Hamilton, Stecher, Klein, Bugliari, & Robyn (2001)	Quantitative non-experimental	National: NCTM	Math	Standardized tests
Rizor (2000)	Quantitative non-experimental	National: NCTM State: WY	Math	State assessment
Saxe, Gearhart & Selzer, (1999)	Quantitative non	National: NCTM	Math	Researcher-developed assessment
Shymansky, Yore & Anderson (2000)	Mixed methods (Quantitative non-experimental)	National: NSES	Science	Researcher-developed assessment
Silver & Lane (1995)	Quantitative non-experimental	National: NCTM	Math	Researcher-developed assessment
Snow-Renner (2000)	Quantitative non-experimental	National: NCTM	Math	Standardized tests
U.S. Department of Education (2001a, 2001b)	Quantitative non-experimental	National: NCTM	LA Math	Standardized tests
Watson (1996)	Quantitative experimental	National: NCTM	Math	Grades Standardized tests

NCTM = National Association of Teachers of Mathematics
NSES = National Science Education Standards

Quantitative experimental studies. The two quantitative experimental studies reviewed yielded mixed results about the impacts of standards-based instructional guidelines on student achievement. This may be due to different study designs, different outcome measures, or different types of treatments. Watson (1996) found that initial participation in an NCTM-influenced alternative algebra class (the *Southern California Regional Algebra Project*, or *SCRAP)* was either neutral or negatively associated with student grades in later, traditional high school mathematics courses. However, although students were described as randomly assigned to comparison and treatment groups in this study, there appeared to be considerable differences between the groups' prior mathematics achievement and ethnicity.

At the elementary level, Ginsburg-Block and Fantuzzo (1998) found that specific mathematics instruction in peer collaboration and problem solving positively affected mathematics achievement for poor and historically underachieving students. This study was notable for its strong design and protection against threats to study validity (such as monitoring any possible diversion from treatment conditions). The researchers randomly assigned matched student pairs to one of three peer-oriented experimental conditions (problem-solving, peer collaboration, or a combination of the two) or a control condition involving individual student instruction. The treatments took place simultaneously over a period of seven weeks. Before and at the end of the program, students were tested on problem-solving and computation measures aligned with the curriculum. Analyses using pretest achievement to control for possible group differences indicated that students in the problem-solving and peer collaboration groups showed significantly higher rates of achievement on both computation and problem-solving measures than did their counterparts in the traditional group. No significant effect was found for the combined condition.

Quantitative quasi-experimental studies. Three quantitative, quasi-experimental studies painted a more promising picture of the impact of standards-based instructional guidelines on student achievement. However, two of them indicated differences in findings related to the nature of the achievement measure.

Adams (1999) found that students who used a district-developed curriculum based on NCTM's *Professional Standards* (1991) had significantly higher mathematics scores on the Iowa test of Basic Skills (ITBS) than students who used the regular curriculum, even after controlling for differences in prior achievement. It should be noted that this study primarily focused on pedagogy consistent with NCTM's instructional guidelines and is addressed in Chapter 3 as a study with a secondary focus on curriculum.

In another study using ITBS scores, Hickey, Moore and Pellegrino (2001) found that fifth grade students of teachers participating in an NCTM-aligned technology program

(*JASPER*) for mathematics instruction had significantly higher increases in mathematics achievement than their nonparticipating peers. However, when ITBS mathematics subscales of problem-solving, concepts/estimation, and computation were used as achievement measures, different achievement patterns emerged. *JASPER* students' achievement increased on the other two subscales, but not on computation. Other interactions were apparent too; the SES of the classroom and the teacher's orientation to NCTM reforms seemed to mediate the effect of the treatment. For instance, high-SES classrooms experienced positive achievement results on the concepts/estimation measure, but not low-SES classrooms (although there was no negative relationship). In reform-oriented classrooms, problem-solving and concepts/estimation achievement increased more than in traditional classrooms, and low-SES students' concepts/estimation achievement decreased significantly less with more reform-oriented teachers. Importantly, computation scores declined significantly in treatment classrooms that were consistent with NCTM reforms and significantly increased in less reform-consistent classrooms.

The findings from the other quasi-experimental study in this section further illustrate the potential tension between higher-order problem solving skills and more basic skills. Brenner et al. (1997) studied the effect of an NCTM-based instructional unit on junior high school regular and ELL student achievement. The unit was a 20-day program focused on multiple representations of concepts and problem solving, and the achievement measure was a researcher–developed assessment that focused on four different mathematics domains: function word problems, word problem representations, word problem-solving, and equation-solving. In this study, both ELL and regular students who participated did significantly better than the comparison group on representing and solving function word problems. They also performed significantly better at problem representation tasks, such as translating word problems into tables and graphs. However, on the equation-solving section of the test, the treatment group did significantly worse than the comparison group, indicating the differences in learning outcomes produced by the treatment.

These findings indicate that the relationship between standards-based mathematics instruction and achievement depends substantially on the achievement measure used. Tests that emphasize higher-order skills are more likely to show a positive relationship with these practices, while tests focusing on more basic skills like computation or solving equations do not.

Quantitative non-experimental and mixed methods studies. The remaining studies of instruction and student achievement show a small positive relationship between standards-based instruction and achievement. There are, again, some inconsistencies across studies. As addressed in the previous section, one apparent reason for inconsistency is the variation across achievement measures. Also, as noted previously, a

number of these studies focus on the relationship between pedagogy consistent with standards and achievement. They are not designed to attribute the pedagogy to specific standards-based instructional guidelines like NCTM (1991), but examine practices congruent with those guidelines' recommendations.

Several non-experimental studies reinforce the theory that, while NCTM-oriented pedagogy may positively impact achievement on higher-order tests, it is not so well-suited to teaching computation. Erickson and Niess (1996), in their study of middle school math teachers' NCTM-oriented instruction, found that problem-solving and computation as mathematical domains, may each require separate instructional attention. They found that the proportion of time teachers spent on problem solving had a significant negative correlation with student computation scores. And the opposite was also true; teachers who spent more weeks per year on numbers and computation saw a significant negative correlation with student problem-solving scores. Part of the challenge for teachers is to use instructional strategies such as small group and manipulative work — which significantly and positively affect student achievement and problem-solving — but also to adequately address the instruction needed for computation. A study by Saxe, Gearhart, and Selzer (1999) had similar findings; their Hierarchical Linear Modeling (HLM) analysis of instruction and achievement in upper elementary mathematics classrooms also found that alignment of pedagogy with NCTM reform principles was positively related to student achievement in problem solving, but this was not the case for computation.

To further explore this potential instruction/assessment mismatch, Mayer conducted several rigorous statistical studies about the relationship between NCTM-oriented pedagogy and achievement on traditional standardized tests. In 1997, Mayer conducted a secondary analysis of data from the Longitudinal Survey of American Youth (LSAY) to examine the effects of active instructional practices on mathematics achievement on the 1986 National Assessment of NAEP, at that time, a typical standardized test. Although the LSAY data predated the publication of the NCTM standards, Mayer made the case that these active instructional practices were parallel with the intent of NCTM reformers. Mayer's HLM identified a negative relationship between active instruction and achievement as measured by the NAEP. The more emphasis reported by teachers on class discussion and small group work, (as opposed to lecture and seatwork), the less students gained on NAEP math achievement ($p<.10$). Based on these findings, Mayer argued that more sensitive measures are needed, and without them, teachers lack incentives to truly change their instruction in the ways envisioned by NCTM.

However, a second study (Mayer, 1998), using a slightly different sample and different measures, demonstrated different findings. This study examined the relationship of instruction consistent with NCTM to achievement, using three criterion-referenced traditional algebra tests. With a sample of middle school (grade 8) and high school (grade

9) students, the study used HLM and found a positive relationship between NCTM-oriented practices and student achievement, but only for the middle school students. Middle school teachers also used much more NCTM-oriented instructional practices than their high school counterparts indicating that achievement and reform practices were associated. Although this study raises interesting ancillary questions about access (i.e., 8th grade algebra students are on a faster track than 9th grade algebra students) to NCTM-oriented instruction, the alignment finding is relatively more positive than Mayer's 1997 study. Traditional multiple-choice tests were not held to be a mismatch with new standards-based instruction, at least not to the extent that they should prove a disincentive for teachers to use such instruction.

Other studies indicate a weak positive relationship between pedagogy consistent with standards and achievement, although this varies in consistency. The Longitudinal Evaluation of School Change and Performance (LESCP) of Title I schools (U.S. Department of Education, 2001a, 2001b), using an HLM to examine a variety of factors as predictors of achievement, found small gains in mathematics and reading achievement on the Stanford 9 Achievement Test when teachers used more exploration in mathematics and less basic skills instruction in reading. Silver and Lane (1995) found that students participating in QUASAR schools (an NCTM-oriented program) performed better than or as well as a national sample of their peers on a NAEP mathematics measure, although the extent to which this is attributable to their instruction is questionable. Other factors besides the nature of the test seem to mediate the relationship between standards-based instruction and achievement, including student grade level (Snow-Renner, 2000) and other organizational factors, such as school scheduling and course structure (McCaffrey et al, 2001).

Several other studies found no relationship between standards-based instruction and achievement. Rizor (2000), in a snapshot study of Wyoming teachers' mathematics pedagogy consistent with NCTM reforms and mean, school-level test scores, found no correlation. However, a number of studies have indicated that within-school variation must be examined to capture the complexities of the instruction–achievement relationship and it is unclear whether achievement measures were based only on mathematics, or also included literacy achievement data. Similarly, Shymansky, Yore, and Anderson's (2000) mixed-methods study found no relationship between teachers' use of science instruction related to the NSES and student achievement, although the study's assessment of instruction (a supervisor's rating) was not well-documented in terms of concrete behavioral indicators.

Several studies focused on standards-based instruction for at-risk students. Findings indicate that the relationship between instruction and achievement is positive for these students, to the extent that they are exposed to such instruction. Lane, Silver, and Wang (1995), in their study of middle school achievement gains in urban QUASAR schools

(organized around NCTM-oriented mathematics instruction), found that, over several years, black students and ELL students experienced achievement gains that paralleled those of white students as long as the at-risk students experienced access to challenging instruction. Lubienski (2002) conducted a secondary analysis of national NAEP data to examine changes in mathematics instruction over time relative to calculator and computer use and student achievement trends by student ethnicity. While all students' achievement improved between 1990 and 2000, black and Hispanic students still lagged behind and, instructionally, they experienced more traditional forms of instruction and assessment. Butty's (2001) analysis of the National Education Longitudinal Study of 1988 data for black and Hispanic students at grades 10 and 12 had mixed findings. Exposure to reform-oriented mathematics practices resulted in significantly higher achievement than traditional instruction for 12th grade students, but not for 10th grade students.

Some noteworthy non-experimental studies of standards-based instruction and student achievement were two large-scale analyses conducted by the RAND Corporation. The studies drew on a large body of data from multiple sites (states, districts, or cities) that had participated in NSF's SI/LSC initiatives, which emphasized the NCTM standards. Both studies had similar findings, but the second used an expanded sample. Klein et al.'s (2000) study measured teacher mathematics practices using a validated teacher survey. Achievement was operationalized with a variety of existing measures, supplemented in some instances with a RAND-developed science measure. Data from approximately 100 elementary and middle schools scattered across six NSF sites were used in the initial study. The researchers found a weak positive relationship between standards-based instruction and achievement for both mathematics and science, with some slight indication that open-response tests were more sensitive to such instruction.

Hamilton et al. (2003) followed up the earlier RAND study using the same measures and analyses, but with a sample of 11 sites and achievement data from over 13,000 students in mathematics and 14,000 students in science. For both studies, the researchers used site-specific covariates to control for student prior achievement, SES and other factors and then conducted linear regression analyses for instruction and achievement. Using the estimated regression coefficients for each site, they conducted pooled analyses to examine the effects of reform-oriented instruction on achievement across sites for mathematics and science and for test format (open-response or multiple-choice). The studies found considerable in-school variation in the extent to which teachers used reform or traditional practices. They also found that use of reform-oriented instructional strategies had weak positive associations with student achievement across mathematics and science, rarely statistically significant and much smaller than the effects of student background characteristics (e.g., SES or ethnicity) on achievement.

Summary of influences on student achievement. The research reviewed here indicated small positive influences of standards-based instructional guidelines (or, in cases where attribution to particular guidelines was not addressed, pedagogy consistent

with the recommendations of standards reformers) on student mathematics achievement, provided that the achievement measure focused on higher-order learning and not basic skills such as computation. Eleven of 20 studies reviewed indicated a positive relationship between standards-based instructional practices and student achievement, although four of these studies noted small effect sizes and three reported on positive but inconsistent effects. Further, four additional studies indicated a positive influence of standards-based instruction on higher-order assessments, but negative influences on measures that emphasized computation or solving equations. In addition, findings were inconsistent across studies, with three studies indicating no relationship and two showing a negative overall relationship between standards-based (e.g., reform-oriented) instruction and student achievement. The nature of the achievement measure is important in examining this relationship, but other factors also seem to play mediating roles. Several studies indicated that student readiness or aspects of differential access may come into play (Mayer, 1998, Lubienski, 2002, Snow-Renner, 2000) as well as other organizational differences such as the ways in which courses are structured (Saxe, Gearhart, & Selzer, 1999).

Studies of Influences on Pedagogy

Sixteen studies met the selection criteria and addressed the relationship between standards-based instructional guidelines and standards-based pedagogy and are listed in Table 4.2. Of those studies, six were quantitative non-experimental in design, three used mixed methods, and seven were qualitative designs.

Table 4.2. Studies of Standards-based Instructional Guidelines and Pedagogy

Author (s) & Year	Method	Standards Addressed	Subject Area(s)	Measures
Barth, Haycock, Jackson, Mora, Ruiz, Robinson, & Wilkins (1999)	Quantitative non-experimental	Multiple levels (general)	General	Principal surveys
Blank, Kim, & Smithson (2000)	Quantitative non-experimental	National: NCTM, NSES	Math Science	Teacher surveys
Bridge, Compton-Hall, & Cantrell (1997)	Mixed methods (Quantitative non-experimental)	State: KY	LA	Teacher surveys Classroom observations
Burian-Fitzgerald, McGrath & Plisko (2003)	Quantitative non-experimental	Multiple levels: National State	Math	Teacher surveys
Flexer, Cumbo, Borko, Mayfield, & Marion (1995)	Qualitative	National: NCTM	Math	Teacher interviews Document review from professional development workshops
Grant & Kline	Qualitative	National: NCTM	Math	Classroom observations

(2000)		District		Teacher interviews
Ivey (1996)	Qualitative	National: NCTM	Math	Classroom observations Document review
Mayrowetz (1999)	Qualitative	National: NCTM	Math	Classroom observations Teacher interviews
McGinnis & Parker (2000)	Quantitative non-experimental	National: NCTM	Math	Teacher surveys
Norman, Stein, Moussiaux, & Clay-Chambers (1998)	Quantitative non-experimental	National: NCTM, NSES	Math, Science	Teacher and student surveys Classroom observations
Smith (2000)	Qualitative	National: NCTM	Math	Classroom observations Teacher interviews Document review
Spillane & Jennings (1997)	Qualitative	District	LA	Classroom observations Teacher interviews Document review
Spillane & Zeuli (1999)	Qualitative	National: NCTM	Math	Teacher surveys Teacher interviews Classroom observations
Swanson & Stevenson (2002)	Quantitative non-experimental	National: NCTM	Math	Teacher surveys
Swierzbin, Liu & Thurlow (2000)	Mixed methods (Quantitative non-experimental)	State: MN	General	Teacher surveys Teacher interviews
Turner (1998)	Mixed methods	National: NCTM	Math	Teacher surveys Teacher interviews Classroom observations

NCTM = National Association of Teachers of Mathematics
NSES = National Science Education Standards

Quantitative non-experimental and mixed methods studies. McREL researchers reviewed six quantitative, non-experimental studies related to pedagogy. Burian-Fitzgerald, McGrath and Plisko (2003), in a comparison of 1995 and 1999 mathematics teacher data from the Third International Mathematics and Science Study (TIMSS), found that teachers' awareness of state assessment specifications and curriculum guides increased considerably from 1995 to 1999. Further, teachers who reported more familiarity with standards were more likely to ask their students to do problem-solving activities. Similarly, Barth et al.'s (1999) widely-disseminated study of high-performing, high-needs schools stated that the use of ditto sheets had increased over the past few years and student discussion of their work had increased. It is not clear,

however, the extent to which these statements reflect the influence of standards-based instructional guidance or of actual teacher practice since the respondent group was composed of principals who were responding to questions about their school as a whole. But it does appear that instruction is changing somewhat to align with the standards, at least as standards have been most broadly interpreted to mean guidelines for content and assessment coverage.

Blank, Kim, and Smithson's (2000) study of mathematics and science instruction in four NSF-funded USIs also indicated that teachers are modifying their pedagogy in ways consistent with the intent of standards-based reforms. This study used a measure much more grounded in actual classroom work than the Barth et al. (1999) study. The researchers' *Survey of Enacted Curriculum* identified systematic differences in instruction between low-implementation and high-implementation sites. Students at high-implementation sites participated in more math analysis, performance assessment, and did more explanations of their reasoning and justifying their answers in mathematics and science than their counterparts in low-implementation sites. They also spent less time on computation and basic memorization skills than their counterparts at low-implementation sites.

Swanson and Stevenson's (2002) multi-level analysis of state, school, and classroom level influences on pedagogy bears further mention as a rigorous study with important implications for policymakers. This study developed a theoretically-grounded hierarchical model designed to explore the roles of different predictors of instructional change, starting with state standards activism. Combining 22 indicators of state-level standards policy activity and coherence into a single measure to classify state policy, the researchers then incorporated state, school, and classroom variables into an HLM analysis using teacher survey data gathered through NAEP. The dependent variable was teachers' use of NCTM-oriented reform pedagogy. Initially researchers examined the partition of variance in instructional practice and found that over three-quarters of this variance was explained at the classroom level, almost 22 percent at the school level, and less than 3 percent by state policy. However, state policy was found to have a modest but significant positive effect on teacher use of standards-based instructional pedagogy in mathematics. This was strongest when mediated through teacher knowledge and attitudes about the reforms. Teacher knowledge about or positive disposition towards NCTM predicted significantly higher levels of standards-based pedagogy, indicating that the ways in which state policy are received depends largely on the extent to which teachers have access to higher levels of standards-relevant professional learning experiences. Recommendations for policymakers were to generate teacher awareness and how-to knowledge around standards-based pedagogy.

Several of the quantitative studies reviewed hinted that instructional change may take some time and require considerable organizational and policy support. For example,

Norman, Stein, Moussiaux, and Clay-Chambers' (1998) evaluation of the Detroit USI indicated changes in teacher practice, but primarily for teachers who had participated for three years. McGinnis and Parker (2000) found that a supportive school environment is key in their survey and follow-up study of mathematics and science teachers who had just completed an NCTM/NSES preparation program. Although the students' survey responses about proposed practices were considerably more in agreement with the standards reforms than a national NSF sample of working teachers, once they began teaching in schools, their work environments mediated and, to a large extent, limited the ways that they put the reforms into practice.

McREL researchers reviewed three mixed methods studies and found that they also support the quantitative findings that general pedagogical change has taken place, although these studies raise questions about how evenly and equitably the reforms are spreading. Bridge, Compton-Hall and Cantrell (1997) in a writing replication study, found considerable changes in instruction in one Kentucky school from 1982 to 1995. Teachers dedicated much more total time to writing than previously and focused more on more high-level activities, portfolio assessments, and open-ended and on-demand writing assignments. They noted that their changes were a direct result of the Kentucky Educational Reform Act. However, other studies highlight unequal access to ambitious instruction (Swierzbin, Liu and Thurlow, 2000) and lack of substantive reform implementation (Turner, 1998).

Qualitative studies. Seven qualitative studies met the synthesis inclusion criteria. They provided detailed information about how instructional changes take place, the required shifts in teachers' beliefs and practices, and the extent to which more subtle changes in instruction may occur (beyond the broad-brush instructional practices measured through teacher surveys).

Generally teachers tend to change their instruction incrementally, here and there, and only where it may seem most appropriate based on their own beliefs and experience. One example of this can be found in Ivey's (1996) description of teacher-centered and student centered cultures co-existing in the same beginning algebra class, depending on a particular day's assignments and tasks. Similarly, Grant and Kline (2000) tracked three teachers over the course of a year as they learned how to implement a standards-based K–2 mathematics curriculum. The teachers' instruction all progressed at different rates, along a continuum of orientation to standards. Their progress varied depending on how well their beliefs aligned with the standards reforms initially and their willingness to challenge those beliefs. All three teachers were challenged by particular tasks like eliciting and engaging with student ideas (without giving them the answers) and learning how to use incorrect student answers as teachable moments. Handling incorrect answers in a way that allowed student development of knowledge (rather than telling students

answers or procedures) and modifying instruction in an ambitious classroom environment also were a problem for teachers of learning-disabled students (Mayrowetz,1999).

Smith's (2000) in-depth study of one middle school mathematics teacher's growth provides valuable insights about how teachers might handle challenges to their old ways of thinking. Based on observations, teacher journal entries, and interviews over the course of a year's professional development on an NCTM program, the study identified changes that went far beyond simple behavioral modifications. The teacher's experience in the training program and with her classes led her to experiences that presented her with new ideas about what students needed to know and be able to do in math. These new ideas demanded reconciliation with the teacher's long-held beliefs. Confronting these dilemmas was difficult, but it allowed the teacher to learn and to change her practice so that it was substantively congruent with NCTM recommendations. However, the experience took a considerable amount of time — at least a year of sustained professional development and reflection.

Flexer, Cumbo, Borko, Mayfield, and Marion's (1995) qualitative study of elementary mathematics teachers' responses to performance assessment training and classroom use reinforced Smith's (2000) findings. Teachers participated in the study over the course of a year. They had considerable resources including access to five assessment experts on a weekly basis. They made many adjustments in their instruction and their assessments that were mutually reinforcing and led to greater alignment with NCTM's vision. Further their descriptions of what they believed about how children learn and how to teach mathematics changed over the course of the year. However, it was a difficult process for many teachers, fraught with anxiety and difficulties in breaking instructional habits, although one that seemed to herald substantive instructional change consistent with NCTM's recommendations. Substantive instructional change was difficult for teachers, taking time and the freedom to question one's assumptions, and being allowed to practice to get better at these types of high-level instructional strategies.

Two related studies by Spillane and Jennings (1997) and Spillane and Zeuli (1999) focused on the actual nature of standards-based classroom instruction. What does it mean in terms of specific indicators when we say that it's changed? And how do we know? Both studies have been meticulously researched, are situated in a strong conceptual base, and emphasize the importance of real changes in instruction from basic skills. Spillane and Jennings (1997) interviewed and observed language arts teachers in a school district that had aligned its policies and curriculum around ambitious language arts standards. In terms of systemic reform, the district seemed an ideal scene. Further, in teacher interviews, instructional similarities seemed apparent; teachers reported that they were emphasizing literature and real books more than in the past, when there was more focus on basals and lower-level skills instruction. Teachers also spoke knowledgeably of reading and standards in relation to their instruction. However, observations indicated

considerable variation in instructional practice, both in the extent to which tasks were challenging of higher-order skills and the extent to which children were required to justify their answers, rather than just to express an opinion on the text. Some teachers emphasized more substantive student comments than others who organized their rooms and talked about them as if they were aligned with ambitious standards but who basically practiced teacher-centered instruction. District policy and resource alignment were insufficient to help teachers change their roles in the ways implied by new ambitious standards. The authors suggested the use of state and district policy as an educative tool. They recommended that policy be designed especially to help teachers learn how to enact the reforms in their classrooms using models and examples.

The Spillane and Zeuli (1999) study extended this inquiry into the area of mathematics and compared observation data with teachers' self-reports using the TIMSS teacher survey. Based on their TIMSS responses about standards and use of reform-oriented mathematics practices, a sub-sample of 25 elementary and middle school mathematics teachers was selected for in-depth observations and interviews. The observations and interviews, once again, focused on the nature the cognitive challenge of classroom tasks required of students and the level of classroom discourse. Although all of the teachers described themselves as very knowledgeable about standards and their teaching as well-aligned with the NCTM standards, in actuality, many of their classroom practices fell well short of the mark. Of the 25 teachers indicating coherence with the standards reforms on the survey measure, only four taught in a way consistent with NCTM instructional recommendations. These teachers taught using conceptually-oriented tasks and conceptually-centered classroom discourse. They focused on students' explanation of their reasoning and justifying their answers, rather than on correct procedures for getting the right answer. The remainder of the teachers had made incremental adjustments to their teaching consistent with the impetus of NCTM; however, they were not as standards-based as might have been thought based on their survey responses. Many teachers focused on conceptually-oriented tasks, but directed student discourse to procedures for getting the correct answer. Still others displayed peripheral changes in instruction, such as using manipulatives as recommended by NCTM, but using them in such a way as to strengthen the understanding of a procedure rather than to build principled conceptual knowledge of the underlying mathematics. Both studies highlight the limitations of survey measures to identify instructional practices. While surveys are useful and relatively inexpensive tools for identifying broad-brush trends of practice, they may not be very effective at measuring the subtle, but important changes implied for instruction in the standards movement.

Summary of influences on teacher pedagogy. The studies exploring the relationship between standards-based instructional guidelines and teacher pedagogy indicate that broad changes are happening in instruction, consistent with the overall intent of the standards. Three studies indicated that this is the case. However, three other studies

noted that not all students have access to such changes and in some sites, they are not happening at all. As highlighted in two studies, the extent to which the more ambitious tenets of standards-based instruction are taking hold in American classrooms is not clear from survey results. Further, nine studies indicated that substantive change to standards-based instruction takes considerable time and teacher learning.

Studies with Secondary Focus on Standards-based Instructional Guidelines

Two studies reviewed in Chapter 3 of this synthesis support the findings described in the previous section. Consistent with Swanson and Stevenson's (2002) findings that state policy had a modest but important effect on standards-based instruction, Lee (1998), in a secondary analysis of 1992 NAEP data, found that principals reported a statistically significant association between state-level curriculum development and school-level instructional change. Again, the relationship was modest. Anderson's (1995a, 1995b,1996) case studies of standards-based instructional change support the conclusion that comprehensive instructional reform is an ongoing process, requiring considerable time and learning experiences to change not only teacher behavior, but also values and beliefs about the goals of instruction.

Summary of Findings from Studies of Standards-based Instructional Guidelines

The studies reviewed in this chapter indicate a weak positive relationship between the use of standards-based instructional strategies (particularly those that emphasize higher-order skills) and student achievement. It should be noted that this relationship involves instructional strategies alone; combined with the effect of curricula, such strategies may show a stronger relationship with achievement. The size of the relationship also may be larger for certain types of assessments (e.g., those that focus on the higher-order skills that such standards are designed to strengthen, rather than basic skills) and may also be mediated through a number of organizational factors, including teacher learning opportunities.

Further, pedagogical change is taking place in the direction pointed by these standards reforms. The extent, however, to which this change meets the intent of standards-based instructional reforms like those advocated by the NCTM is unclear, based on the limitations of surveys for measuring instruction. Sweeping changes of the scope implied by standards-based reforms take considerable time and investment in teacher learning.

IMPLICATIONS FOR PRACTICE AND POLICY

The research reviewed in this chapter has several key implications for practice and policy. Instruction is changing in ways consistent with the intent of standards-based reforms such as those advocated by the NCTM, but it is changing slowly. Teachers and other educators know the words to use to describe a standards-based classroom, but actual instructional changes vary in terms of fidelity to the standards, similar to findings described in Chapter 3. Variations are related to individual teachers' beliefs about and support for the reforms, as well as their instructional skills. The changes implied in standards-based instruction require far more than adopting new strategies in class; rather they require teachers to question their own beliefs about content, students, and the learning process.

In order for teachers to make instructional changes that go beyond a cosmetic adjustment, considerable investment in their learning is required. For teachers to teach in the ways envisioned by the NCTM and other reformers, they need chances to learn the strategies, practice them, see the effects on their students' learning, and reflect on what that means. Changes in instruction can lead to changes in belief in and support for the reforms, as shown by Flexer, et al (1995). However, few teachers are given the opportunity to make such principled changes and follow them through for long enough to see the effects, particularly in today's environment of accountability. States, districts, and schools need to re-examine the ways in which teachers' professional needs are met and how accountability is operationalized.

This is particularly important since current accountability policies focusing on test results may be a disincentive for teachers to use more ambitious, standards-based instruction. Results-oriented legislation like NCLB imposes political pressure on schools and teachers to raise student test scores quickly. There are deadlines for performance and heavy sanctions for schools and teachers if performance targets are not met, including threats of school reorganization and closure. Yet we have seen from the research that standards-based instruction has a weak positive relationship with student achievement, at least as it has been operationalized on a number of large-scale tests. This is not an environment in which teachers have the freedom, time, or safety to change their practice in the revolutionary ways envisioned by the standards reformers. Test scores may increase, but a number of assessment studies have shown that raised scores do not necessarily indicate higher-level learning, but other practices, possibly including "teaching to the test" (Frederiksen & Collins, 1989; Haladyna, Nolen, & Haas, 1991). Such a policy environment provides teachers with few incentives to teach high-level skills and further, may unfairly reward teachers and schools that have learned how to raise test scores without raising instructional discourse.

In order to change instruction substantively, so that the standards movement is not yet another in a series of reforms that stop at the classroom door, teachers need to learn about and support the reform. As shown in the research, state policy has its strongest influence on ambitious standards-based instruction when it is mediated through the support of more localized teacher learning opportunities (Swanson & Stevenson, 2002). This means that a revised state policy role is called for in order to increase teacher knowledge of and receptivity to the reforms. This should focus on building local capacity and on adjusting policy so that it helps teachers better learn how to implement the reforms (Spillane & Jennings, 1997). One step for states could be to better attend to issues of instruction, rather than simply to test results. Another would be for policymakers to ensure that the assessments used for accountability are of high quality and test the desired skills — not only in terms of content, but also in terms of whether they measure higher-order skills as well as important basic skills.

CONCLUSIONS

This research review indicates that standards-based pedagogy consistent with the reform orientation of the NCTM can lead to improved student achievement, particularly on higher-order skills. Further, the research indicates that pedagogy is undergoing some changes towards more coherence with the recommendations of standards-based reforms. However, the extent to which pedagogy has changed is unclear, partially because a broad repertoire of pedagogical approaches (ranging from teacher-centered to more student-centered instruction) is appropriate for use with different aspects of content, and current measures do not adequately capture the use of appropriate pedagogy in a standards-based system. Compounding these problems, many achievement measures are not designed to measure whether students have attained ambitious learning goals, and the current accountability context emphasizes test performance.

Increased state emphasis on issues of pedagogy is likely to focus attention on limitations in the data that curtail the extent of the research in this area. For instance, survey measures of instruction should be validated and tested for the quality of information they provide, in particular, relative to whether approaches are appropriate and communicate high expectations for all students' learning. Similarly, achievement measures should be evaluated for the type of content they measure, the performance standards that they address, the levels of aggregation at which results can be determined, and the level of analysis at which it is appropriate to determine consequences. For large-scale studies, assessments that can be aggregated across a number of sites and still measure the ambitious goals of standards-based reform are needed. Lacking these, cross-site analyses like the ones employed by RAND (Klein et al., 2000) may be used as design models in further large-scale research. However, regardless, there are important database issues to resolve, including the problems of linking student achievement and teacher instruction data, preferably over time.

CHAPTER 5: THE INFLUENCE OF STANDARDS-BASED ASSESSMENT ON TEACHING AND STUDENT LEARNING

This chapter defines standards-based assessments as mandated student tests produced at a variety of different levels — school, district, regional, state, or national — that reflect or are aligned with established content standards (Clune, 2001). Thus, it uses the definition described in Chapter 1 in which "standards-based" refers to a policy approach organized around agreements about the content that students should know. Such assessments may or may not reflect reform tenets of inquiry and higher-order thinking, depending on the content and performance standards on which they are based. They are unified by how their results are used; these assessments provide the data upon which accountability decisions are based. They are to be distinguished from standards-based assessments that are used primarily at the classroom level in order to diagnose and assist student learning; rather, they are large-scale external assessments (Pellegrino, 2004).

State-level standards-based assessment programs are now in place in all but a few states, and most of the research on their effects focuses on the state–level. These programs vary considerably in terms of their comprehensiveness, aims, and quality, as well as in the relative emphases of the standards on which they are based. States are still dealing with the challenge of bringing curricula, standards, assessments, and instruction into alignment, an effort that is complicated by the need for effective professional development programs and comprehensive curriculum materials that will support such alignment (Linn, 2000; Darling-Hammond, 2003). It is also clear that the states have entered into reform with varying degrees of haste and enthusiasm. While considering this variation as an acceptable and inevitable characteristic of the reform movement, McREL seeks, however, to understand the common influences of standards-based assessments on teaching and student learning.

Standards-based assessment programs are controversial. Supporters point to the need for accountability in public school systems that are not meeting societal expectations (National Commission on Excellence in Education, 1983; Resnick & Resnick, 1992). Further, this call for accountability is amplified by the current emphasis on student equity in NCLB. Critics of standards-based assessments, on the other hand, contend that the tests provide only a limited view of what children know and are able to do and that

attaching high stakes to these tests will encourage teachers to narrow their own curricula and lose sight of the long-term goals of their students' lessons (Shepard & Dougherty, 1991; Stecher, Barron, Kaganoff, & Goodwin, 1998).

Regarding the stakes attached to standards-based assessments, student scores on these assessments are used to varying degrees in judging the success of districts, schools, administrators, teachers, and students. There are lower-stakes environments in which there are few consequences for students, teachers, schools, or districts related to test scores (Clarke et al., 2003). There are also higher-stakes environments in which students face grade-level retention (Roderick, Jacob, & Bryk, 2002), and teachers, schools, and districts face professional and financial repercussions, such as those specified under NCLB (Clarke et al.). The specific effects of these varying stakes attached to standards-based assessments have been studied by others and are beyond the scope of the current work here (see Kiplinger & Linn, 1993; Koretz, McCaffrey, & Hamilton, 2001; Clarke et al.).

What is the influence of standards-based assessments on K–12 teaching and student learning? The studies reviewed provide findings and implications that help to answer this question. The first section of this chapter concerns the influences of standards-based assessment on teacher instruction, and the second section concerns the influences on student achievement. The chapter concludes with overall findings and implications for policy and practice.

OVERVIEW OF STUDIES

A total of 31 research studies related to standards-based assessments to either teacher instruction or student achievement were identified. Three reports concern a series of survey studies based on the same group of teacher respondents (Stecher & Barron, 1999; Stecher, Barron, Kaganoff, & Goodwin, 1998; Koretz, Barron, Mitchell, & Stecher, 1996). We treat these three research reports as one extended study, which leaves 29 studies that inform the conclusions drawn in this chapter. All the studies are presented with details on the findings in Appendix E.

As a body of evidence, these 29 studies represent a wide variety of large-scale standards-based assessments. Twenty-seven of the studies examined influences of state-level testing in 19 different states. Three of these 27 studies tied influences to national standards as well as state standards (one to the NRC, one to the NCTM, and one to unidentified national standards in language arts and mathematics). Two studies focused solely on the influences of local standards (Chicago).

The studies also represent a wide variety of public school classrooms. Twenty-four studied elementary classrooms (grades K–5), 15 studied middle-school classrooms (grades 6–8), and 12 studied high school classrooms (grades 9–12). Eight of these studies included two education levels (e.g., elementary and middle-school), and another seven studies included all three levels. The different subject areas examined included 24 studies of language arts (i.e., reading, writing, and English), 23 studies of mathematics, 18 studies of science, and 18 studies of social studies. There was considerable crossover with 16 studies addressing all of these subject areas and only 7 that examined a single subject area (two in language arts, two in mathematics, two in science, and one in social studies).

Sixteen of the 29 studies reported sample and/or population representation of at-risk students. Two studies examined influences of standards-based assessments on ELLs, one investigated influences on special education students, and 13 studies reported at-risk student representation in terms of socioeconomic status, percentages of students who receive free or reduced-priced lunches, or urban-school settings.

Twenty-six of the 29 studies examined teacher instruction in response to standards-based assessment. Most of the studies relied on multiple measures including information based on surveys (n=15), teacher interviews (n=13), classroom observations (n=8), document review (n=8), and focus groups (n=3).

In contrast to the area of teacher instruction, there were relatively few studies addressing student responses to mandated assessments. Only six studies examined the influence of standards-based assessments on student achievement. (Three of the six examined both teacher instruction and student achievement.) Two studies relied solely on standards-based assessment scores as a study measure. Student assessment scores were combined with other measures in two studies, and teacher assessment of student achievement was the measure used in the remaining two studies.

Ten of the studies were technical reports, nine were published in journals, one was a published book, five were conference presentations, and four were dissertations. None of the 29 studies employed experimental designs, but two employed quasi-experimental designs. Six employed quantitative non-experimental designs, seven used mixed quantitative and qualitative methodologies, and 14 of the studies were qualitative approaches. Thus, the vast majority of the studies reviewed in this chapter used descriptive methods.

With regard to systemic influences, one study examined teacher responses to a state systemic initiative (Falk & Drayton, 2004). Mention of district-level responses to state assessments was made in other studies (Doran, 2001; Pedroza, 1998), but specific systemic influences were not the focus of these investigations.

STUDY FINDINGS

The following sections first review findings from the 26 studies that addressed teacher instruction, followed by a review of the six studies that addressed student achievement (three studies included both measures). Findings specific to subject areas (e.g. influences on mathematics achievement) are not presented here due to the large number of multi-subject studies and the overall lack of subject-specific findings.

Studies of Influences on Teacher Instruction

The 26 studies that concern the influence of standards-based assessment on teacher instruction are listed in Table 5.1. Thirteen of these studies employed qualitative designs, eight used mixed methods, and five relied solely on quantitative data. One of the studies employed a quasi-experimental research design (Doran, 2001).

The findings across the 26 studies fall into four categories: (1) changes in content coverage (or the teaching curriculum), (2) changes in classroom pedagogy, (3) changes in classroom assessment, and (4) other classroom changes. Studies in each of these categories are described in the subsections that follow.

Table 5.1. Studies of Standards-based Assessment and Teacher Instruction

Author (s) & Year	Method	Standards Addressed	Subject Area(s)*	Outcome(s)	Measures
Avery, Beach, & Coler (2002)	Mixed methods (non-experimental)	State: MN	LA SS	Teacher instruction	Teacher surveys Teacher interviews
Barksdale-Ladd & Thomas (2000)	Qualitative	State: unidentified	LA Math Science SS	Teacher instruction	Teacher interviews Teacher focus groups
Beran (2003)	Quantitative non-experimental	State: NE	LA Math Science SS	Teacher instruction	Teacher surveys
Clarke, Shore, Rhodes, Abrams, Miao, & Li (2003)	Qualitative	State: KS, MI, MA	LA Math Science SS	Teacher instruction	Teacher surveys
Daniels (1995)	Qualitative	State: VT	LA	Teacher instruction	Teacher & administrator interviews Classroom observations

Din (1996)	Quantitative non-experimental	State: KY	LA Math Science SS	Teacher instruction	Teacher surveys
Doran (2001)	Quantitative quasi-experimental	State: AZ	LA Math Science SS	Teacher instruction	Teacher surveys
Falk & Drayton (2004)	Qualitative	National: NRC State: MA	Science	Teacher instruction	Teacher interviews Classroom observations Document review
Fernandez (2004)	Qualitative	State: FL	LA Math Science SS	Teacher instruction	Teacher interviews Classroom observations Document review
Firestone, Mayrowetz, & Fairman (1998)	Qualitative	State: ME, MD	Math	Teacher instruction	Teacher & administrator interviews
Grant (2000)	Qualitative	State: NY	LA Math Science SS	Teacher instruction	Teacher focus groups
Grant (2001)	Qualitative	State: NY	SS	Teacher instruction	Teacher interviews Teacher surveys Document review
Janson (2002)	Mixed methods (non-experimental)	State: OH	Science	Teacher instruction	Teacher interviews Teacher surveys Document review
Khattri, Reeve, Kane, & Adamson (1995)*	Qualitative	various national, state, district, local	LA Math	Teacher instruction Student achievement	Teacher interviews Classroom observations
Koretz, Mitchell, Barron, & Keith (1996)	Mixed methods (non-experimental)	State: MD	LA Math	Teacher instruction	Teacher surveys
Louisville University, School of Education (1995)	Mixed methods (non-experimental)	State: KY	LA Math Science SS	Teacher instruction	Teacher surveys

McDonnell & Choisser (1997)	Qualitative	State: KY, NC	LA Math Science SS	Teacher instruction	Teacher interviews Document review
McMillan, Myran, & Workman (1999)	Mixed methods (non-experimental)	State: VA	LA Math Science SS	Teacher instruction	Teacher surveys
Moon, Brighton, & Callahan (2002)	Mixed methods (non-experimental)	State: states nationwide	LA Math Science SS	Teacher instruction	Teacher surveys Teacher focus groups Classroom observations
Schorr, Firestone, & Monfils (2003)	Qualitative	National: NCTM State: NJ	Math	Teacher instruction	Teacher interviews Classroom observations
Smith (1997)	Qualitative	State: AZ	LA Math Science SS	Teacher instruction	Teacher interviews Classroom observations Document review
Stecher & Barron (1999); Stecher, Barron, Kaganoff, & Goodwin (1998); Koretz, Barron, Mitchell, & Stecher (1996)	Quantitative non-experimental	State: KY	LA Math	Teacher instruction	Teacher surveys
Stecher & Chun (2001)*	Quantitative non-experimental	State: WA	LA Math Science SS	Teacher instruction Student achievement	Teacher surveys
Stone & Lane (2000)*	Quantitative non-experimental	State: MD	LA Math Science SS	Teacher instruction Student achievement	Teacher surveys State assessment data
Taylor, Shepard, Kinner, & Rosenthal (2003)	Mixed methods (non-experimental)	State: CO	LA Math Science SS	Teacher instruction	Teacher surveys Teacher interviews
Wong, Anagnostopoulos, Rutledge, & Edwards (2001)	Qualitative	Local: Chicago	LA Math	Teacher instruction	Classroom observations Document review

* Study is presented in both tables in this chapter because it informs both the teacher instruction and student achievement analyses

NCTM = National Council of Teachers of Mathematics

Changes in the teaching curriculum. Teachers described in this research were actively pursuing alignment between the content they were teaching (at the classroom level, we refer to this as "teaching curriculum") and the content measured by the various state tests. Doran (2001), in a quasi-experimental study that compared teachers in tested grades to those in non-tested grades, revealed the strong influence of the Arizona tests on teaching curricula in that state. The 153 elementary teachers surveyed for this study reported that they altered their teaching curricula to match the tests, and the teachers in non-tested grades reported significantly less alteration than did the teachers in non-tested grades. Across the studies in this category, there is evidence that teachers are actively changing the content that they cover to align with state assessments (Clarke et al., 2003; Firestone, Mayrowetz, & Fairman, 1998; Grant, 2001; Wong, Anagnostopoulos, Rutledge, & Edwards, 2001). The state testing program in Kentucky, for example, has reportedly encouraged those teachers to make significant changes in curricular content and objectives (Din, 1996). The Maryland teachers studied by Koretz, Mitchell, Barron, and Keith (1996) made changes in their teaching curricula, and apparently many saw these as significant changes in their teaching practice.

The findings of McMillan, Myran, and Workman (1999) further illustrate this struggle to align the teaching curriculum and the tested curriculum and also illustrate an accompanying issue of superficial, broad coverage. The Virginia teachers surveyed in the study gave priority to covering the state-assessment. For some, however, covering the required content meant adopting a pace or superficiality that made them uncomfortable. “We were pushed to get as much taught before the test as possible....we were not teaching for understanding” (p 10). Teachers in Kentucky also were aligning their teaching curricula with the state test (Stecher, Barron, Kaganoff, & Goodwin, 1998; Koretz, Barron, Mitchell, & Stecher, 1996), although a subsequent analysis suggested that the teachers’ efforts to align were misplaced and demonstrated a focus on the array of state test measures rather than on any greater student understanding of underlying content and concepts (Stecher & Barron, 1999).

Teachers, especially those expected to specialize in one or two subject areas, noted that the mandated testing programs encouraged them to broaden their teaching curricula beyond reasonable bounds. For example, in a case study by Fernandez (2004), one Florida teacher's response to her teaching in the months leading up to the state assessment was “too much to cover in such a short time frame” (p. 111). In another case study, Falk and Drayton (2004) found that Massachusetts middle school teachers were likely to respond to the state assessments by broadening the number of topics they covered. They also reported that teachers in low-SES schools, although feeling pressure to engage in this same broadening effort, were afraid to do so for fear of overwhelming their students. Ironically, both groups of teachers described feelings of having left their students ill-prepared for state testing.

In contrast to increasing the number and decreasing the superficiality of topics covered, some studies also indicate that responses to standards-based tests also can narrow the teaching curriculum. For example, Stecher and Chun (2001) reported on curriculum narrowing in the state of Washington. Two-thirds of the elementary and middle-school teachers surveyed in the study had made conscious efforts to conform to the curricular standards of the Washington state assessment, but the authors provide evidence that the teachers were covering tested topics by borrowing instructional time from non-tested subjects. This is another consequence of the use of standards-based assessments, particularly in self-contained (typically elementary) classrooms where teachers are in a position to choose between time spent on tested subjects (i.e., mathematics, reading) and non-tested subjects, especially those that may be tested in later grades (i.e., science, social studies). Teachers in these studies made curricular choices that favored core subjects, a logical extension of the focusing effects of standards. However, they also tended to exclude science experiments (Barksdale-Ladd & Thomas, 2000) and other activities that promote depth in curricular understanding (Beran, 2003; Clarke et al., 2003).

The studies presented here support the conclusion that teaching curricula are influenced substantially by standards-based assessments. The research that supports this conclusion, however, brings to light several issues around the nature of the choices that some teachers are making. Teachers may be influenced by these assessments to broaden their coverage of particular subjects and deemphasize topical depth. They also may narrow the content that they teach and deemphasize the instruction they devote to non-tested subjects.

In summary, 15 of the 26 studies analyzed for changes in the teaching curriculum provided evidence that teachers were adapting content to align with that covered by mandated standards-based assessments. In this sense, such assessments seem to have a powerful role in shaping what students learn. However, of these, four studies presented findings that teachers cover a broader range of content in response to these assessments, and four studies presented findings that demonstrated teachers narrowing the content they teach. Seven studies documented teachers' efforts to align their curricula with the assessments. Teacher choices, such as those that lead to curricular broadening or narrowing the scope of content coverage, are trends that merit attention.

Changes in classroom pedagogy. In addition to changes in the teaching curriculum (or content) in response to standards-based assessments, there is some evidence that changes also are occurring in teaching practice. We refer to teachers' choices of instructional approaches as pedagogy. Research suggests that, in some cases, standards-based assessment programs have influenced changes in pedagogical approaches to more closely approximate reform-based teaching. Stecher et al. (1998) reported that in response to the state test, surveyed teachers in Kentucky were adopting reform-oriented approaches such as open-ended questioning, extended written responses, and the use of manipulative activities in mathematics. Similarly, Koretz, Mitchell, Barron, and Keith

(1996) found that the fifth- and eighth-grade teachers surveyed in their study of the Maryland assessment program were influenced by the state assessment to teach their students higher-order skills. Teachers reported employing activities that promoted problem-solving skills, application of skills in varying contexts, and communication of mathematical concepts, all areas of emphasis on the state test

In another study of Maryland teachers, Stone and Lane (2000) used structural equation modeling to explain differences in student scores on the state assessment. The authors found that teachers' claims of using reform-oriented approaches were correlated with higher student assessment scores. In the McMillan et al. (1999) survey and case study of Virginia teachers, the authors indicated that elementary teachers were relying less heavily on whole-class lecture and seatwork in response to the state assessment, although their new and presumably reform-oriented approaches were not described. The research also included studies that described teachers who chose to develop or use new writing activities in response to the state assessment programs (Clarke et al., 2003; Daniels, 1995).

A second type of pedagogical response documented in the research is that teachers are moving their pedagogy away from reform-oriented practices in response to standards-based assessments. For example, Grant (2000), in a study using focus groups of teachers in New York state, reported that the state assessment program was encouraging teachers to employ what they saw as test-practice activities rather than the higher-order thinking activities that they preferred. Wong et al. (2001) indicated that the Chicago district standards called for interpretive and implied reasoning skills that were not reflected in the district assessments and as a result, the teachers were not emphasizing these higher-order skills in their pedagogy. Similarly, Taylor, Shepard, Kinner, and Rosenthal (2003) in their survey of Colorado teachers, identified negative changes in pedagogy related to test preparation and practice due to the state test, although they identified some positive practices as well, such as more emphasis on writing instruction.

A third type of response occurs when teachers do not change their pedagogy to align with standards-based assessments (Avery, Beach, & Coler, 2002; Firestone et al., 1998; Grant, 2001; Smith, 1997). Complicating the matter, some teachers may think they're changing pedagogy, but they are only making superficial changes. For instance, New Jersey teachers who were interviewed and observed by Schorr, Firestone, and Monfils (2003) described their teaching in terms of reform-based approaches, but these approaches were not observed by the researchers in the teachers' classrooms, a finding also described in Chapters 3 and 4. It is worth noting that some researchers who have reported changes in pedagogy have attributed these changes to sources other than standards-based testing programs. Falk and Drayton (2004), for example, linked teachers' responses to district influences rather than the Massachusetts assessment program. In Colorado, Taylor et al.

(2003) reported reform-oriented effects on pedagogy from the standards and the state test, but noted that the test resulted in test preparation activities too.

The findings across the studies demonstrate less consistency in teacher changes in pedagogy in response to standards-based assessments than changes in the teaching curriculum. Some of this may be due to the wide variability in the nature of the assessments themselves. In Kentucky and Maryland, which has tests emphasizing higher-order skills, some studies have indicated pedagogical change to support the development of those skills. But state tests are very different from one another. Further, the decision to make pedagogical changes may depend on the stakes attached to the tests, teachers' individual capacities to accommodate changes, or to the strength of other influences on pedagogical change. These findings suggest that, while the emphasis of standards-based assessments may be important in terms of their influence on pedagogy, the tests themselves may not be enough to encourage pedagogical changes.. Comprehensive efforts including professional development programs, opportunities for staff collaboration, and aligned curricula may be necessary to produce change at this depth, particularly if the changes desired include those advocated by standards reformers like the NCTM.

There is also evidence that standards-based assessments can result in pedagogical consequences inconsistent with the intent of standards as high expectations for all students' learning. First, it is clear from the research that under certain conditions, standards-based assessment programs promote a teaching methodology of their own, one most often referred to as "test preparation." These test-preparation sessions are designed to simulate the actual test, without necessarily providing learning that will transfer into other settings (Barksdale-Ladd & Thomas, 2000; Fernandez, 2004; Smith, 1997; Stecher & Chun, 2001; Wong et al., 2001). In one study, these strategies are described as an effective way to improve students' test scores (Stone & Lane, 2000). Second, some teachers also reported increasing their use of recitation as a way to cover material and to satisfy the perceived intent of the testing programs (Moon, Brighton, & Callahan, 2002; Wong et al., 2001). The authors of these studies do not indicate that these recitation or whole-class lecture sessions were demonstrated as productive ways to improve student test scores.

In summary, six studies indicated that teachers were encouraged by standards-based assessments to align their pedagogy with reform-oriented pedagogy aligned. Two additional studies demonstrated that standards-based assessment programs were encouraging teachers to depart from reform-oriented teaching toward more traditional teacher-centered practices. In contrast, seven studies noted no pedagogical change in response to the assessments. In terms of unintended consequences of the assessment programs, six studies described teachers who were choosing to employ test-preparation activities in response to standards, and two studies described teachers who were choosing

recitation as a pedagogical approach that would best prepare students for the mandated standards-based assessments.

Changes in classroom assessment. Among the other findings addressed in these 26 studies, the research provides evidence that teachers adapted their classroom testing formats to mirror state assessment formats. The survey research conducted by Koretz et al. (1996) revealed that Kentucky teachers decreased their dependence on multiple-choice items in their classroom tests in order to better prepare their students for the open-ended format of the state assessments. In a qualitative interview study, McDonnell and Choisser (1997) compared the practices of Kentucky teachers to their counterparts in North Carolina and found that the Kentucky teachers were more likely to employ extended written response formats in classroom testing. The North Carolina teachers, on the other hand, were more likely to employ multiple-choice testing consistent with the testing format in their state.

In fact, most of the studies that provided evidence regarding classroom assessments found that teachers responded to standards-based assessments by aligning their classroom tests with the mandated assessment format in that particular state. McMillan et al. (1999) reported that elementary teachers were using multiple-choice tests in the classroom more often to mirror the format of the Virginia assessment. Clarke et al. (2003) indicated that most of the teachers responding to the authors' survey described themselves as designing classroom assessments that emphasized writing activities, encouraged critical-thinking skills, and asked students to explain results, all in response to the influence of the state testing programs in Kansas, Michigan, and Massachusetts. Similar responses were documented for teachers in Kentucky (Din, 1996) and Maryland (Stone & Lane, 2000).

Although these studies suggest that the changes made by teachers may be limited to changes in assessment formats, one study described teachers who made deeper changes in their classroom testing practices (Louisville University, School of Education, 1995). In this survey study of Kentucky teachers, the authors noted that testing-year teachers were more likely to employ performance assessments as an integrated part of their classroom testing and instructional processes than teachers who were teaching a non–tested grade.

Without question, the research is mixed; in some cases, teachers did nothing to adapt assessment practices. Minnesota teachers surveyed by Avery et al. (2002) were reticent to change any of their classroom practices including approaches to classroom assessment. A similar result was documented by Janson (2002) in a case study of Ohio elementary teachers who depended on publisher-produced curricula and assessments and were therefore unwilling to make any significant changes in assessment practices despite state-level pressure to do so.

In summary, seven studies described teachers who were adjusting their assessment formats in response to the state's mandated assessment programs. Three of these studies specifically referred to teachers moving toward or away from multiple-choice formats (depending on the nature of the state test), and one study indicated that standards-based assessment was encouraging teachers to adopt performance testing as an integrated part of their instruction in Kentucky. Conversely, two studies identified no changes in classroom testing approaches.

Other classroom changes. Three studies noted that standards-based assessment programs placed a burden on teachers in terms of time and/or logistics. The teachers studied by Avery et al. (2002), who reported making few if any changes in practice, also expressed a need for more planning time to accommodate the requirements of the state-level assessment program in Minnesota. Nebraska teachers who were surveyed by Beran (2003) identified an increased workload that accompanied the state assessment program in that state. In a case study by Daniels (1995), teachers in Vermont identified their workload increase as a function of the paperwork that resulted from the state's portfolio assessment requirements. None of the studies analyzed suggested that teachers perceived the mandated testing programs as making their jobs easier.

Summary of influences on teacher instruction. The 26 studies related to teacher instruction that were reviewed support several findings. The evidence suggests that standards-based assessments are strong influences on teachers' decisions about what they teach in terms of content. There is also a need for attention to the consequences of these decisions, particularly with respect to broadening and narrowing the curricula. The evidence also suggests that standards-based assessments influence pedagogy in some classrooms by encouraging reform-based or assessment-aligned approaches, although this influence also can result in test preparation approaches. It is likely that the nature of the particular test in question, the stakes attached to test results, and the available support for change all play some role in this relationship. Finally, standards-based assessments encourage changes in classroom testing practices, but these changes may be limited to shifts in testing formats and are highly dependent on the format of particular tests in particular locations.

Studies of Influences on Student Achievement

Six studies that addressed the influence of standards-based assessments on student learning are listed in Table 5.2. Of these six studies, one employed a quasi-experimental design, three employed a quantitative non-experimental design, and two were solely qualitative approaches. Three of the six studies analyzed here also identified influences on teacher instruction and were described in the previous section (Khattri et al., 1995; Stecher & Chun, 2001; Stone & Lane, 2000).

Table 5.2 Studies of Standards-based assessment and Student Achievement

Author (s) & Year	Method	Standards Addressed	Subject Area(s)	Outcome(s)	Measures
Khattri, Reeve, Kane, & Adamson (1995)*	Qualitative	various national, state, district, local	LA Math	Teacher instruction Student achievement	Teacher interviews Classroom observations
Pedroza (1998)	Qualitative	State: TX	LA Math Science SS	Student achievement	Teacher & administrator interviews State assessment data
Roderick, Jacob, & Bryk (2002)	Quantitative quasi-experimental	Local: Chicago	LA Math	Student achievement	Local assessment data
Schulte, Villwock, Whichard, & Stallings (2001)	Quantitative non-experimental	State: NC	LA	Student achievement	State assessment data
Stecher & Chun (2001)*	Quantitative non-experimental	State: WA	LA Math Science SS	Teacher instruction Student achievement	Teacher surveys
Stone & Lane (2000)*	Quantitative non-experimental	State: MD	LA Math Science SS	Teacher instruction Student achievement	Teacher surveys State assessment data

*Study is presented in both tables in this chapter because it informs both the teacher instruction and student achievement analyses

Roderick, Jacob, and Bryk (2002) conducted the only quasi-experimental study in this set. This study examined the influences of Chicago's gate promotion system (testing for promotion in grades 3, 6, and 8) on Chicago students' reading and mathematics test scores. They compared student achievement in the targeted grades in the years prior to and after the establishment of high-stakes testing for promotion. The study found that students' scores had increased in response to the advent of the testing program, and that low-performing schools were most positively affected. The lowest-performing reading students and the highest-performing math students showed the largest achievement gains under the testing policy. Overall the high-performing students showed such small gains under the policy that the authors questioned the universal effectiveness of the testing program. Due to the nature of the study, it was difficult for the researchers to attribute the increases in achievement observed to a particular factor.

On the other hand, Stecher and Chun (2001) combined their analyses of student test score data with teacher and principal survey data, as well as student demographics. Their goal was to understand the student test score differences between schools and the changes

observed in student scores over one year in the state of Washington. The authors identified student demographics as the only significant predictor of student test score differences despite the researchers' attempts to link these differences to reform-based teaching practices.

The six studies in this group provide few coinciding results, which is predictable, given the considerable variation in achievement measures. The exception to this finding is that Roderick et al.'s (2002) results were supported by the results of an earlier study by Khattri et al. (1995). The authors of the earlier study combined teacher interview data with classroom observations from several states and found that students were improving in reform-emphasized skills such as reading, writing, and mathematics.

The other researchers represented in this set of six studies made a variety of claims. Pedroza (1998) reported that the Texas testing program had no positive influences on the achievement levels of the English-language learners in a southern-border district. In North Carolina, elementary school students with learning disabilities were reported to be responding positively to that state's testing program (Schulte, Villwock, Whichard, & Stallings, 2001). Stone and Lane (2000), as described previously, indicated that students in Maryland produced higher achievement scores after being exposed to test-preparation activities.

Summary of influences on student achievement. The lack of evidence and the diversity in research findings make it difficult to draw overall conclusions regarding the influence of standards-based assessments on student achievement. However, it seems clear that, since testing alone does not substantively affect pedagogy, it is unlikely that testing alone would affect student achievement.

Studies with a Secondary Focus on Standards-based Assessment

Studies presented in the previous chapters of this synthesis also inform a discussion about the influences of standards-based assessments on teacher instruction and student achievement. A review of these studies further confirms the findings that are summarized and the conclusions that are made here. For example, the studies by Firestone, Camilli, Yurecko, Monfils, and Mayrowetz (2000) and Johns (2004) both noted that mathematics teachers were actively choosing their teaching curricula to mirror the curricula covered on the standards-based assessments. In a study of special education teachers and students, McLaughlin (2000) confirmed the notion that teachers broaden their teaching curricula in response to the pressures of standards-based assessments. Smith (2003) described social studies teachers who resorted to "cramming," a test-preparation activity, in an effort to cover the topics to be tested.

Summary of Findings from Studies of Standards-based Assessments

Three major findings related to the influence of standards-based assessments on teacher instruction are provided. First, teachers in these studies aligned the coverage and sequence of their teaching curriculum to the content of standards-based assessments. The research, however, identifies several choices teachers in these studies are making for consideration by policy makers. On one hand, teachers seem to be narrowing the curriculum so that they focus on topics that are tested — which could reflect more emphasis on content areas that have been determined more important. Whether this is a positive or negative outcome needs to be further examined, especially in terms of the value of those content areas that are being deemphasized. Teachers in some cases are broadening the number of topics they teach to accommodate mandated tests that cover large numbers of topics, which decreases curricular depth and full student comprehension.

Second, changes in pedagogy were less consistent than in content coverage. Changes in instruction, it seems, are deeper and less likely to be influenced by standards-based assessment programs. It is important to note that although the research reveals examples of teachers who do not make pedagogical changes, the research also describes teachers who made substantial changes in practice by adopting approaches that enhanced higher-order thinking skills, as well as approaches that enhanced creative skills such as writing. In each case, these changes were in response to standards-based assessments that encouraged the development of these student skills; the nature of the test is an important factor in interpreting these results. The research also illustrated the concern that teachers may be adopting test-preparation and other non-reform-oriented practices that are unintended consequences of standards-based assessment programs.

Third, standards-based assessment programs appear to encourage changes in classroom assessment. Teachers are choosing to adopt the formats of standards-based assessments for their own classroom-level tests. In some cases this involves eliminating or adding a common test format (e.g. multiple-choice items), and in other cases the change is a deeper one aligned with the concepts for learning represented in a standards framework. In the best of cases, the teachers moved beyond mere format changes to adopting a substantial change in their classroom assessment strategy to better align their goals to those of the state or district.

IMPLICATIONS FOR PRACTICE AND POLICY

The research reviewed in this chapter indicates that tests matter. Teachers are actively modifying the content they teach and the sequence in which they teach it in response to standards-based assessments. Due to these responses, top-to-bottom curricular alignment

is readily attainable. However, the functionality and stability of the overall system is dependent on the standards on which it is based, so it is of paramount importance that these standards are of high quality. The nature of the assessments that measure these standards is also important; they should be well-aligned in terms of content and should reflect high expectations for all students' learning.

Attention must be given to issues of curricular "broadening" and "narrowing" as well as other consequences of standards-based assessment programs. The quality of standards-based assessment programs should be judged in terms of how well they induce intended classroom-level curricular decisions. This means that policymakers need to have substantive discussions about what the intent of these reforms are and to examine whether what's happening is in line with their intent. For instance, it is unlikely that policymakers want to encourage superficial learning at the expense of deep understanding, although the research indicates that tests covering a broad number of topics encourage this effect. Assessment programs may need modification to correct for some of these effects and should be evaluated on an ongoing basis for their effects on classroom practice.

Teachers respond to assessment formats used, so testing programs must be designed and administered with this influence in mind. Tests that emphasize inquiry, provide extended writing opportunities, and use open-ended response formats or a portfolio approach tend to influence instruction in ways quite different from tests that use closed-ended response formats and which emphasize procedures. Any change in testing content or format should consider the potential results on classroom practice, and should evaluate those changes in light of whether such results are desired or not. Potential changes to be considered include the use of practice testing and teacher burden of the assessment program.

Changes in pedagogy and classroom assessment strategies consistent with the intent of standards reformers like the NCTM may be more or less encouraged by standards-based assessments, depending on the individual nature of the assessment. Further, these changes may be encouraged by stakes or rewards, but the test alone is insufficient to support sweeping pedagogical changes like these; such deeper changes also need to be supported through additional training and resources. There is no indication in the reviewed research that deep changes in teaching practice will happen easily or quickly, consistent with findings in Chapters 3 and 4.

CONCLUSIONS

Standards-based assessments are influential in changing teacher practice in surface ways, such as changes in the content taught or the sequence in which it is taught. They also have changed instruction in some ways that may or may not be consistent with the idea of

standards expressing high expectations for all students' learning. The most notable examples of this is through practice-testing or format familiarization. The influence of such assessments on the deeper changes in practice implied by NCTM and other reformers is less consistent. Teachers do not consistently use reform-oriented pedagogy or aligned classroom assessment strategies in response to standards-based assessments. This may be due to the considerable variation in the nature of standards-based assessments; the content they cover and the formats they use do not consistently require that teachers adopt such pedagogy. In cases where testing content and format do encourage problem-solving and open-ended reasoning, there is some indication that pedagogy is adjusted accordingly.

In the best of cases, standards-based assessment programs appear to influence teachers to adopt a range of reform-based classroom practices that are consistent with these assessments. Some teachers are teaching reading with an emphasis on understanding, some are teaching mathematics with an emphasis on higher-order thinking, and some are teaching writing with an emphasis on extended and thoughtful expressions.

The available research, however, reveals several issues with respect to the choices teachers are making in response to standards-based assessments. In these studies, teachers are neglecting non-tested topics and concepts, they are responding to the breadth of curricular coverage by choosing not to pursue depth in understanding, and they are teaching students to take tests well rather than to be better learners. While a greater focus on certain content areas may be a positive thing, a focus on surface knowledge and test scores at the expense of deeper understanding goes against the concept of standards as high expectations for all students' learning.

These conclusions are tempered by the obvious need for better research on the impacts of standards-based assessments on instruction and student achievement. Future research efforts should cover the complete spectrum of potential influences and use rigorous research designs that include matched comparison groups that do not receive the treatment. However, this raises a dilemma under NCLB, which requires standards-based assessment programs in each state. For, in order to identify the influence of standards-based assessments on student achievement or teacher instruction, and, more importantly, to attribute that influence specifically to the role of standards-based assessments, researchers need access to comparison sites that do not use these types of assessments. Perhaps federal entities and other policy agencies might consider granting states and local agencies waivers to NCLB and state assessment requirements as a condition of their full participation in rigorous research about alternate assessment models.

Such studies would identify the specific influences of standards-based assessments on teaching practice in a variety of different settings. They would address the role of stakes, and how that relates to the influence of assessments on instruction. They would then

continue on to examine the changes in student learning that result from these changed practices, using multiple sources of evidence about student learning.

CHAPTER 6: SUMMARY AND CONCLUSIONS

This chapter begins with a summary of findings on the influences of K–12 standards on teaching and student learning. This summary is followed by a discussion of research issues related to standards-based education. The final section presents conclusions and implications of this research synthesis.

SUMMARY OF FINDINGS

The studies reviewed in this synthesis varied in the methods they used and the variables they addressed. This section describes that variation and summarizes the findings across the studies.

Overview of Reviewed Studies

A total of 113 different research studies that addressed the influence of standards-based variables were reviewed: 48 studies of standards-based curriculum, 36 studies of standards-based instructional guidelines, and 29 studies of standards-based assessment. Of the total, 71 studies examined influences on teacher instruction, and 56 examined influences on student achievement, including some studies that measured both. The majority of the studies on curriculum and instructional guidelines addressed national content standards, while almost all the studies of assessment concerned state content standards (as represented on mandated state assessments). None of the studies specifically addressed the influence of performance standards. Across the three variables, both elementary and secondary education levels were represented. About 40 percent of the studies included at-risk student populations, with most identifications based on SES. There were only four studies that examined influences on learning disabled students, and nine studies included English language learners. In the areas of curriculum and instructional guidelines, the research focused on mathematics and science. In the research on standards-based assessments, most studies addressed multiple subject areas, particularly mathematics and language arts.

The methodology used in the reviewed research was predominantly descriptive. researchers identified only two studies that used an experimental research design and 17 studies that employed quasi-experimental designs. An additional four studies using mixed methods included a quasi-experimental design. McREL researchers reviewed 34 studies

that were quantitative but non-experimental, 33 qualitative studies, and 27 studies that used mixed methods in which both quantitative and qualitative data were collected. Although few of the studies used experimental methods, 47 were published in journals. Other sources were technical reports, obtained mainly from websites (n=33), conference presentations obtained through the ERIC database (n= 21), dissertations (n=10), and books (n=2).

Synthesis of Study Findings

A narrative synthesis such as this does not lend itself easily to quantitative summaries of findings. However, it is possible to identify trends in the data across the studies and to point out possible moderators or factors that influence the findings based on individual research studies. With this approach in mind, the following sections examine the reviews of the studies that addressed curriculum, instructional guidelines, and assessment.

The review of standards-based curriculum studies revealed predominantly positive influences on student achievement, including that of at-risk students (Woodward & Baxter, 1997). The majority of these studies used quantitative quasi-experimental designs comparing students who were learning a standards-based (i.e., reform-oriented) curriculum with those learning a more traditional curriculum. Because this design is more appropriate for identifying effects of an intervention than non-experimental or qualitative designs, there is reason to be confident in the overall findings. Elements that may affect the relationship between the use of standards-based curriculum and student achievement are the time spent on the curriculum and how well-aligned the achievement measure is to the curriculum. Several studies indicated that student achievement improved more with longer exposures to a standards-based curriculum (Marx et al., 2004; Speas, 2004). Some studies suggested that a disparity between the achievement measure and curricular goals initially negatively influenced the student achievement outcome (Ridgway et al., 2003), although this effect disappeared over time, and other studies provided evidence that students can demonstrate the knowledge gained from a standards-based curriculum on traditional standardized tests (Carroll, 1997).

Standards-based curricula had a less consistent influence on teacher instruction, although in most of the studies reviewed, teachers were changing their instruction to reflect more reform-oriented approaches. A possible modifier of this effect was the extent to which systemic supports were in place and aligned with the curriculum; a standards-based curriculum alone did not influence instruction unless assessments and materials were aligned with that curriculum. Another modifier may be teachers' perceived lack of time for preparation of the instruction required for a standards-based curriculum (VanSledright, 2002). Consistent with studies of instructional guidance, McREL's inquiry into curriculum documented a mismatch between teachers' expressed knowledge and awareness of standards-based reform and their reported and/or observed instructional

practices (Schneider, et al., 2005). Awareness and knowledge emerge prior to implementation, which, according to some teachers, is influenced by issues of time and stress.

The reviewed studies of standards-based instructional guidelines found evidence for a small positive relationship between standards-based (i.e., reform-oriented) instructional practices and student achievement. The main moderator of this finding was the variable nature of achievement measures used. For example, in studies of mathematics instruction, achievement tests that emphasized higher-order skills were more likely to show a positive relationship with standards-based instruction than tests that focused more on basic skills such as computation (Brenner et al., 1997). However, other studies indicated that student performance on traditional assessments is not negatively affected by reform-oriented instruction (Mayer, 1998. Another possible moderator identified in some studies was differential student access to reform-oriented pedagogy, suggesting, for example, that historically low-performing students may experience more traditional forms of instruction than other students (Lubienski, 2002, Mayer, 1998).

As was the case for studies of curriculum, results concerning the influences of standards-based instructional guidelines on teacher instruction were inconsistent. Some studies indicated broad changes in pedagogy, consistent with a reform orientation (Swanson & Stevenson, 2002), while other studies noted that such changes are not widespread, and that not all students have access to them. An important element identified in some studies was the measure of instructional practices. On surveys, teachers reported being knowledgeable about standards and using standards-based practices in their classrooms, but observations of their instruction indicated otherwise (Spillane & Zeuli, 1999). Teachers appear to be changing their pedagogy in ways that are generally consistent with the intent of reformers, but there is considerable variability among teachers. This might be due to the degree of difficulty that teachers experience when they attempt to align their practices with the intent of standards, or it might relate to teachers' variable access to learning opportunities and supplemental instructional resources. Substantive instructional change toward a reform-oriented pedagogy takes considerable time and learning resources (Flexer et al., 1995; Smith, 2000).

All but three of the studies of standards-based assessment addressed influences on teacher instruction, and the six studies that examined student achievement were inconsistent in their findings. The latter might be due to the variation in the content and format of different standards-based assessments currently in use. Almost all the assessment studies used descriptive methods, with only one employing a quasi-experimental research design.

Based on this body of evidence, it is clear that standards-based assessments have a strong impact on instruction. Across the studies there was evidence that teachers changed what they were teaching in the classroom (i.e., their "teaching curriculum") to align with state assessments (Clarke et al., 2003; Firestone et al, 1998), but this took a variety of different

forms, much of which was due to the nature of the specific assessment. In terms of content coverage, some secondary teachers were broadening their teaching curricula to include more topics in response to testing pressures (Falk & Drayton, 2004), while other studies reported that elementary and secondary teachers were narrowing their teaching curricula to the exclusion of non-tested subjects (Stecher & Chun, 2001). There were indications that aligning teaching curricula with the state test led some teachers to focus on testing rather than on greater student understanding of underlying content and concepts (Stecher & Barron, 1999). Test content and format appear to moderate the influences of standards-based assessments on instruction. In some studies, teachers reported using test-preparation sessions to simulate standards-based assessments (Barksdale-Ladd & Thomas, 2000; Fernandez, 2004), and some teachers reported increasing their use of recitation as a way to cover all the material tested (Moon et al., 2002).

RESEARCH ISSUES

The review of research on standards-based education was a daunting task. McREL researchers started with over 4,500 abstracts of studies that purported to investigate the influence of K–12 standards. The majority of these studies were eliminated from further consideration because they were not empirical and/or did not actually investigate standards; nearly 700 studies were considered for possible inclusion. Of these, 113 separate studies met the final inclusion criteria.

Overall, the research base on standards-based education is limited in both quality and breadth. Only 17 percent of the studies reviewed used experimental or quasi-experimental methods, and most of these studies investigated the influence of standards-based curriculum. On the other hand, McREL's review does include many well-designed survey and qualitative studies that met the criteria for rigor appropriate for the methodology employed. Perhaps with the NCLB mandate for standards, it is important to ask not only whether standards work but also under what conditions they work best. Nonetheless, the small number of studies that met our inclusion criteria indicates the need for more rigorous research on standards-based education.

The vast majority of the studies McREL reviewed focused on mathematics. The reason for this is most likely the strong influence of the NCTM (1989), which led to implementation of standards-based reforms in mathematics earlier than in the other content areas. Another likely influence was NSF's support of systemic initiatives designed to reform mathematics education as well as science education. Evaluations of the NSF SIs have produced many research studies related to standards-based education in mathematics and science. In contrast, the development of standards in language arts lagged the development of mathematics standards by several years, and the consensus around these standards generally has been less strong and slower to develop than that for

mathematics (Valencia & Wixson, 2000). This may be one reason for the lack of studies related to language arts standards.

Chatterji's review (2002) demonstrated that there has been little coherence in the research and evaluation of the influence of standards. Chatterji and others contend that a systems perspective is needed to study standards-based reform (Puma et al. 2000; Dutro, 2002). The synthesis confirms the lack of coherence in the research on standards-based education, but based on the studies of systemic reform that McREL screened for inclusion, the synthesis authors agree with Dutro that with systemic research approaches, it is difficult to attribute observed teacher and student outcomes to the influence of standards. Synthesis authors excluded studies that described the general aspects of systemic reform exhibited by schools or districts in an initiative and then attempted to connect the whole of these to instruction and/or student achievement. Such studies do not investigate the relative importance of the components of systemic reform nor the quality of these components, making it difficult to generalize the findings to other schools and districts. For example, it is impossible to know from some studies of systemic reform to what degree the outcomes are due to the standards framework, the curriculum, the assessment, the accountability policy, the professional development, the school governance, or more than one of these elements. Without knowledge about the relative influence of various components of reform, it is difficult to make recommendations to administrators and policymakers regarding the sequence of strategic action or optimal resource allocation.

This synthesis does indicate, however, that in a standards-based system, curriculum, instructional guidelines, and assessment should be aligned to obtain favorable outcomes. Perhaps most importantly for schools today, studies of standards-based assessment demonstrate that, when the test is not well-aligned with high expectations for student performance, it may influence teachers to teach in ways inconsistent with standards as envisioned by reformers like the NCTM (1991). It may encourage pedagogy that focuses on test preparation, that is, instruction designed to raise scores on the test without a parallel increase in learning (Grant, 2000; Wong et al, 2001; Taylor et al, 2000). However, the opposite also has occurred with teachers using reform-oriented practices because related content and test formats are emphasized on the state test (Koretz et al., 1996).

The findings of many of the studies reviewed for this synthesis indicate that teachers need professional development to overcome barriers to standards-based reform (Ross et al., 2002). For example, the NSF SIs embraced several elements of systemic reform, but most have implemented professional development as their primary means of influencing instruction and student achievement (Kahle & Kelly, 2001). Although a review of research on standards-based professional development was beyond the scope of this synthesis, it is a research area that deserves further attention.

CONCLUSIONS AND IMPLICATIONS

The results of this synthesis lead to several conclusions and implications for practice, policy, and research related to standards-based education:

- *Standards-based curricula and standards-based instructional guidelines can have positive influences on student achievement.* This is particularly apparent when curricula and instructional guidelines are consistent with the more constructivist definition of "standards-based." The measure used to determine student achievement is important , with tests that measure higher-order skills showing more improvement relative to reform-oriented content and instruction. But in general, student experience with more reform-oriented content and instruction is associated with better learning of higher-order skills and does not negatively affect performance on measures of important basic skills. However, the effect of standards-based curricula and instruction is mediated through a number of factors, including teacher receptivity and knowledge of the reforms, as well as how well policies and materials are aligned with the standards. All of these affect the relationship between standards and student achievement and should be taken into account as part of a system-wide approach to education.

- *Standards-based curricula and standards-based instructional guidelines can influence teachers toward adopting reform-oriented instructional practices.* High-quality standards, translated into curricula that are well-aligned with instructional materials and student assessments, are necessary but not always sufficient to help teachers to change their practice. For teachers to teach in the ways envisioned by standards reformers, they need opportunities to learn reform-oriented strategies, practice them, and observe their effects on student learning. If such changes in instruction are to become widespread, it may be necessary to re-examine the ways in which teachers' professional needs are met and how accountability is operationalized. Clearly, changing instruction to align with the constructivist interpretation of standards is both personally challenging and time-consuming for teachers.

- *Standards-based assessments, as interpreted in state accountability programs, influence both the content and pedagogy of classroom instruction.* Tests matter — the content covered, the format used, and the application of their results — all influence teacher behavior. Depending on the particular assessment, teachers may broaden or narrow their teaching curriculum, use more or less reform-oriented instructional practices, and emphasize more or less test preparation in their instruction. Therefore, the quality of assessments and the standards on which they are based is of great importance. State education administrators should evaluate the quality of their state's standards-based assessments not only in terms of reliability and

validity, but also in terms of their potential to induce teachers to make favorable curricular and pedagogical decisions. This means that, at the policy level, leaders need to discuss substantively what is desired in terms of teaching and learning. Given the evidence that state standards-based assessment programs can have strong effects (in different directions) on instruction, it is the responsibility of policy to consider how to encourage desired effects and discourage harmful ones in evaluating current assessment programs and when considering changes to those programs.

- *At-risk students may experience less access to reform-oriented instruction than more advantaged students.* The research shows that at-risk students can benefit from reform-oriented instruction, but that their access to it is not assured. Administrators and policymakers need to find ways to make instruction equitable among diverse groups of students if standards really mean high expectations for all students to learn.

- *Results from studies of standard-based education are dependent on how the outcomes are measured.* In studies of instruction involving self-report, teachers tend to overestimate their use of reform-oriented standards-based instructional practices compared to classroom observations of their instruction. Disparity between a student achievement measure and the goals of a standards-based curriculum can negatively influence student scores. Similarly, achievement tests that emphasize higher-order skills are more likely to show a positive relationship with standards-based instruction compared to tests that focus on basic skills. In conducting research on standards-based education, researchers and evaluators should consider the influence and limitations of the measures on outcomes, and so should those who are using the research. Multiple sources of evidence should be used to the greatest extent feasible in order to validate the measures being used.

- *The breadth and quality of research on standards-based education needs to improve.* Now that America is entrenched in standards-based reform, the research should address not only the question of "does this work?" but also "how can we make it work it better?" Research can better help answer these questions, but more studies are needed that (1) more clearly define the indicators of standards-based education, (2) address language arts and social studies, and (3) use more rigorous research methodology. In particular, more experimental and quasi-experimental studies are needed in order to attribute changes in teaching and learning to standards-based curricula, instructional guidelines. and assessments. Within the current policy context, it may require creative steps to encourage participation in such studies. For example, in studying the effects of standards-based assessments, which are required of all states under NCLB, comparison sites typically are not available. One strategy might be for federal entities and other policy agencies to consider granting waivers to NCLB and state assessment requirements as a

condition of schools' full participation in rigorous, possibly randomized research about policy effects on teaching and learning.

Standards-based policies influence teaching and student learning in K–12 classrooms, but the nature of these influences depends on how the policies are perceived and implemented by teachers. According to Resnick and Zurawsky (2005), if standards-based reforms are to achieve their promise of high standards for all students, then more attention and resources are needed for the instructional support system in schools, including curriculum, instruction, professional development, and interventions for struggling students. The results of this research synthesis support this observation and suggest that the next step in improving standards-based education is to help teachers in their efforts to implement standards policies in their classrooms.

REFERENCES

(References marked with an asterisk indicate studies reviewed for the synthesis.)

Achieve, Inc. (2002). *Staying on course: Standards-based reform in America's schools: Progress and Prospects.* Washington, DC: Author.

*Adams, A. A. (1999, November). *An evaluation of a district-developed NCTM standards-based elementary school mathematics curriculum.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Point Clear, AL. (ERIC Document Reproduction Service No. ED435759)

*Adams, K., Brower, S., Hill, D., & Marshall, I. (2000). *The components of an effective mathematics and science middle school: Standards, teaching practices, and professional development*. Corpus Christi, TX: The South Texas Research and Development Center. (ERIC Document Reproduction Service No. ED449032)

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

American Association for the Advancement of Science. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

American Federation of Teachers. (2001). *Making standards matter 2001: A fifty-state report on efforts to implement a standards-based system.* Washington, DC: Author. Retrieved June 30, 2005, from http://www.aft.org/pubs-reports/downloads/teachers/msm2001.pdf

*Anderson, R. D. (1996). *Study of curriculum reform. Volume I: Findings and conclusions.* Boulder, CO: University of Colorado, Boulder. (ERIC Document Reproduction Service No ED397535)

*Anderson, R. D. (1995a). *Study of curriculum reform. Volume II: Case studies. Studies of education reform*. Boulder, CO: University of Colorado, Boulder. (ERIC Document Reproduction Service No. ED397536)

*Anderson, R. D. (1995b). *Study of curriculum reform. Volume III: Technical appendix--research design and methodology. Studies of education reform.* Boulder, CO: University of Colorado, Boulder. (ERIC Document Reproduction Service No. ED397537)

*Avery, P. G., Beach, R., & Coler, J. (2002). *The impact of Minnesota's "Profile of Learning" on teaching and learning in English and social studies classrooms*. Retrieved May 7, 2004, from http://epaa.asu.edu/epaa/v11n7/

Apthorp, H. S., Dean, C. B., Florian, J. E., Lauer, P. A., Reichardt, R., & Snow-Renner, R. (2001). *Standards in classroom practice: Research synthesis* (REL Deliverable #2001-17). Aurora, CO: Mid-continent Research for Education and Learning.

*Ba, H., Admon, N., & Anderson, L. (2002). *A quantitative investigation of teachers and the JASON multimedia science curriculum: Reported use and impact: Year two evaluation report*. Retrieved May 12, 2005, from http://www2.edc.org/CCT/admin/publications/report/QuantitativeinvestJason.pdf

*Ba, H., Martin, W., & Diaz, O. (2001). *The JASON project's multi-media science curriculum impact on student learning: Final evaluation report year one.* Retrieved May 12, 2005, from http://www2.edc.org/CCT/admin/publications/report/jason-yr1.pdf

*Barksdale-Ladd, M. A., & Thomas, K. F. (2000). What's at stake in high-stakes testing: Teachers and parents speak out. *Journal of Teacher Education, 51*(5), 384–397.

Barley, Z., Lauer, P. A., Arens, S. A., Apthorp, H. S., Englert, K. S., Snow, D., & Akiba, M. (2002). *Helping at-risk students meet standards: A synthesis of evidence-based classroom practices* (REL Deliverable #2002-20). Aurora, CO: Mid-continent Research for Education and Learning.

*Barth, P., Haycock, K., Jackson, H., Mora, K., Ruiz, P., Robinson, S., et al. (Eds.). (1999). *Dispelling the myth: High poverty schools exceeding expectations*. Washington, DC: The Education Trust. (ERIC Document Reproduction Service No. ED445140)

*Baxter, J. A., Woodward, J., & Olson, D. (2001). Effects of reform-based mathematics instruction on low achievers in five third-grade classrooms. *Elementary School Journal, 101*(5), 529–547.

*Ben-Chaim, D., Fey, J. T., Fitzgerald, W. M., Benedetto, C., & Miller, J. (1998). Proportional reasoning among 7th grade students with different curricular experiences. *Educational Studies in Mathematics, 36*(3), 247–273.

*Beran, R. J. (2003). *Fourth-grade teachers' perceptions of Nebraska's state standards process in elementary schools.* Unpublished doctoral dissertation, University of Nebraska, Omaha, Omaha, NE.

*Blank, R. K., Kim, J. J., & Smithson, J. (2000). *Survey results of urban school classroom practices in mathematics and science: 1999 report. Using the survey of enacted curriculum conducted during four USI site visits. How reform works: An evaluative study of national science foundation's urban systemic initiatives. Study monograph no. 2*. Norwood, MA: Systemic Research, Inc. (ERIC Document Reproduction Service No. ED454361)

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73*(2), 125–230.

*Brenner, M. E., Mayer, R. E., Moseley, B., Brar, T., Duran, R., Reed, B. S., et al. (1997). Learning by understanding: The role of multiple representations in learning algebra. *American Educational Research Journal, 34*(4), 663–689.

*Briars, D. J., & Resnick, L. B. (2000). *Standards, assessments--and what else? The essential elements of standards-based school improvement. (CSE technical report 528).* Los Angeles, CA: UCLA Center for the Study of Evaluation (ERIC Document Reproduction Service No. ED450137)

*Bridge, C. A., Compton-Hall, M., & Cantrell, S. C. (1997). Classroom writing practices revisited: The effects of statewide reform on writing instruction. *Elementary School Journal, 98*(2), 151–170.

*Burian-Fitzgerald, M., McGrath, D. J., & Plisko, V. (2003). Mathematics teachers' familiarity with standards and their instructional practices: 1995 and 1999. *Education Statistics Quarterly, 5*(1), 53–56.

Buttram, J .L. & Waters, J. T. (1997, September). Improving America's schools through standards-based education. *NASSP Bulletin, 81*(590), 1–6.

*Butty, J. L. M. (2001). Teacher instruction, student attitudes, and mathematics performance among 10th and 12th grade black and Hispanic students. *Journal of Negro Education, 70*(1–2), 19–37.

*Cain, J. S. (2002). An evaluation of the connected mathematics project. *Journal of Educational Research, 95*(4), 224–233.

*Caron, E. J. (2002). *Standards-in-practice: The classroom implications of Indiana standards-based reform.* Unpublished Ph.D., The University of Iowa.

*Carroll, W. M. (1997). Results of third-grade students in a reform curriculum on the Illinois state mathematics test. *Journal for Research in Mathematics Education, 28*(2), 237–242.

*Chance, P. L., & Anderson, R. B. (2003, April). *The principal's role in standards-based reform: Linking accountability to instructional improvement.* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED478258)

Chatterji, M. (2002). Models and methods for examining standards-based reforms and accountability initiatives: Have the tools of inquiry answered pressing questions on improving schools? *Review of Educational Research 72*(3), 345–386.

*Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J., & Li, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from interviews with educators in low-, medium-, and high-stakes states.* Chestnut Hill, MA: National Board on Educational Testing and Public Policy. (ERIC Document Reproduction Service No. ED474867)

Clune, W. H. (2001). Toward a theory of standards-based reform: The case of nine NSF statewide Systemic Initiatives. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 13–38). Chicago: University of Chicago Press.

Cohen, D. K. (1995). What is the system in systemic reform? *Educational Researcher, 24*(9), 11–17.

Conley, D. T. (1993). *Roadmap to restructuring.* Eugene, OR: ERIC Clearinghouse on Educational Management.

Cooney, B. (2001, April). *The widening gap of intolerance: An analysis of standards-based reforms and special education in secondary schools.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. (ERIC Document Reproduction Service No. ED456619)

Cooper, H. (1998). *Synthesizing research, third edition: A guide for literature reviews*. Thousand Oaks, CA: Sage Publications.

Cuban, L. (1989). Reforming again, again, and again. *Educational Researcher, 19*(1),3–13.

*Daniels, P. A. (1995, November). *A case study: The meaning and use of portfolios in the classroom of a more conservative (curriculum oriented) teacher.* Paper presented at the Annual meeting of the National Reading Conference, New Orleans, LA. (ERIC Document Reproduction Service No. ED395322)

Darling-Hammond, L. (2003). Standards and assessments: Where we are and what we need. *Teachers College Record* [online]. Retrieved July 11, 2005, from http://www.tcrecord.org/content.asp?contentid=11109

David, J. L., & Shields, P. M. (2001). *When theory hits reality: Standards-based reform in urban districts. Final narrative report.* Menlo Park, CA: SRI International. Retrieved December 14, 2004 from http://www.sri.com/policy/cep/pubs/pew/pewfinal.pdf

*Din, F. S. (1996, February). *The impact of Kentucky state testing on educational practices in Kentucky public schools.* Paper presented at the Annual meeting of the Eastern Educational Research Association, Cambridge, MA. (ERIC Document Reproduction Service No. ED405352)

*Doran, H. C. (2001). Evaluating the consequential aspect of validity on the Arizona instrument to measure standards. *Dissertation Abstracts International, 63*(01), 35A. (UMI No. AAT 3040140)

Dutro, E. (2002). Do state content standards make a difference? An illustration of the difficulties of addressing that pressing question. *Mid-Western Educational Researcher, 15*(4), 2–6.

Education Week. (2005). *Quality counts 2005. No small change: Targeting money toward student performance.* January 6, 2005.

Education Week. (2004). *Quality counts 2004: Count me in.* January 8, 2004.

Education Week. (2001). *Quality counts 2001. A better balance: Standards, tests, and the tools to succeed.* January 11, 2001

*Erickson, D. K., & Niess, M. L. (1996). Focusing on NCTM's standards: Teachers' choice and decisions related to students achievement in middle school mathematics. *Research in Middle Level Education Quarterly, 19*(2), 23–42. (ERIC Document Reproduction Service No. EJ524895)

*Falk, J., & Drayton, B. (2004). State testing and inquiry-based science: Are they complementary or competing reforms? *Journal of Educational Change, 5*(4), 345–387.

*Fernandez, A. Y. (2004). A qualitative study of selected fourth-grade teachers' perceptions of the sunshine state standards. *Dissertation Abstracts International, 65*(04), 1232A. (UMI No. AAT 3130419)

*Firestone, W. A., Camilli, G., Yurecko, M., Monfils, L., & Mayrowetz, D. (2000). *State standards, socio-fiscal context and opportunity to learn in New Jersey*. Retrieved November 28, 2004, from http://epaa.asu.edu/epaa/v8n35/

*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.

*Flexer, R. J., Cumbo, K., Borko, H., Mayfield, V., & Marion, S. F. (1995). *How "messing about" with performance assessment in mathematics affects what happens in classrooms (CSE technical report 396)*. Los Angeles, CA: UCLA Center for the Study of Evaluation.

Fraser, K. (1996). *What will it take? Standards-based education reform for all students.* Alexandria, VA: Center for Policy Research on the Impact of General and Special Education Programs. (ERIC Document Reproduction Service No. ED401695)

Frederiksen, J.R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27–32.

*Fuller, J. L. (2001, April). *An integrated hands-on inquiry based cooperative learning approach: The impact of the palms approach on student growth.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA. (ERIC Document Reproduction Service No. ED453176).

*Fuson, K. C., Carroll, W. C., & Drueck, J. V. (2000). Achievement results for second and third graders using the standards-based curriculum everyday mathematics. *Journal for Research in Mathematics Education, 31*, 277–295.

Gaddy, B.B., Dean, C.B., & Kendall, J.S. (2002). *Noteworthy perspectives: Keeping the focus on learning.* Aurora, CO: Mid-continent Research for Education and Learning.

*Ginsburg-Block, M. D., & Fantuzzo, J. W. (1998). An evaluation of the relative effectiveness of NCTM standards-based interventions for low-achieving urban elementary students. *Journal of Educational Psychology, 90*(3), 560–569.

*Goertz, M. E., Floden, R. E., & O'Day, J. (1996). *Systemic reform. Volume I: Findings and conclusions. Studies of education reform*. New Brunswick, NJ: Consortium for Policy Research in Education. (ERIC Document Reproduction Service No. ED397553)

*Goertz, M. E., Floden, R. E., & O'Day, J. (1995a). *Systemic reform. Volume II: Case studies. Studies of education reform*. New Brunswick, NJ: Consortium for Policy Research in Education. (ERIC Document Reproduction Service No. ED397554).

*Goertz, M. E., Floden, R. E., & O'Day, J. (1995b). *Systemic reform. Volume III: Technical appendix--research design and methodology. Studies of education reform*. New Brunswick, NJ: Consortium for Policy Research in Education. (ERIC Document Reproduction Service No. ED397555).

*Goldenberg, L. B., Ba, H., Heinze, J., & Hess, A. (2003). *Jason multimedia science curriculum's impact on student learning: Final evaluation report year three*. Retrieved May 12, 2005, from http://www2.edc.org/CCT/admin/publications/report/JASON_03FinalReport.pdf

*Grant, S. G. (2000, February). *Teachers and tests: Exploring teachers' perceptions of changes in the New York state-mandated testing program*. Retrieved March 23, 2005, from http://epaa.asu.edu/epaa/v8n14/

*Grant, S. G. (2001). An uncertain lever: Exploring the influence of state-level testing in New York State on teaching social studies. *Teachers College Record, 103*(3), 398–426.

*Grant, T. J., & Kline, K. (2000, April). *Understanding teachers' changing beliefs & practice while implementing a reform curriculum.* Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED440965)

*Grant, T. J., & Kline, K. (2001, October). *What impacts teachers as they implement a reform curriculum? The case of one fifth grade teacher.* Paper presented at the Annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Snowbird, UT. (ERIC Document Reproduction Service No. ED476635)

*Grisham, D. L., & Brink, B. (2003). Teachers reflect on the reform movement and its impact on their practice. *Journal of Reading Education, 28*(3), 15–25.

Haladyna, T.M., Nolen, S.B., & Haas, N.S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher, 20*(5) 2–7.

*Hamilton, L. S., McCaffrey, D. F., Stecher, B. M., Klein, S. P., Robyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis, 25*(1), 1–29.

*Hannafin, B. (2002). *Mashpee high school, Mashpee, MA.* Bloomington, MN: Plato Learning Inc. (ERIC Document Reproduction Service No. ED472043)

*Haug, C. (1998, April). *Variability in teachers' responses to standards-based mathematics reform.* Paper presented at the Annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED425067)

*Hickey, D. T., Moore, A. L., & Pellegrino, J. W. (2001). The motivational and academic consequences of elementary mathematics environments: Do constructivist innovations and reforms make a difference? *American Educational Research Journal, 38*(3), 611–652.

Hill, D. (2003). Workshop summary. In K. S. Hollweg & D. Hill (Eds.), *What is the influence of the National Science Education Standards? Reviewing the evidence, a workshop summary* (pp.3–20). Washington, DC: The National Academies Press.

Hollweg, K. S., & Hill, D. (2003). *What is the influence of the National Science Education Standards? Reviewing the evidence, a workshop summary.* Washington, DC: The National Academies Press.

*Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Santong, J., & Fey, J. T. (2000). Effects of standards-based mathematics education: A study of the core-plus mathematics project algebra and functions strand. *Journal for Research in Mathematics Education, 31*(3), 328–361.

*Indiana University, Center for Innovation in Assessment. (1998). *A study of the instructional effectiveness of the signatures program. Signatures technical report #1.* Bloomington; IN: Harcourt Brace. (ERIC Document Reproduction Service No. ED431172)

Isaacs, A., Carroll, W., & Bell, M. (2001). *A research-based curriculum: The research basis for the UCSMP Everyday Mathematics Curriculum.* Retrieved on June16, 2005, from http://everydaymath.uchicago.edu/educators/Revised_Research_Base_5.pdf

*Ivey, K. M. C. (1996). Mechanistic and organic: Dual cultures in a mathematics classroom. *Journal of Research & Development in Education, 29*(3), 141–151.

*Janson, D. C. (2002). *Assessment practices of third and fifth grade science teachers: A comparison to the style/format, process, and content of Ohio's proficiency tests.* Unpublished doctoral dissertation, Ohio Wesleyan University, Delaware, OH. (ERIC Document Reproduction Service No. ED478989)

*Jerome, K., & Gilman, D. A. (2003). *Writing improvement programs: Does this type of intervention really work?* Terre Haute, IN: Indiana State University. (ERIC Document Reproduction Service No. ED474935)

Joftus, S., & Berman, I. (1998). *Great expectations? Defining and assessing rigor in state standards for mathematics and English language arts.* Washington, DC: Council for Basic Education. (ERIC Document Reproduction Service No. ED416080)

*Johns, C. M. (2004). Success or survival: Beginning elementary teachers' experience with the mathematics curriculum. *Dissertation Abstracts International, 65*(05), 1668A. (UMI No. AAT 3134484)

Johnson, J., & Duffett, A. (2003). *Where we are now: 12 things you need to know about public opinion and public schools. A digest of a decade of survey research.* New York: Public Agenda Foundation. (ERIC Document Reproduction Service No. ED477867)

Kahle, J. B., & Kelly, M. K. (2001). Science teacher professional development: A researcher's perspective. In J. Rhoton & P. Bowers (Eds.) *Professional development leadership and the diverse learning: Issues in science education* (pp. 101–113). Arlington, VA: National Science Education Leadership Association and National Science Teachers Association.

Kendall, J. S. (2001). *A technical guide for revising or developing standards and benchmarks.* Aurora, CO: Mid-continent Research for Education and Learning. Retrieved June 30, 2005, from http://www.mcrel.org/PDF/Standards/5011TG_TechnicalGuide.pdf

Kendall, J. S., & Marzano, R. J. (2000). *Content knowledge: A compendium of standards and benchmarks for K–12 education* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

*Kerr, R. D. (1999). Implementing constructivism to improve the mathematics achievement of inner city third-grade students. *Dissertation Abstracts International, (59)*12, 4351A. (UMI No. AAT 9913906)

*Khattri, N., Reeve, A. L., Kane, M. B., & Adamson, R. J. (1995). *Assessment of student performance. Volume I: Findings and conclusions. Studies of education reform.* Washington, DC: Pelavin Research Institution. (ERIC Document Reproduction Service No. ED397532)

Kiplinger, V. L., & Linn, R. L. (1993). *Raising the stakes of test administration: The impact on student performance on NAEP.* CSE Technical Report 360. Los Angeles, CA: National Center for Research on Evaluation, Standards, and student testing.

*Klein, S., Hamilton, L., McCaffrey, D., Stecher, B., Robyn, A., & Burroughs, D. (2000). *Teaching practices and student achievement: Report of first-year findings from the 'mosaic' study of systemic initiatives in mathematics and science.* Santa Monica, CA: Rand Corporation. (ERIC Document Reproduction Service No. ED445914)

Koretz, D. M., McCaffrey, D. F., & Hamilton, L. S. (2001). *Toward a framework for validating gains under high-stakes conditions.* CSE Technical Report 551. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

*Koretz, D., Mitchell, K., Barron, S., & Keith, S. (1996). *Final report: Perceived effects of the Maryland school performance assessment program*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

*Koretz, D. M., Barron, S., Mitchell, K., & Stecher, B. (1996). *Perceived effects of the Kentucky instructional results information system (KIRIS)*. Santa Monica, CA: RAND. (ERIC Document Reproduction Service No. ED405365)

Lachat, M.A. (1999). *Standards, equity, and cultural diversity.* Providence, RI: LAB at Brown University.

*Lane, S., Silver, E. A., & Wang, N. (1995, April). *An examination of the performance gains of culturally and linguistically diverse students on a mathematics performance assessment within the quasar project.* Paper presented at the Annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED390927)

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. A., Snow, D., & Martin-Glenn, M. (2003). *The effectiveness of out-of-school-time strategies in assisting low-achieving students in reading and mathematics: A research synthesis*. Aurora, CO: Mid-continent Research for Education and Learning. Retrieved June 30, 2005, from http://www.mcrel.org/topics/productDetail.asp?productID=151

*Lawrenz, F., Huffman, D., & Lavoie, B. (2001). *Evaluating the long term effects of teacher enhancement*. Minneapolis, MN: University of Minnesota. (ERIC Document Reproduction Service No. ED455101)

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 20*(2), 4–16.

*Lee, J. (1998). The impact of content-driven state education reform on instruction. *Research in Middle Level Education Quarterly, 21*(4), 15–26.

*Louisville University, School of Education. (1995). *The implementation of performance assessment in Kentucky classrooms.* Frankfort, KY: Kentucky Institution for Education Research. (ERIC Document Reproduction Service No. ED394978)

*Lubienski, S. T. (2002, April). *Are we achieving "mathematical power for all?" a decade of national data on instruction and achievement.* Paper presented at the Annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No ED463166)

Manouchehri, A., & Goodman, T. (1998). Mathematics curriculum reform and teachers: Understanding the connections. *Journal of Educational Research, 92*(1), 27–41.

*Marx, R. W., Blumenfeld, P. C., Krajcik, J., Fishman, B. J., Soloway, E., Geier, R., et al. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Teaching, 41*(10), 1063–1080.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us.* Aurora, CO: Mid-continent Research for Education and Learning. Retrieved June 30, 2005, from http://www.mcrel.org/PDF/SchoolImprovementReform/5002RR_NewEraSchoolReform.pdf

Massell, D., Kirst, M., & Hoppe, M. (1997). *Persistence and change: Standards-based systemic reform in nine states.* Philadelphia: Consortium for Policy Research in Education.

*Mayer, D. P. (1997, March). *New teaching standards and old tests: Dangerous mismatch?* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED413348)

*Mayer, D. P. (1998). Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis, 20*(2), 53–73.

*Mayrowetz, D. (1999, October). *The interaction of standards-based reform and inclusion in New Jersey's fourth grade classrooms: Are they perfect together?* Paper presented at the Annual Conference of the University Council for Educational Administration, Minneapolis, MN. (ERIC Document Reproduction Service No. ED439804)

*McCaffrey, D. F., Hamilton, L. S., Stecher, B. M., Klein, S. P., Bugliari, D., & Robyn, A. (2001). Interactions among instructional practices, curriculum, and student achievement: The case of standards-based high school mathematics. *Journal for Research in Mathematics Education, 32*(5), 493–517.

McClure, P. (2005). Where standards come from. *Theory Into Practice, 44*(1), 4–10.

*McDonnell, L. M., & Choisser, C. (1997). *Testing and teaching: Local implementation of new state assessments. CSE technical report 442*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the Center for the Study of Evaluation (CSE).

*McGinnis, J. R., & Parker, C. (2000, April). *What happens to first year teachers prepared to make connections between science and mathematics when they enter the workplace?* Paper presented at the Annual meeting of the National Association for Research in Science Teaching, New Orleans, LA. (ERIC Document Reproduction Service No. ED442674)

*McLaughlin, M. J. (2000). *Reform for every learner: Teachers' views on standards and students with disabilities*. Alexandria, VA: Center for Policy Research on the Impact of General Special Education Reform. (ERIC Document Reproduction Service No. ED444302)

McLaughlin, M. W., & Shepard, L.A. (with O'Day, J. A.). (1995). *Improving education through standards-based reform*. Stanford, CA: National Academy of Education. (ERIC Document Reproduction Service No. ED387867)

*McMillan, J. H., Myran, S., & Workman, D. (1999, April). *The impact of mandated statewide testing on teachers' classroom assessment and instructional practices.* Paper presented at the Annual meeting of the American Educational Research Association., Montreal, Quebec, Canada. (ERIC Document Reproduction Service No. ED431041)

*Michlin, M., Seppanen, P., & Sheldon, T. (2001). *Charting a new course: A study of the adoption and implementation of standards-based mathematics curricula in eight Minnesota school districts. Final report to scimath[mn], fall 2001.* Minneapolis, MN: University of Minnesota. (ERIC Document Reproduction Service No. ED463321)

Mid-continent Research for Education and Learning. (2000). *Noteworthy perspectives on implementing standards-based education.* Aurora, CO: Author. Retrieved June 30, 2005, from http://www.mcrel.org/PDF/Noteworthy/5993IR_NWImplementingStandards.pdF

*Moon, T. R., Brighton, C. M., & Callahan, C. M. (2002). State standardized testing programs: Friend or foe of gifted education? *Roeper Review, 25*(2), 49–60.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. Washington, DC: Government Printing Office.

National Council on Education Standards and Testing (1992, January). *Raising standards for American education*. Washington, DC: U.S. Government Printing Office.

National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics.* Reston, VA: National Council of Teachers of Mathematics

National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.

National Education Goals Panel. (1991). The National Education Goals Report: Building a nation of learners. Washington, DC: Author.

National Research Council. (2002). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education.* I. R. Weiss, M. S. Knapp, K. S. Hollweg, & G. Burrill (Eds.), Committee on Understanding the Influence of Standards in K–12 Science, Mathematics and Technology Education, Center for Education, Division of Behavioral and Social Sciences Education. Washington, DC: National Academy Press.

National Research Council. (1996). *National science education standards.* Washington, DC: National Academy Press.

National Science Education Standards. (2005). National Science Education Standards. Retrieved on June 16, 2005, from http://www.nap.edu/readingroom/books/nses/

Nave, B., Miech, E., & Mosteller, F. (2000). A lapse in standards: Linking standards-based reform with student achievement. *Phi Delta Kappan, 82*(2), 128–132.

North Central Regional Educational Laboratory. (2000). Critical issue: Integrating standards into the curriculum. Retrieved June 16, 2005, from http://www.ncrel.org/sdrs/areas/issues/content/currclum/cu300.htm

*Norman, J., Stein, M., Moussiaux, S., & Clay-Chambers, J. (1998, April). *The effect of the Detroit urban systemic initiative on perceived instructional practice and curriculum adequacy.* Paper presented at the Annual meeting for the National Association for Research in Science Teaching, San Diego, CA. (ERIC Document Reproduction Service No. ED417979)

*Parker, V., & Gerber, B. (2000). Effects of a science intervention program on middle-grade student achievement and attitudes. *School Science and Mathematics, 100*(5), 236–242.

*Pedroza, A. (1998, April). *Bordering on success: Mexican American students and high stakes testing.* Paper presented at the Annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED420713)

Pearsall, M. K. (Ed). (1993). *Scope, sequence, and coordination of secondary school science. Vol. 1. The content core: A guide for curriculum designers.* Washington, DC: National Science Teachers Association.

Pellegrino, J. W. (2004). *The evolution of educational assessment: Considering the past and imagining the future.* Princeton, NJ: Educational Testing Service.

Puma, M., Raphael, J., Olson, K., & Hannaway, J. (2000). *Putting standards to the test: A design for evaluating the systemic reform of education.* Washington, DC: Urban Institute. Retrieved February 3, 2005, from http://www.urban.org/Uploadedpdf/410385_Standards.pdf

Raber, S., Roach, V., & Fraser, K. (Eds.). (1998). *The push and pull of standards-based reform. How does it affect local school districts and students with disabilities?* Alexandria, VA: The Center for Policy Research on the Impact of General and Special Education Reform. (ERIC Document Reproduction Service No. ED417531)

Ravitch, D. (1995). *National standards in American education: A citizen's guide.* Washington, DC: Brookings Institution.

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 instruction* (pp. 37–75). Boston: Kluwer Publishing.

Resnick, L. B., & Zurawsky, C. (2005). Getting back on course: Standards-based reform and accountability. *American Educator* [online]. Retrieved May 31, 2005, from http://www.aft.org/pubs-reports/american_educator/issues/spring05/resnick.htm

*Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education, 34*(1), 74–95.

*Ridgway, J. E., Zawojewski, J. S., Hoover, M. N., & Lambdin, D. V. (2003). Student attainment in the connected mathematics curriculum. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? Do they work?* (pp. 193–224). Mahwah, NJ: Lawrence Erlbaum Associates.

*Riordan, J. E., & Noyce, P. E. (2001). The impact of two standards-based mathematics curricula on student achievement in Massachusetts. *Journal for Research in Mathematics Education, 32,* 368–398.

*Rivet, A. E., & Krajcik, J. (2004). Achieving standards in urban systemic reform: An example of a sixth grade project-based science curriculum. *Journal of Research in Science Teaching, 41*(7), 669–692.

*Rizor, D. M. (2000). Wyoming teacher survey of instruction in elementary mathematics. *Dissertation Abstracts International, 61*(06), 2191A. (UMI No. AAT 9973572)

*Roderick, M., Jacob, B. A., & Bryk, A. S. (2002). The impact of high-stakes testing in Chicago on student achievement in promotional gate grades. *Educational Evaluation & Policy Analysis, 24*(4), 333–357.

Ross, J. A., McDougall, D., & Hogaboam-Gray, A. (2002). Research on reform in mathematics education, 1993–2000. *Alberta Journal of Educational Research, 48*(2), 122–138.

*Sandholtz, J. H., Ogawa, R. T., & Scribner, S. P. (2004). Standards gaps: Unintended consequences of local standards-based reform. *Teachers College Record, 106*(6), 1177–1202.

*Saxe, G. B., Gearhart, M., & Seltzer, M. (1999). Relations between classroom practices and student learning in the domain of fractions. *Cognition and Instruction, 17*(1), 1–24.

*Schneider, R. M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching, 42*(3), 283–312.

*Schneider, R. M., Krajcik, J., Marx, R. W., & Soloway, E. (2002). Performance of students in project-based science classrooms on a national measure of science achievement. *Journal of Research in Science Teaching, 39*(5), 410–422.

*Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education, 34*(3), 228–259.

Schoen, H. L., Fey, J. T., Hirsch, C. R. & Coxford, A. F. (1999). Issues and options in the math wars. *Phi Delta Kappan, 80*(6), 444–453.

*Schoen, H. L., & Hirsch, C. R. (2003). The core-plus mathematics project: Perspectives and student achievement. In S. L. Senk & D. R. Thompson (Eds.), *Standards-based school mathematics curricula: What are they? What do student learn?* (pp. 311–343). Mahwah, NJ: Lawrence Erlbaum Associates.

Schoenfeld, A.H. (2002). Making mathematics work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*, 13–25.

*Schorr, R. Y., Firestone, W. A., & Monfils, L. (2003). State testing and mathematics teaching in new jersey: The effects of a test without other supports. *Journal for Research in Mathematics Education, 34*(5), 373–405.

*Schulte, A. C., Villwock, D. N., Whichard, S. M., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review, 30*(4), 487–506.

Shanahan, T. (2000). Research synthesis: Making sense of the accumulation of knowledge in reading. In M. L. Kamil, P .B. Mosenthal, P. D. Pearson, and R. Barr (Eds.), *Handbook of reading research, volume III* (pp. 209–226). Mahwah, NJ: Lawrence Erlbaum.

Shepard, L. A., & Dougherty, K. C. (1991, April). Effects of high-stakes testing on instruction. Paper presented at the annual joint meetings of the American Educational Research Association and the National Council on Measurement in Education, Chicago, IL.

*Shymansky, J. A., Yore, L. D., & Anderson, J. O. (2000, April). *A study of changes in students' science attitudes, awareness and achievement across three years as a function of the level of implementation of interactive-constructivist teaching strategies promoted in a local systemic reform effort.* Paper presented at the Annual meeting of the National Association for Research in Science Teaching, New Orleans, LA. (ERIC Document Reproduction Service No. ED439954)

*Silver, E. A., & Lane, S. (1995). Can instructional reform in urban middle schools help students narrow the mathematics performance gap? Some evidence from the QUASAR project. *Research in Middle Level Education Quarterly, 18*(2), 49–70.

*Smith, A. M. (2003, April). *Negotiating control and protecting the private: History teachers and the Virginia standards of learning.* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED475360)

*Smith, M. L. (1997, March). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program (ASAP): Equity and teacher capacity building. CSE technical report 425*. Retrieved January 21, 2005, from http://www.cse.ucla.edu/CRESST/Reports/TECH425.PDF

*Smith, M. S. (2000). Balancing old and new: An experienced middle school teacher's learning in the context of mathematics instructional reform. *Elementary School Journal, 100*(4), 351–375.

Smith, M.S., & O'Day, J.A. (1990). Systemic school reform. In S. Fuhrman & B. Malen (eds.), *Politics of Curriculum and Testing: Politics of Education Association Yearbook* (pp. 233–267). *London: Taylor and Francis Limited.*

*Snow-Renner, R. L. (2000). The technical feasibility of using large-scale measures of opportunity-to-learn to inform policy and practice: Illustrations using the Colorado TIMMS. *Dissertation Abstracts International, 61*(08), 3017A. (UMI No. AAT 9983963)

Snow-Renner, R. (2001). *Teachers' perspectives on standards-based education: Initial findings from a high-performing, high-needs school district*. Aurora, CO: Mid-Continent Research for Education and Learning.

*Speas, C. (2003). *Project Achieve evaluation report: Year one, 2001-2002. E & R Report no. 02.35*. Raleigh, NC: Wake County Public School System. (ERIC Document Reproduction Service No. ED477126).

*Speas, C. (2004, January). *Project Achieve evaluation: Year two, 2002-03. E & R report no. 03.16*. Retrieved March 21, 2005, from http://www.wcpss.net/evaluation-research/reports/2003/0316year2_achieve.pdf

*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.

*Stecher, B., & Chun, T. (2001). *School and classroom practices during two years of education reform in Washington state. CSE technical report 550.* Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing. (ERIC Document Reproduction Service No. ED459194)

*Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing. CSE technical report no. 482*. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing. (ERIC Document Reproduction Service No. ED426070)

*Stecher, B. M., & Barron, S. I. (1999). *Quadrennial milepost accountability testing in Kentucky. CSE technical report 505*. Los Angeles, CA: Center for Research on Evaluation, Standards Student Testing. (ERIC Document Reproduction Service No. ED440149)

Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125–138). New York: Russell Sage Foundation.

*Stone, C. A., & Lane, S. (2000, April). *The relationship between changes in MSPAP school performance over time and teacher, student, and school factors.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA. (ERIC Document Reproduction service No. ED446152)

*Swanson, C. B., & Stevenson, D. L. (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation & Policy Analysis, 24*(1), 1–27.

*Swierzbin, B., Liu, K., & Thurlow, M. (2000). *Initial perceptions of English as a Second Language educators on including students with limited English proficiency in Minnesota's high standards. State assessment series, Minnesota report 33.* St. Paul, MN: Minnesota State Department of Children, Families, and Learning. (ERIC Document Reproduction Service No. ED457691)

*Taylor, G., Shepard, L., Kinner, F., & Rosenthal, J. (2002). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost. CSE technical report 588*. Los Angeles, CA: Center for Research on Evaluation, Standards Student Testing. (ERIC Document Reproduction Service No. ED475139)

Thompson, S. (2001, January). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 358–362.

Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001*. Minneapolis, MN: National Center on Educational Outcomes. (ERIC Document Reproduction Service No. ED475047)

*Thompson, C. L., Zeuli, J. S., & Borman, J. (1997, March). *The science reforms: Mapping the progress of reform and multiple contexts of influence.* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, IL (ERIC Document Reproduction Service No. ED409168)

*Turner, H. M., III. (1998, April). *Mathematics teaching in the 1990's: Are pacesetter mathematics teachers teaching from the standards?* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED428044)

*U. S. Department of Education. (2001a). *The longitudinal evaluation of school change and performance (LESCP) in Title I schools: Final report. Volume 1: Executive summary*. Retrieved May 19, 2005, from http://www.ed.gov/offices/OUS/PES/esed/lescp_vol1.pdf

*U. S. Department of Education. (2001b). *The longitudinal evaluation of school change and performance (LESCP) in Title I schools. Final report. Volume 2: Technical report.* Retrieved May 19, 2005, from http://www.ed.gov/offices/OUS/PES/esed/lescp_vol2.pdf

Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literacy standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research. Volume III* (pp. 909–935). Mahwah, NJ: Lawrence Erlbaum Associates.

*VanSledright, B. A. (2002). Fifth graders investigating history in the classroom: Results from a researcher-practitioner design experiment. *Elementary School Journal, 103*(2), 131–160.

*Watson, M. L. (1996). The impact of a new program for teaching algebra to increase achievement and persistence in mathematics. *Dissertation Abstracts International, 57*(09), 3864A. (UMI No. AAT 9703817)

Wayne, A. J., & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73*(1), 89–122.

Weiss, I. R. (1994). *A profile of mathematics and science education in the United States.* Chapel Hill, NC: Horizon Research.

*Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. (2003, May). *Looking inside the classroom: A study of K–12 mathematics and science education in the United States*. Retrieved December 22, 2004, from http://www.horizon-research.com/insidetheclassroom/reports/looking/complete.pdf

Wheelock, A. (1995). *Standards-based reform: What does it mean for the middle grades?* New York: Edna McConnell Clark Foundation.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001, February). *Teacher preparation research: Current knowledge, gaps and recommendations* (Document R-01-3). Seattle, WA: Center for the Study of Teaching and Policy.

*Wong, K. K., Anagnostopoulos, D., Rutledge, S., & Edwards, C. (2001). *The challenge of improving instruction in urban high schools: Case studies of the implementation of the Chicago academic standards.* Chicago, IL: University of Chicago. (ERIC Document Reproduction Service No. ED460186)

*Woodward, J., & Baxter, J. (1997). The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings. *Exceptional Children, 63*(3), 373–388.

*Woodward, J., Monroe, K., & Baxter, J. (2001). Enhancing student achievement on performance assessments in mathematics. *Learning Disabilities Quarterly, 24*(1), 33–46.

Woodward, J., & Montague, M. (2002). Meeting the challenge of mathematics reform for students with learning disabilities. *Journal of Special Education, 36*(2), 89–101.

Zucker, A. A., Shields, P. M., Adelman, N. E., Corcoran, T. B., & Goertz, M. E. (1998). *A report on the evaluation of the National Science Foundation's Statewide Systemic Initiative (SSI) program.* Menlo Park, CA: SRI International. Retrieved February 14, 2005 from http://www.sri.com/policy/cep/pubs/ssi/ssifinal.pdf

APPENDIX A
WEBSITE AND TABLE OF CONTENT SEARCHES

Websites Searched	Tables of Content Searched
Achieve	American Educational Research Journal
American Federation of Teachers	American Journal of Education
Center on Education Policy	Education and Urban Society
Center on Reinventing Public Education	Educational Administration Quarterly
Center for Research on Evaluation, Standards, and Student Testing	Educational Evaluation and Policy Analysis
Council of Chief State School Officers	Education Next
Consortium for Policy Research in Education	Education Policy Analysis Archives
Education Commission of the States	Education Statistics Quarterly
Education Trust	Educational Assessment
Horizon Research, Inc.	Educational Policy
Mid-continent Research for Education and Learning	Educational Researcher
National Education Association	The Elementary School Journal
Policy Studies Associates, Inc.	Equity and Excellence in Education
RAND Corporation	Harvard Educational Review
Southern Regional Educational Board	Journal for Research in Mathematics Education
Systemic Research, Inc.	Journal of Curriculum and Supervision
SRI International	Journal of Educational Research
Thomas B. Fordham Foundation	Journal of Education for Students Placed At Risk
Urban Institute	Journal of Research in Rural Education
What Works Clearinghouse	Journal of Research in Science Teaching
	NAASP Bulletin
	Reading Research Quarterly
	Research in Middle Level Education
	Review of Educational Research
	Teachers College Record

APPENDIX B
CODING INSTRUMENT

McREL 2005 Research Synthesis
The Influence of Standards on K-12 Teaching and Student Learning

Primary SB input
(circle one)

Curriculum
Instruction
Assessment

Consult the 2005 Research Synthesis Coding Manual for instructions and definitions of coding terms.
For direct quotes, use quotation marks and cite page umbers.
For ERIC documents, cite the ERIC page number, not the original page number.

INITIAL CODER name and date: _____

FINAL CODER name and date: _____

1. STUDY SUMMARY *(cite page numbers where appropriate)*

1.01 Author(s): _____

1.02 Year: _____

1.03 Title (*or first several words*): _____

_____

1.04 Source: ☐Journal ☐ Dissertation ☐ Conference presentation ☐Technical report

☐ Book or book chapter

1.05 Are there multiple reports related to this study that have been screened for the synthesis? ☐ Yes ☐ No

If yes, name the author(s) and year(s) of the reports: _____

_____

1.06 Additional studies *(reference chasing; cite page no.)*:_____

_____

1.07 Additional information in the study that informs the synthesis *(cite page no.)*: _____

_____

1.08 ☐ **EXCLUDE STUDY - SCREENING LEVEL**	
a) ☐ Lacks empirical data.	b) ☐ Reported before 1991
c) ☐ Doesn't involve K–12 students	d) ☐ Not conducted in U.S. or its territories
e) ☐ No assessment/documentation of student or teacher outcomes	f) ☐ Not related to K–12 standards (see 2.08)
g) ☐ Not in core subject area (see 2.07)	h) ☐ Does not address research questions

1.09 ☐ **EXCLUDE STUDY - CODING LEVEL**	
a) ☐ Qualitative study lacking criteria for rigor	b) ☐ Quantitative study lacking criteria for rigor
c) ☐ Mixed methodology study lacking criteria for rigor	d) ☐ Study of CSR
e) ☐ Alternate version of the same report	f) ☐ Summary of other included reports

For a, b, c, complete the following:

1.09a/c **Qualitative study lacking the following criteria for rigor** (*check those that the study does **not** include*)

- ☐ The study sufficiently describes the processes and methods used for data collection, such as a description of an interview process and protocol.
- ☐ The study sufficiently describes the processes used for data analysis, such as coding methods.
- ☐ The study includes sufficient evidence of the results, such as detailed descriptions of events or observations and/or samples of responses from interviewees.
- ☐ The study describes one or more processes used to validate the results, such as multiple sources of evidence, a search for disconfirming evidence, generation of rival hypotheses or explanations, negative case analysis, member checking.

1.09b/c **Quantitative study lacking the following criteria for rigor** (*check those that the study does **not** include*)

- ☐ The study describes instrument development, such as where the items were obtained or, if items were developed, there was a pilot test of the items and their validity or an expert check.
- ☐ For results based on a survey, the response rate was reported or could be obtained.
- ☐ The study sufficiently describes the processes used for data collection and analysis, including sample size.
- ☐ The study includes sufficient evidence of the results such as tables of means or frequencies and descriptive statistics, as opposed to summarizing findings without details.

NOTE: No need to continue coding excluded studies.

2. METHODOLOGY *(cite page numbers where appropriate)*

2.01 Type of research and method *(check only one and describe the method or design if applicable, e.g., pretest-posttest survey design, qualitative case study design with interviews and document review)*:

☐ Quantitative experimental _____

☐ Quantitative quasi-experimental _____

☐ Quantitative non-experimental _____

☐ Qualitative _____

☐ Mixed methods _____

_____

2.02 Study Type: ☐ Research study ☐ Evaluation study ☐ Not indicated

2.02a If an evaluation, was the evaluator external to the program? ☐ Yes ☐ No ☐ Not indicated

2.03 Longitudinal study? ☐ Yes ☐ No 2.03a If yes, duration of study: _____

2.03b If yes, number and frequency of measurements: _____

2.04 Primary standards-based (SB) input variable(s) addressed by the study *(check only one)*:

☐ SB curriculum ☐ SB instructional guidance ☐ SB assessment

Define/describe the primary input variable (*include duration if applicable, e.g. a 5-week curriculum, 9 monthly professional development meetings; cite page no.*:_____

_____

_____

_____

_____

2.04a Other (secondary) SB input variables addressed by the study *(check all that apply)*:

☐ SB curriculum ☐ SB instructional guidance ☐ SB assessment

Define/describe the other input variable(s) (*include duration if applicable; cite page no.):*

_____

_____

_____

_____

2.05 Outcome(s) studied and how measured (*check all that apply and describe how measured):*

2.05a Student achievement ☐ Yes ☐ No If yes, describe measure:

☐ standardized test ☐ state assessment ☐ curriculum-developed assessment

☐ teacher developed classroom assessment ☐ researcher (study author) developed assessment

☐ grades in subject areas ☐ teacher reports (qualitative)

☐ observations (qualitative) ☐ Other *(describe):* _____

2.05b Teacher instruction ☐ Yes ☐ No If yes, describe measure:

☐ previously validated teacher survey ☐ researcher developed teacher survey

☐ principal surveys ☐ teacher interviews ☐ principal interviews ☐ teacher logs or journals

☐ classroom observations ☐ Other *(describe)*: _____

2.06 What analytic methods are used to measure the link between the SB input variable(s) and the outcome(s)?

☐ in-depth description (qualitative) ☐ counts/frequencies (qualitative)

☐ counts/frequencies (quantitative) ☐ descriptive statistics (quantitative) ☐ simple correlations

☐ inferential statistical comparisons ☐ multiple regression

☐ Other *(describe):* _____

2.07 Does the study compare different influences on teaching or student learning? ☐ Yes ☐ No

If yes, what is being compared? ☐ Different aspects of the input variable

☐ Different teacher characteristics ☐ Different school characteristics or contexts

☐ Other *(describe):* _____

2.08 Subject area: ☐Reading/Language Arts ☐ Math ☐ Science ☐ Social Studies ☐ Missing

☐ Other or general (*explain):* _____

2.09 Level of standards *(check all that apply):* ☐ school ☐ district ☐state *(name)* _____

☐ national *(name)* _____

2.10 Does the study include systemic standards-based implementation(s) or influences? ○ Yes ○ No

If yes, check all that apply: ☐accountability policies ☐ administrator leadership

☐administrator support for teachers ☐college entrance policies ☐educator expectations for students

☐ governance ☐ learning environment ☐personnel ☐resources ☐ school-community relationships

☐teacher certification policies ☐ teacher education ☐teacher leadership ☐technology ☐time

☐working relationships ☐ Other *(describe)*: _____

2.11 Were there indications of standards-based alignment? ☐ Yes ☐ No

2.10a If yes, what was aligned with what? _____

_____

2.12 Other characteristics of methodology, including factors that influence research quality:

_____

_____

3. STUDY POPULATION/SAMPLE *(cite page numbers where appropriate)*

3.01 School/district locale: ☐ Urban ☐ Suburban ☐ Rural ☐ Missing

3.02 At-risk population or sample? ☐ Yes ☐ No If yes, describe at-risk indicators.

At-risk indicators of *school:* ☐ Percent FRL _____% ☐ Percent minority _____%

☐ Other *(describe)*: _____

At-risk indicators of *district*: ☐ Percent FRL _____% ☐ Percent minority _____%

☐ Other *(describe)*: _____

Addl. information re at-risk:_____

_____

3.03 Sample size(s) and description: _____

_____

_____

_____

3.04 Sample selection and/or assignment process: _____

_____

_____

3.05 Survey response rate(s): _____ ☐ Not applicable

3.06 Other attrition indicators: _____

_____

_____

4. RESULTS *(cite page numbers where appropriate)*

4.01 Was an effect size reported? ☐ Yes, on page(s): _____ ☐ No

4.02 Could this study be used for a meta-analysis? ☐ Yes, see data on page(s): _____ ☐ No

4.03 Summary description of relevant findings: *(Describe the results — both positive and negative — that pertain to the influence of the standards-base input variable(s) on teaching and student learning. For direct quotes, use quotation marks and cite page umbers.)*

Primary standards-based input variable: _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Other (secondary) standards-based input variable(s):

_____

_____

_____

_____

_____

_____

_____

Other important results: _____

_____

_____

_____

_____

_____

4.04 Could this study inform a research-to-practice publication? ☐ Yes ☐ No

4.04a If yes, in what way? *(cite page numbers)*

☐ Careful description of standards-based implementation: _____ ☐ Vignette(s): _____

☐ Dialogue transcript: _____ ☐ Other: _____

4.05 Implications for education policies and practices: *(Distinguish between those that the author describes and those that can be validly inferred from the results)* _____

_____

_____

APPENDIX C
REVIEWED STUDIES OF STANDARDS-BASED CURRICULUM (CHAPTER 3)

Study Source Standards	Standards-based (SB) Variable(s) Methods Relevant to Research Synthesis Questions Measures and Sample	Findings
Adams, Brower, Hill, & Marshall (2000) Technical report National standards: NCTM, NSES State standards: TX	SB Curriculum: Texas state standards and national standards in math and science Study identified reform issues that affect student achievement in math and science by examining math and science standards and reforms, teaching instruction, and professional development. Method: Quantitative non-experimental Measures: A teacher survey was administered to measure teachers' knowledge of state and national standards, teaching practices, and professional development Sample: A stratified random sample of 100 schools from four Texas regions (1, 2, 3, and 20) was identified. This included approximately 350 math and science teachers in grades 4 through 8. 177 surveys were completed.	Overall, survey results indicated that teachers rated instructional strategies such as hands-on activities, problem solving activities, and technology integration as important in creating an ideal classroom but when asked what strategies they used during the week teachers were using more traditional strategies. Teachers frequently reported that hands-on activities and technology were the most important factors in an ideal classroom. However, teachers reported using more seatwork and seldom reported using math manipulatives, calculators, and computers in their classrooms. Results analyzed by subject showed that math teachers applied what they knew about effective practices in their classrooms more than science teachers. Problem solving was used frequently in math classes while science teachers reported using more seatwork. Both math and science teachers reported using technology the least in their classrooms.
Anderson (1995) Vol 2 (1995) Vol 3 (1996) Vol 1 Technical reports National standards: NCTM AAAS NSTA NSES	SB Curriculum: National math and science standards Study examined how schools engaged in curriculum reform using case studies with cross case analysis. (Also included references to assessment influences). Method: Qualitative Measures: Classroom observations, interviews with teachers, principals, students, and parents, document review Sample: Schools were selected based on their success in implementing reforms 3 middle schools	Examined effects of curricular changes in 9 schools over a 4 year period of time. Four of the 9 case study schools specifically addressed national standards as the framework for the new curriculum. In the cross case analysis, the researchers found the following: As a result of the new curriculum, the teacher's role in the classroom had changed from being a dispenser of knowledge to being a coach and a facilitator. Teachers must learn how to effectively asses their students' learning. The students can no longer be passive receivers of information but must become self-directed learners. Student work had changed from teacher-prescribed

	6 high schools *Note: detailed description of site selection in Vol. 3)*	activities to student directed learning. Outcomes for the students were greater student engagement, development of thinking skills in an embedded, applicable context, new student learning roles.
Ba, Admon, & Anderson (2002) Technical report Year 2 National standards: NSES	SB Curriculum: *JASON Project*- multimedia science curriculum for middle grades. Second year evaluation of a SB multimedia science curriculum focused on its use. Method: Quantitative non-experimental Measures: An initial survey and a follow-up survey on teacher backgrounds, use of *JASON* curriculum, teaching practices, and experience with *JASON.* Samples: Initial survey: Approximately 25,000 *JASON* teachers were surveys. 849 online surveys and 1,047 paper surveys were completed (8% response rate). Teachers had an average of 14.5 years of teaching experience. 44.5% had used *JASON* for 2 to 4 years. 53% taught elementary and 37% taught middle school. Follow-up survey: 1,133 of the above teachers.	Over nine out of ten of teachers used *JASON's* print curriculum and said that it helped them teach science in an exciting way. Other curriculum components were used by 7 to 8 out of ten teachers. Cramer's V showed a significant relationship between the percentage of honor students and teachers' greater use of the curriculum components. Nine out of ten teachers said the curriculum helped them to meet national/local standards. In the follow up survey: around three quarters of teachers reported that they had their students revise their work and used group research activities.
Ba, Martin, & Diaz (2001) Technical report Year 1 National standards: NSES	SB Curriculum: *JASON Project*- multimedia science curriculum for middle grades. First year evaluation of the impact of a SB science curriculum on student achievement and instruction. Method: Mixed methods(quasi-experimental) Measures: Pre/post inquiry test, videotaped student presentations, classroom observations, principal and teacher interviews, and teacher and student surveys. Sample: Schools: 8 middle schools in 8 states with diverse SES (mainly white and low to middle class). Teachers: 9 science teachers with an average of 5 years experience with the *JASON* Project. Students: 269 students Comparison: In 2 of the 8 schools comparison classrooms were matched on grade and general ability level for pre-post inquiry test	Teachers reported that the *JASON* curriculum made using hands-on and project based learning easier to implement and increased collaboration. Teachers reported using more assessment techniques such as portfolios and presentations. They also reported using more technology in the classroom. Based on surveys, interviews, and observations researchers concluded that *JASON* students used hands on learning activities and made real world connections helping them understand complex science concepts. 66% of *JASON* students made an overall gain of 1 to 10 points on the inquiry test. More made gains in process (67%) than in content (46%). Two schools were matched with a control classroom within their school. In these two schools, 59% of the 90 *JASON* students made overall compared to 20% of the 63 comparison group students. It was not reported if these gains were significantly different.

Baxter, Woodward, & Olsen (2001) Journal article National standards: 1989 NCTM	SB Curriculum: *Everyday Mathematics* (EM) Study examined how low achieving students learn in the context of using standards-based curriculum and instruction. Method: Qualitative Measures: Data were collected over one school year through classroom observations, weekly conversations with teachers, and a semi-structured teacher interview at the end of the year. Sample: Two schools in Pacific Northwest were selected because they were using EM: 104 3rd grade students and 5 3rd grade math teachers. 16 students were identified as low achieving or at risk. Target students were identified based on a learning disabled classification on an IEP, performance at or below 34th percentile on the ITBS math section or classified by teachers to be low achieving. There were 2 to 4 target students in each class.	A typical math lesson included whole group work and pair work. Half of each lesson included "student talk" where students talked about their solutions and how they got them. During whole group discussions low achievers were passive, unengaged, and often off task. During pair work low level students were more engaged but at a lower level and most often participated in a nonmathematical capacity. Across the five classrooms all teachers used manipulatives, but only two classrooms used manipulatives to engage students in mathematical reasoning and to represent different ways of thinking. One teacher used “ad hoc” groups to focus on a skill or problem. This involved 8 to 11 lower achieving students. The teacher’s aide worked with the rest of the class while the teacher worked with the ad hoc group. There was high involvement for all students using this method.
Ben-Chaim, Fey, Fitzgerald, Benedetto, & Miller (1998) Journal article National standards: NCTM	SB Curriculum: *Connected Mathematics Program* (CMP) Study examined student performance on rate, density, ratio, and scaling problems comparing students who were taught using Connected Mathematics for one year and students using a traditional curriculum Method: Mixed methods (quasi-experimental) Measures: A post-test of proportionality problems presented in three forms distributed randomly in each class was administered. 25% of the students were selected for interviews to gain a better understanding of their thinking. Sample: A control group was selected from a control population identified by the CMP evaluation team. Equivalence of the two samples was assessed using standardized test results that were reported for both groups. These results showed that the control students were slightly higher at the beginning of the year and slightly lower at the end of the year compared to CMP students. Treatment group- 187 students, 7 teachers, 8 7th grade classrooms located in Michigan, San Diego, and Pittsburgh.	Problems were scored as correct with correct support work, correct answer with incorrect support work, correct answer only, incorrect answer, incorrect answer with partial understanding, and incorrect with incorrect thinking. CMP students outperformed the control students. CMP students outperformed control students on all individual problems. The majority of students in both groups provided some support work with their answers but CMP students were more proficient in their writing and used a wider variety of strategies to solve problems. The control group students had twice as many correct answers with incorrect support work than CMP students. CMP students used strategies that were identified as more effective to solve problems more often than students in the control group.

	Control group- 128 students, 6 teachers 6 7th grade classrooms located in Michigan, Toledo, San Diego, and Pittsburgh.	
Briars & Resnick (2000) Technical report National standards: NCTM District Standards: Pittsburg Public Schools	SB Curriculum: *Everyday Mathematics* Study examined weak implementer schools compared to strong implementer schools on SB assessment and Iowa Test of Basic Skills (ITBS) Method: Quantitative quasi-experimental SB Assessment: New Standards Mathematics Reference Examination (NSMRE) SB Instructional Guidelines: NCTM Measures: Student achievement scores for 3 years on NSRME and the ITBS Survey Battery, Form K. Sample: Demonstration teachers in each Pittsburgh elementary school rated all 3rd and 4th grade teachers' use of Everyday Mathematics over two years. Teacher ratings were aggregated to identify strong or weak implementation schools. 3 weak implementer schools were identified (182 students) and 7 strong implementer schools were identified (291 students). All schools were in the Pittsburg Public School District.	Significant differences in 4th grade students' math achievement scores on both tests (NSMRE & ITBS) were found between strong and weak implementer schools favoring strong implementer schools. The results were most significant on the NSMRE which measured skills, concepts, and problem solving. Performance increased dramatically in the group of 4th graders who were taught with the full SB policy (curriculum, assessment and professional development) since kindergarten. The achievement gap between white and African American students was closed in the strong implementer schools. Professional development was essential for successful strong implementation of a SB Curriculum.
Cain (2002) Journal article National standards: NCTM	SB Curriculum: Connected Mathematics Project (CMP) – a middle school curriculum Formative evaluation of a standards-based math curriculum used in a school district. Method: Mixed methods (non-experimental) Measures: Scores from the Louisiana Educational Assessment Program (LEAP 21) and the Iowa Test for Basic Skills (ITBS) were obtained for the school years 1998-1999 and 1999-2000. Interviews of CMP teachers, district superintendent, and math/science coordinator were conducted. Weekly site visits were conducted and a teacher and student attitude questionnaire was distributed Sample: Schools included in the study were located in a school district in Louisiana. Approximately 3,500 students in the district were involved in CMP. There was a wide range of SES and demographics across the district. There were 34 CMP teachers. Four of the middle schools had full CMP implementation, another five middle schools were in the process of implementing CMP, four middle schools piloted the program seven	ITBS math results for the years 1998-1999 and 1999-2000 showed that CMP students' average score was higher than non-CMP students in grades 6 and 7 for the four schools using CMP beginning in 1998-99. This was also true for the six schools using CMP in the 6th grade in 1999-2000. LEAP 21 results for 8th grade showed that CMP students had a higher passing rate than non-CMP students. Teacher questionnaires (N=28) indicated that the majority of teachers liked CMP better than other math programs and reported that students improved their communication and reading skills, CMP helped improve their understanding of basic math concepts, and helped students to be better problem solvers. All of the teachers reported that professional development in CMP was very beneficial. Student questionnaires (N=300) indicated that the majority of students reported that CMP helped them to be better problem solvers and that CMP activities were helpful in learning math. Teachers new to CMP reported that it was

	years ago, one school is its first stage of implementation, and two will begin implementation in the next year.	challenging to teach. It was concluded that it takes 1 year to be comfortable with the materials and format of CMP. *Note: Evaluator was a lead CMP teacher in the district and interacted with teachers and students during classes.*
Carroll (1997) Technical report National standards: NCTM	SB Curriculum: *Everyday Mathematics* (EM) Study examined the impact of a standards-based math curriculum on student achievement on a state test. Method: Quantitative non-experimental Measures: 1993 scores from the Illinois Goal Assessment Program (IGAP) were obtained. Scores for Cook County (excluding Chicago city schools because no 3rd grade classes were using the curriculum in Chicago at the time of the study) and state scores were used as comparisons. Sample: 26 schools were chosen in Illinois that were using EM in all of their 3rd grade classes. These schools were a cross section of suburban schools in Chicago. There were 1,885 students using the reform in these schools. 14 schools used the EM curriculum since kindergarten and 12 schools had added EM in 1992 or 1993.	EM schools scored higher on the IGAP than the state mean and the mean score for Cook County. All 26 schools scored above the state average and 23 schools scored above the average for Cook County. The 14 schools that had implemented EM since kindergarten scored higher than the state average and also had a higher mean score than the schools that had recently implemented EM. Schools with the highest poverty levels scored well above the state average as well as the average for Cook County on the IGAP. Half of the students showed no change in scores between 1992 and 1993 but the 14 schools that had implemented EM since kindergarten showed more of an increase in scores from 1992 to 1993. Half of the students in the 14 schools that had implemented EM since kindergarten exceeded state goals and only 2% failed to meet state goals.
Caron (2002) Dissertation State standards: IN	SB Curriculum: Indiana State Standards Study on the impact of state standards on high school English teachers' attitudes of standards and their instruction. Method: Mixed methods (non-experimental) Measures: A researcher-developed teacher survey measured demographics, teacher attitudes, and instruction and assessment practices. Teacher interviews were also conducted Sample: Surveyed 238 high school English teachers who were members of the National Council for the Teachers of English (NCTE) and interviewed 14 teachers from three schools in three different districts varying in SES and state assessment scores.	Almost half of the teachers reported no significant change in their practices due to standards. Surveyed teachers reported increases in their use of cooperative groups, collaboration, and interactive discussions. Teachers reported a slight decrease in lecture, text-based worksheets, and restricted response exams. A third of teachers indicated an increase in their use of newspaper and magazine reading but that students were not spending more time reading in general. Teachers reported using more short essays, 3+ page reports, letter writing, and creative writing. One third of teachers reported using oral reports and interactive discussions more often. However, the same amount said that they were able to individualize instruction "much less

		often" or "less often". Interview data indicated that teachers reported little change in their practices because they perceived that these skills were already present in their repertoire.
Chance & Anderson (2003) Conference presentation State standards: NV	SB Curriculum: Nevada science standards Study on teachers' and principles' perceived impact of the Nevada science standards across six areas: instruction, curriculum, assessment, accountability, professional development, and supervision. Method: Mixed methods(non–experimental) Measures: The Nevada Science Standards Questionnaire (teacher and administrator versions) and interviews Sample: Questionnaires were sent to all secondary science teachers and principals in Nevada; 195 teachers (46% response rate) and 56 administrators (43% response rate) responded. A random sample of these teachers and principals were interviewed.	Interview data suggested that teachers concentrated more on documenting standards than actually changing their instruction. Only two-fifths felt that the standards impacted their instruction. Principals had significantly higher perceptions than teachers on the impact that state standards had on: helping teachers develop integrated approaches to content areas, positively impacting science instruction, allowing science teachers to emphasize teaching and learning, and providing common expectations for all students.
Firestone, Camilli, Yurecko, Monfils, Mayrowetz (2000) Electronic Journal State standards: NJ	SB Curriculum: NJ state standards The study examines how the introduction of state standards three yrs prior affects the teaching of 4th grade math and science teachers. Method: Quantitative non-experimental Measures: Classroom practice was measured through surveys, interviews and reviews of teacher lessons. Sample: A statewide sample of 600 4th grade teachers were asked to respond to a request for information about the new standards. Of these 245 competed telephone interviews, 172 completed mailed questionnaires and 110 provided examples of classroom activities.	A small percentage of teachers reported teaching more math and more science since the standards were implemented. Researchers identified topics in both mathematics and science that represented topics that would have been part of the curriculum before the state standards were implemented and topics that were likely added in response to the state standards. In math, teachers reported spending more time on the traditional topics than the new topics but did report that they increased the time spent of the new topics. Teachers in wealthier and poorer school districts reported spending more time on the new topics than those in middle income districts. In science, teachers reported spending more equivalent amounts of time on old and new topics in science than they did in math. There were no differences between districts based on the average SES. Teachers reported that access to materials was improving in lower SES school districts. Teachers in lower SES school districts reported more teaching to the test than those in the wealthiest districts.

Fuller (2001) Conference presentation National standards: NSES State standards: MA	SB Curriculum: *Partners Advancing Learning of Math and Science (PALMS)*, a state wide systemic initiative funded through NSF. The study examines teachers' perceptions of changes in their teaching practices in the 6th year of SB Curriculum implementation. Method: Mixed methods (non-experimental) Measures: Teacher survey was administered to measure classroom practices and student growth. The survey included quantitative questions and qualitative open ended questions. Sample: 62 K -12 teachers in 1 school district that were in their 6th year of PALMS implementation. They had an average of 21.61 years of experience, were 91.8 % female, and 93.9% white.	Teachers reported using the *PALMS* approach most often in science and a little less in math. They rated the overall effectiveness of PALMS as 3.92 on a 5 pt. scale. ANOVAs and paired t-tests between teachers' ratings of their practice before *PALMS* and in their 6th year suggested that teachers perceived significant improvement in their classroom management, resource utilization, classroom and school culture, student growth in their classes, and their ability to help students develop to their fullest potential. 6 themes emerged on open-ended questions about *PALMS'* effects: increased emphasis on reflective teaching, benefits for students, changed teaching methods, increased stress, increased preparation, teaching and remediation time, and increased use of teaching materials. Teachers who believed that *PALMS* was good for the students were still using the approach 6 yrs after its implementation.
Fuson, Carroll & Drueck (2000) Journal article National standards: NCTM	SB Curriculum: *Everyday Math (EM)* The study follows a group of students in SB math curriculum and compares them with other national samples. Study 1 compares second graders in their second year of EM with second graders in other studies and study 2 compares a sub set of these students in third grade with a NAEP comparison group. Method: Quantitative quasi-experimental Measures: In Study 1 math achievement was measured through items from a number sense test and math achievement test used in the comparison study. In Study 2 math achievement was measured on the NAEP test and items from a comparison study (Wood & Cobb, 1989). Sample: The original *EM* sample came from 11 schools in six school districts spread across urban, suburban and rural areas: Study 1 included 343 out of the original sample of 496 and compared them to 29 US and 33 Japanese students in another study (Okamoto, Miura & Tajika, 1995). Study 2 included 236 of the *EM* students and compared them to a subset of national NAEP data and to students in another study (Wood and Cobb, 1989)	Study 1: *EM* students scored significantly higher than the US. comparison despite having a lower SES background, but scored lower than the Japanese comparison on the math achievement test. On the number sense test, the EM group scored significantly higher on 2 items than the U.S comparison, ,but scored lower on one item than both the U.S. and Japanese comparison groups. Study 2: *EM* third graders scored higher overall than the NAEP comparison and significantly higher on 6 items of the number and computation test. On the test used by Wood and Cobb (1998) *EM* students scored significantly higher than the comparison on 8 of 9 items in Geometry and reasoning.

Goertz, Floden & O'Day (1995) Technical reports/ book State standards for Language Arts and Math: California, Michigan and Vermont	SB Curriculum: state math and language arts curriculum in California, Vermont and Michigan. Study of state approaches to reform including capacity, policy, governance, and district, school, teacher practice and teacher responses to reform. Method: Mixed methods (non experimental) Measures: Teacher interviews, teacher surveys - Survey questions were drawn from previous studies including the Student and Staffing Survey, the National Education Longitudinal Study, the National Survey of Science and Math Education, and others. Sample: purposive sample of 12 reforming schools (4 in each of three states). Each states' sample included two districts each with 2 schools, one elementary and one middle school. Total of 60 teachers (5 in each school) were interviewed and surveyed.	Teachers reported being influenced by state policy, but report that their own knowledge and beliefs have a larger influence on their teaching practice. Language Arts: Across states, teachers report teaching topics that are in alignment with the recommendations of State and National reforms, but continue to also teach traditional topics. Practices vary somewhat by state. Math: Across states: there was a similar pattern of topic emphasis across states. Teachers report spending about a third of instruction time on problem solving and communication about math— which is an increase from earlier studies (Porter, 1989). Traditional topics were also taught.
Goldenberg, Ba, Heinze, & Hess (2003) Technical report Year 3 National standards: NSES	SB Curriculum: *JASON* Project- multimedia science curriculum for middle grades. Third year evaluation: Case studies of 9 schools were conducted to examine of a SB multimedia science curriculum and its impact on different types of students' learning. Method: Mixed methods (non–experimental) Measures: Interviews, observations, teacher use surveys, filed logs, student feedback survey, and student content activity (teachers chose a lesson on kelp, forests, seals, or land movement.) Sample: Follow up with sub–sample of teachers who participated in the second year evaluation. 9 public school sites, 12 teachers, 30 classes, and 608 students were involved. Students were identified as mainstream, at-risk, gifted and talented, and special ed. Teachers were from grades 5-8, and had an average of 6 years experience with the *JASON* project.	Teachers reported that they incorporated more hands-on activities and that there was more collaboration among students. Teachers also reported more collaboration with other teachers around *JASON* activities. Teachers adapted the *JASON* curriculum to fit student needs. Teachers of special education students reinforced literacy by using visual aids, highlighted key vocabulary words, and modified worksheets, activities, and *JASON* novels. Students of all literacy levels were able to use scientific, sophisticated vocabulary during the student content activity. Overall, students demonstrated good content knowledge. Students were able to ask relevant questions about the activity. About half of the students could explain how technology was used in the activity but not how it would help answer the research question. Interview and survey data across sites suggested that support for hands-on activities increased student motivation, engagement, and knowledge acquisition.
Grant & Kline (2001) Conference presentation National standards:	SB Curriculum: *Investigations In Number Data and Space* (NSF funded) Qualitative Ethnographic study examined the implementation of a math curriculum.	Despite agreeing with the reform philosophy in theory, the teacher struggled with implementation. Some difficulties he experienced included: inconsistent focus on student learning, use of counter reform practices such as teaching terminology before

NCTM	Method: Qualitative Measures: Daily classroom observations for 8 weeks, teacher interviews after before and after lessons, and a final teacher interview Sample: An experienced 5th grade teacher in the first year of implementing a new reform math curriculum	the unit, focus on just one or two strategies, and not pursuing student's incorrect solutions. The authors conclude that the ability to engage with student's ideas and the belief in the importance of developing a variety of ways of reasoning are important factors in a teacher being able to implement the reform curriculum.
Grisham & Brink (2003) Journal article State standards: WA	SB Curriculum: Washington literacy standards Case studies of three teachers and the impact that the Washington State reform has had on their classroom practices in literacy. Method: Qualitative Measures: Teacher interviews and classroom observations were conducted over the 1997-98 school year. Follow-up interviews were conducted with each teacher in 2000. Sample: Teachers were randomly selected from those who participated in the school/university partnership and the Year of the Reader Workshop series. Teachers were from the same district but different elementary schools. Two teachers had Master's degrees and were involved in some capacity in the reform movement at their school. Years of teaching experience ranged from 11 years to over 30 years.	Themes emerged from the data across teachers. First, teachers were supportive of the reform and standards but interpreted them to fit their own personal beliefs. Second, teachers indicated that the Washington standards impacted their practice, however teachers' actual practice lagged behind their knowledge of SB reform. Third, teachers were concerned about the impact of the state assessment and accountability on students and their instruction. For example, one fourth grade teacher moved to fifth grade because students were tested in the fourth grade and she felt that there was a narrowing of the curriculum to focus on how to take the test. Lastly, teachers valued learner-centered professional opportunities. They felt there was a need for more practical and hands-on professional development. Reported changes in instruction by teachers included having students explain their thinking and write answers in complete sentences. Teachers also reported using more guided reading instruction and small groups.
Hannafin, (2002) Technical report State standards: MA	SB Curriculum: *Plato* remedial program in math for sophomores whose scores were at or below passing on their 8th grade state assessment scores. The study used a pre-post comparative design to examine the effect of the SB Curriculum on state assessment scores. The *Plato* group's scores were compared to a non-*Plato* group's scores from the same school. Overall school scores were compared to state scores. Method: Quantitative Quasi-experimental Measures: Pre-post student gain scores on the Massachusetts Comprehensive Assessment System exam between 8th and 10th grade used as student achievement measure. Sample: 1 low performing high school in Cape Cod: 87 at risk students in *Plato* program were compared to a convenience sample of 39 students	Gain scores from the 8th grade pretest to the 10th grade post test were computed. Student scores in both the *Plato* group and non-*Plato* comparison group increased significantly. The *Plato* groups' gain scores were significantly higher than the non-*Plato* comparison groups' gain scores. The sample school's 10th grade test scores were lower than the state average before PLATO was introduced but improved and were in-line with the state average after PLATO was implemented (note: not all students used PLATO). *Note: The Plato and non-Plato comparison group were not matched or randomly assigned.*

	not in the *Plato* remedial math program.	
Haug (1998) Conference presentation State standards: CO National standards: NCTM	SB Curriculum: Colorado standards-based education policies in math Case studies of two school districts examined how teachers' instruction and curriculum was impacted by math standards. Method: Qualitative Measures: Data were collected from teacher and administrator interviews, classroom observations, and district and classroom documents. Sample: School districts were recommended by experts (reputational sampling) based on their demonstrated implementation of SB education. An elementary, middle, and high school were chosen in each district and the principal at each school nominated two teachers that were representative of the school. In the Northern School District, the Director of Assessment, 3 principals, and 12 teachers participated. In the Eastern School District, 3 administrators, 3 principals, and 12 teachers participated. Northern School District- served between 10,000 to 20,000 students, 10% minority, and 13% FRL Eastern School District – served between 5,000 to 10,000 students, 50% Hispanic, and 60% FRL	Teachers' understanding of SB education varied. In Northern, teachers saw reform as more simplistic; standards "give a new name to old practice" (p. 14). In Eastern, teachers viewed SB education as complex and saw it as a huge shift in their teaching. However, some teachers thought the standards helped provide more meaningful math education. Teachers reported more problem solving activities, since the advent of standards but expressed concern that students were not getting enough of the "basics" like computation and faced issues in integrating problem-solving and computation. Teachers in both districts integrated writing in their math instruction. At the elementary level teachers integrated math across subjects but also reported that they had to drop topics. The majority of teachers in Northern (10 teachers) and very few in Eastern incorporated statistics and probability in the curriculum. Those who did include statistics and probability in their math class taught it in isolation instead of integrating it with other math topics. The majority of teachers in both districts reported that they used curriculum maps and unit organizers to organize their material and document their use of standards. There were teachers in both districts (5 in Northern and 3 in Eastern) who reported no impact from SB education.
Huntley, Rasmussen, Villarubi, Sangtong, & Fey (2000) Journal article National standards: NCTM	SB Curriculum: *Core-Plus Mathematics Project* (CPMP) – a high school program Study examined the effects of a SB math curriculum on students' algebraic understanding, skill, and problem solving ability compared to students using a traditional math curriculum. Method: Mixed methods (quasi-experimental) Measures: Teacher interviews and student assessments with 3 parallel forms were developed. Part 1 included contextual problems, part 2 included symbolic manipulations with no contextual information, and part 3 included open ended contextual problems to be worked on in pairs. Sample: Six of the 36 CPMC field test schools	Interview data indicated that implementation of CPMP and traditional programs varied greatly across sites. The majority of students across both groups were below average on their math achievement. Neither group performed well on basic symbolic calculations found on college placement tests. There were significant differences between the control classrooms and the CPMP group. CPMP students outperformed control students on part 1 and part 3 of the student assessments. Control students outperformed CPMP students on problems related to algebraic symbol manipulations such as, testing equivalences of expressions and solving

	accepted an invitation to participate. 2 schools were located in the Southeast, 2 in the Midwest, 1 in the South, and 1 in the Northwest. There were two CPMP teachers and between 90 to 180 students at each site. There were between 1 and 3 control teachers at each site. Control classes were drawn from advanced algebra classes and matched on standardized test scores from 8th grade in four sites. One site randomly assigned students to CPMP or traditional math classes at the beginning of the 8th grade. Control classes at the sixth site did not have achievement data available to match on but the school made assurances that the groups were comparable.	equations and inequalities. Overall, CPMP students performed better on tasks that were applied problems within a contextual setting and students could use graphing calculators. Control students performed better on traditional symbol manipulation tasks when there was no context and no calculators.
Indiana University (1998) Technical report State standards: Unspecified; "most states" National standards NCTE International Reading Association	SB Curriculum: *Signatures* (Harcourt Reading Program) Study examined the effects on student achievement and attitudes towards reading before and after exposure to one 4-6 week theme from the SB Curriculum. Method: Quantitative non-experimental Measures: Student scores on Stanford 9 reading test and the program's "means skills test". Sample: 18 volunteer teachers from Illinois, Massachusetts, Minnesota, Missouri, Ohio, and New Jersey and who had not previously used the *Signatures* program (9 first grade and 9 4th grade)	The raw score on the Stanford 9 test and the program's mean skills test increased significantly for both grade 1 and grade 4 students
Jerome & Gilman (2003) Technical report State standards: Indiana	SB Curriculum: a school-wide writing improvement program Study compared third grade students in 2000/2001 who had received a schools two-year SB writing improvement curriculum to those in 1999/2000 who had not received the program. Method: Quantitative quasi-experimental Measures: Student scores in writing on the Indiana state achievement test (ISTEP+) Sample: sample of 23 student's ISTEP+ writing test scores were taken from three classes in 2000/2001 and were compared to a control group sample of 23 students in three classes in 1999/2000.	The control group had a significantly higher average score on the achievement test than the group who used the writing program. *Note: The control group comes from a previous year. No attempt was made to measure or control for prior achievement.* *Only a small sample of student's test scores were examined for both control and treatment groups.*
Johns (2004) Doctoral	SB Curriculum: *Everyday Mathematics* SB Assessment: state assessment A study to examine how beginning elementary	Differences in the math curriculum and the way the teachers had been taught math caused a conflict for group 2 and 3 teachers who then deviated from the curriculum's intentions

Dissertation National standards: NCTM State standards: MD	teachers understand and use a SB math curriculum (also briefly addresses assessment influences on teacher instruction). Method: Qualitative (cross-case analysis) Measures: 2 in-depth interviews with each teacher, at least 1 classroom observation per teacher (with pre & post interview), and document reviews (e.g., state standards, curriculum materials, lesson plans) Sample: 6 elementary schools in same district Total of 11 second-year elementary teachers in grades 1-5 were divided into groups based on comfort level and expertise in math Group 1 = frustrations with math – both learning & teaching (n = 5) Group 2 = enjoy math & confident teaching (n= 3) Group 3 = very confident with math & specialized training (n = 3)	when teaching. Group 1 teachers used the curriculum to enhance their knowledge of math so they were less conflicted and had less deviation When using the curriculum, group 1 teachers were simultaneously learning math content and how to teach math which can be problematic in terms of being able to interpret student work. Content of state and local test influenced teachers' decisions about what elements to teach in their classrooms and what elements to leave out. Teachers in all 3 groups expressed a strong desire for more professional development around the curriculum.
Kerr (1999) Dissertation National standards: NCTM State standards: IL District Standards: Chicago Public Schools	SB Curriculum: *Math Trailblazers* Study examined the influence of a standards-based math curriculum on elementary student achievement. Method: Mixed methods (non-experimental) Measures: classroom observations (one per classroom, once a month for the school year), 4 focus groups (1 hour duration each) with all 6 teachers, student and parent attitude survey, and teacher survey Student achievement data: Illinois Goal Assessment Program (IGAP) – criterion-referenced, mandated by the state and district. Iowa Test of Basic Skills (ITBS) – standardized, norm-referenced test mandated by district. New Standards (America's choice) Mathematics Performance Reference Examination TIMS Multidimensional rubric Sample: 147 3rd grade students and their 6 teachers in one elementary school in Chicago	There was an increase in student test scores on IGAP, ITBS, and New Standards Mathematics Performance Reference Examination after the first year of implementation. TIMS multidimensional rubric was only used in 1998, after implementation of the curriculum. It showed that 80% of students met the 3 criteria for demonstrating emerging math power (math content, problem solving, and math communication). Post implementation student attitude toward math surveys indicated generally positive attitudes. Post implementation surveys of parents' attitudes toward reform math education were generally positive, with the exception of the statement "my child is smart at mathematics" where less parents agreed. Teacher observations showed a constructivist approach to math. Teacher survey and focus groups indicated positive attitudes regarding aspects of the SB Curriculum such as use of calculators, using multiple problem solving strategies, and students working together. *Note: Researcher was the principal at the school where the program was implemented. No validity evidence for any of the surveys used.*
Lawrenz, Huffman,	SB Curriculum*: Scope, Sequence, and*	There were problems sustaining

& Lavoie (2001) Technical report National standards: NSES	*Coordination* (SS&C), an inquiry based 9th and 10th grade science curriculum Six year study of a science reform and its impact on students who participated in the reform compared to those not in the reform. Method: Mixed methods (quasi–experimental) Measures: Student and teacher questionnaires, classroom observations, student, teacher, and principle interviews, researcher developed tests, and laboratory experiments (randomly selected students participated in laboratory assessment). Sample: Five case study schools in TX, WA, NY, CA, IA, and MT were chosen as representative of locale, population, and ethnicity. The 5 case study sites included 20 9th grade science teachers and over 1,500 students. A comparison group was drawn from the same schools the year prior to the reform.	implementation: Among the 5 case study schools 3 were implementing the program at the end of the study, but one had modified it. Classroom observations showed that across the five case-study sites the amount of time spent on student centered activities increased (significance was not reported). The two sites that had continued using the unmodified reform showed more gains in time spent on student-centered activities. Many teachers reported that they adapted the SS&C curriculum to fit their classroom needs Results were mixed regarding student achievement. The results of test items were inconsistent, but with a trend towards higher achievement in the comparison year. The general pattern for the laboratory measures of achievement showed positive gains in reform years/groups from the comparison year. Researchers concluded there was a trend for student achievement to improve slightly if they had been exposed to the reform for two consecutive years. *Note: the comparison and treatment groups did not occur simultaneously.* *Results for each individual case-study site are included in the report.*
Lee (1998) Journal Article National standards: NCTM State standards: CA, MN Also includes Lee (1996) Conference Presentation	SB Curriculum: Content-driven state policies Study used secondary data analysis of national survey data to examine the influence of state education reform on instruction in middle grade math (includes instructional guidance influences). Method: Quantitative non-experimental Measures: National surveys: 1991-1992 Council of Chief State School Officers (CCSSO) state policy survey 1992 National Assessment of Educational Progress Trial State Assessment (NAEP TSA) 8th grade mathematics teacher and school survey MN and CA state level policy documents were compared	Teachers reported using more SB instructional practices when student assessments *and* textbooks were linked to a SB Curriculum. Implementation of a new SB Curriculum alone was not shown to have an impact on teacher instructional practices. Principals perceived an association between state-level curriculum policies and teacher instruction, however analysis showed this was not a significant relationship. In states with a direct link between student assessment and a curriculum framework, principals reported a greater impact on instruction than principals in states with no such link. MN focused on a short term approach and created outcome based practices that were broad in terms of content. CA had a long term approach with a state framework that outlined specific guidelines with examples of how to implement content and SB instructional practices. CA had shown more changes toward SB instructional practices by early 1990s than MN did based on principal and

		teacher reports.
Marx, Blumenfeld, Krajcik, Fishman, Soloway, Geier, & Tal (2004) Journal National standards: AAAS, NSES	SB Curriculum: four 8-10 week SB science units developed by the NSF funded Center for Learning Technologies in Urban Schools The study examined student learning over the course of three years in which a total 4 SB science units were taught. Method: Quantitative (non–experimental) Measures: Student learning on each unit was measured by a pre-post test developed for that unit. Each test was divided into content and process scores. Sample: 8000 6^{th}-8^{th} grade students in 14 schools in the Detroit school district (over half below the poverty line and 96% minority)	Within subject t-tests were significant and showed pre-post student gains for 7 of the 8 content and process tests. The effect sizes were stronger for content scores than for process scores. The weighted average of effect sizes grew stronger across the three years. The researchers suggest this could reflect a cumulative positive effect of repeated exposure to SB units.
McLaughlin, (2000) Technical report State standards: unspecified	SB Curriculum: varied – the state and/or district curriculum in math/ language arts. Study examined district's implementation of reform curriculum and accountability policies on instruction received by special education students. Method: Qualitative Measures: Classroom observations of 28 elementary school and middle school language arts and math classrooms; 60 interviews of principals, teachers who taught the observed classes, and/or affiliated special education teachers. Sample: 4 districts in 4 states in different stages of SB Curriculum implementation: 1 urban, at-risk (70% FRL, 88% minority), 1 rural, 1 suburban, 1 affluent	Urban district: Both general and special education teachers report teaching to the test due to high accountability. Teachers reported relying on district curriculum. Observations suggested that teaching was primarily teacher driven lecture with very little small group or technology time and variable. Some classes seemed to have no purpose. Special education was mostly not integrated and teachers taught to lower level goals. Rural district: Observed variability in classroom instruction in terms of traditional vs. student centered instruction. Most of the responsibility for adapting materials and instruction for special education fell on the special education teachers even though students were often integrated. Suburban district: Observed teachers using standards to guide instruction including small groups. Observed a wide range of accommodation and modifications for special education students—usually at the time of instruction. District use of portfolios worked well for special education students. Affluent district: Observed student directed instruction and a variety of groupings. Special education was

		integrated and technology was used extensively
Michlin, Seppanen, & Sheldon (2001) Technical report National standards: NCTM State standards: MN	SB Curriculum: variety of NSF developed math curriculum Study of the impact of SB Curriculum on teacher attitudes and practices. Phase 1 of a multi-year study. Method: Mixed methods (non-experimental) Measures: A teacher survey, teacher interviews, and classroom observations were used to collect data on teacher activities and current teaching and classroom activities. Sample: Districts and school sites with strong SB Curriculum implementation were identified. Eight districts were chosen and 2 schools in each district served as case study sites. District characteristics were diverse:0% to 22% LEP 9% to 66% FRL 27% to 97% White 500 to 50,000 students	Results should be considered as baseline data because sites varied in the number of years (1 to 4 years) that the curriculum had been implemented (researchers assumed that full implementation takes 5 or more years). Teachers reported that they were making changes in their instruction mainly in three areas: 1) improvement in teacher questioning and listening, 2) less lecture and more facilitating, and 3) physical environment of classroom. However, teachers continued to supplement their instruction with more traditional curriculum materials. Classroom observations indicated that teachers were still in the early phase of making changes in their practices using traditional practices most of the time and reform practices less than half of the time. The survey data indicated that the majority of teachers used lecture every day but also reported that they had student's explain their reasoning every day and addressed alternative solutions to problems.
Parker & Gerber (2000) Journal Article National standards: NSES State standards: GA	SB Curriculum: Science program aligned with state and national standards. Ten, 2-hour lessons delivered over a 5 week period. The study measured the results of the SB program on student learning and attitudes towards science. Methods: Mixed methods (non–experimental) Measures: Student scores on a pre and post researcher developed criterion-referenced test, teacher logs of student behavior, pre-post student attitudes toward science survey Sample: 11 4th and 5th grade African American students (5 boys, 6 girls)	T-tests on pre and post scores were statistically significant and showed improvement for all 11 students. Teacher logs indicated students were acquiring science knowledge over the course of 5 weeks. Student attitudes toward science improved during the 5 week program as evidenced by the student survey and teacher logs.
Reys, Reys, Lapan, & Holliday (2003) Journal article National standards: NCTM State standards: MO	SB Curriculum: *Math Thematics* (MT) and *Connected Mathematics* (CMP), both NSF curricula Study examined achievement levels in districts using SB Curriculum compared to districts using more traditional curriculum. Method: Quantitative quasi-experimental Measures: 8th grade state assessment results in	No significant differences were found between SB districts and comparison districts on the state assessment prior to the SB Curriculum being introduced. All significant differences favored SB students after two years of the SB Curriculum. Based on a chi–square test, in the first matched pair there was a significantly greater number of students who scored in the highest

	mathematics used as student achievement measure Sample: Districts that used SB curricula beginning in fall 1996 were matched with comparison districts on middle school organization, prior math achievement, locale, and FRL. District matched pairs: 1) SB district- two middle schools, 2000 students, 30% FRL Comparison district- one middle school, 800 students, 24% FRL 2) SB district- two middle schools, 1000 students, 25% FRL Comparison district- middle schools, suburban, 1000 students, 20% FRL 3) SB district- one middle school, suburban , 600 students, 13% FRL Comparison district- one middle school, 800 students, 11% FRL	achievement level of the state mathematics test than the comparison district. There were no significant differences in the two other district matched pairs. T-tests on the six content standards of the state test showed that students in all three SB districts scored significantly higher on the data analysis, probability, and statistics and algebra sections of the state mathematics test than students in comparison districts. In the first two district matched pairs, SB students scored significantly higher three additional content strands, number sense, geometric and spatial sense, and discrete mathematics. There was no significant difference between any of the matched districts on mathematical systems.
Ridgway, Zawojewski, Hoover, & Lambdin (2003) Journal article National standards: NCTM	SB Curriculum: *Connected Mathematics* (CMP) – a middle school program Evaluation study of the effect of *Connected Mathematics* on student achievement in grades 6-8. Looked at growth over one year. Method: Quantitative quasi-experimental Measures: The Iowa Test of Basic Skills (ITBS) and the Balanced Assessment were used as achievement measures. Sample: Volunteer samples of CMP and non-CMP populations. Nine sites participated across the country. Two sites dropped out in second year. Groups were matched on ability, locale, and diversity. CMP group- 338 6th grade students, 627 7th grade students, 820 8th grade students Non-CMP group- 162 6th grade students, 234 7th grade students, 275 8th grade students	Scores from a pretest were used as a covariate to control for any differences in achievement between CMP and non-CMP. On the Balanced Assessment, an ANCOVA, using Fall ITBS scores as a covariate, indicated that CMP students had significantly higher scores in the Spring than non-CMP students. CMP students started behind non-CMP students but finished ahead. ITBS results were mixed. CMP students started behind non-CMP students in both sixth and seventh grades and showed no signs of catching up. There was no significant difference between groups at the eighth grade. Additional analyses were conducted with the school that had implemented CMP all three years. An ANOVA indicated by the end of year 3, CMP students were significantly ahead of non-CMP students on the Balanced Assessment. An ANOVA of ITBS results showed that CMP students scored significantly higher than non-CMP students .
Riordan & Noyce (2001) Journal article	SB Curriculum: *Everyday Mathematics* (EM) – an elementary program and *Connected Mathematics* (CMP)- a middle school program Study examined student achievement of students	Students using *Everyday Mathematics* or *Connected Mathematics* scored significantly higher than their comparison group on the math section of the Massachusetts

National standards:| NCTM State standards: MA	exposed to SB Curriculum compared to students using more traditional curriculum. Method: Quantitative quasi-experimental Measures: 1999 Massachusetts Comprehensive Assessment System (MCAS). scores in math (only regular education students included in analysis) Sample: Schools were chosen to represent those who implemented SB Curriculum programs (early implementers-4 or more years and late implementers-2 to 3 years). Schools were matched with comparison schools based on state test scores and FRL. All schools ranged from 0-10% F/RL and 81 – 95% white students	Comprehensive Assessment System (MCAS). In addition, student results were higher the longer the school had been implementing the SB Curriculum. Overall, students in the SB Curriculum groups had higher achievement than the comparison groups regardless of ethnicity, gender, FRL, or math ability. However, positive gains were higher for Blacks and Hispanics compared to White students and for FRL students compared to non-FRL students. EM and CMP schools outperformed comparison schools on most question types including number sense, and patterns and functions. There was no difference in performance on geometry and statistics.
Rivet & Krajcik (2004) Journal article National standards: NSES, AAAS State standards: MI District Standards: Detroit	SB Curriculum: *Big Things* project, an 8 week project for sixth grade students focused on developing a machine. Study of students' learning after 8 week SB science project. Method: Quantitative non-experimental Measures: Student learning measured on a researcher created pre/post test designed to be aligned with curriculum. Sample: The SB project was taught for four years by 24 teachers in 15 schools in lower SES neighborhoods, (91% African American, 70% FRL). Sample students: Year 1 -179 sixth grade students, Year 2 -299 sixth grade students, Year 3 -859 sixth grade students, Year 4 -1239 sixth grade students	In each of the four years, students showed statistically significant achievement gains overall and in each of the four learning goals of the project (balanced and unbalanced forces, simple and complex machines, mechanical advantage, and inquiry process) The learning goal related to mechanical advantage had the largest effect size all four years.
Sandholtz, Ogawa, & Scrubner (2004) Journal article State standards: CA	SB Curriculum: District level standards Case study of school district in which there was a standards gap between state standards and the standards adopted by the district. Method: Qualitative Measures: interviews of administrators, school principals, and teachers, classroom observations. A teacher survey on the influence of the district standards on practices and open ended questions related to curriculum and instruction were distributed. A document review was also conducted.	There was a gap between district standards and the state standards. The district's curriculum standards lagged behind the state standards and were differentiated by ability (there were different sets of standards for students of differing ability). The district assessment only covered math and language arts and was based on the district's minimal standards. As a result, teachers in elementary schools limited their teaching to primarily math and LA and their instruction was directed towards low ability levels. Teachers reported that there was no time to use supplemental materials to

	Sample: Medium- sized school district in working-class community: 2 comprehensive high schools, 1 continuation high school, 3 middle schools, 16 elementary schools, 52% FRL, 24% LEP, 65% minority, and 19,000 students	improve their instruction and cover more content. Teachers in elementary schools and teachers of low level secondary courses used more drill and kill, practice, routine teaching strategies, and less hands-on learning. Ninety percent of elementary teachers reported that they focused on basic skills compared to 40% who reported working on higher order thinking skills with their students. Students scored higher on the district assessment than the state assessment reflecting the district's lower standards than the state standards.
Schneider, Krajcik, & Blumenfeld (2005) Journal Article National standards: AAAS, NSES	SB Curriculum: an 8 week Project-Based Science unit funded through NSF's Urban Systemic Initiative (USI). Study examined teachers' implementation of SB Curriculum and classroom instructional practices over an 8 week period. Method: Qualitative Measures: Ratings of teachers' instructional practices as observed in videotaped classroom observations. Instructional ratings were made in 7 categories: accuracy of science ideas presented, completeness of science ideas presented, quality of adoptions, amount of instructional supports offered, and appropriateness of instructional supports. Sample: one 8th grade science teacher in each of 4 Midwestern, urban middle schools. At risk indicators: students were 29-66% free or reduced lunch and 95%-100% minority.	Findings were mixed. Half of the teachers' instructional practices averaged high ratings in all 7 analysis categories — indicating their actions were consistent with the intent of the curriculum. The other half had medium to low ratings in the 7 analysis categories — indicating their actions were less consistent with the intent of the curriculum. The researchers suggest that professional development and instructional guidelines are required to help teachers plan and use SB Curriculum effectively in their classroom. *Note: Detailed description of analysis and ratings of lessons is given*
Schneider, Krajcik, Marx, Soloway (2002) Journal National standards: AAAS, NSES	SB Curriculum: *Foundations of Science* (A three year integrated *Project Based Science* (PBS) curriculum) High-school students participating in the PBS curriculum for 2 to 3 years were compared to three national NAEP Samples: the overall national sample, a sample of white students and a sample of students not eligible for FRL. Method: Quantitative quasi-experimental Measures: Student achievement was measured using the 12th grade NAEP science test.	Multivariate revealed that the PBS students outscored all three NAEP samples overall, and item by item t– tests suggest they outscored the comparison samples on just under half to just over half of the items depending on which comparison is used. Researchers suggest that this study shows that students in PBS science courses will not be disadvantaged on large scale achievement tests.

	Sample: 42 10th and 11th graders in an urban high school: predominately white and middle to upper class	
Schoen, Cebulla, Finn, Fi (2003) Journal article National standards: NCTM	SB Curriculum: *Core-Plus Mathematics Project* (CPMP) – a high school program Study examined teacher variables associated with student achievement in the context of using a standards-based curriculum. Method: Mixed methods (non-experimental) Measures: The Ability to Do Quantitative Thinking (ATDQT), a subtest of the Iowa Tests of Educational Development (ITED) was administered as a pretest and posttest after each of the three courses. Classroom observations were conducted and teachers completed an implementation survey and a concerns survey. The student achievement index was computed using the ITED posttest score (removing the variance due to the pretest). Sample: 40 teachers and 1,466 students across 26 field-test schools. Teachers needed to complete at least five of seven Course 1 units and students needed to complete both the achievement pretest and posttest. The mean student enrollment was 1294 (ranged from 431 to 2,777). The average FRL was 16.1% (range of 1.1% to 46.1%) and the average percent minority was 20.4% (range of 1% to 97.2%).	Regression techniques were used and several teacher practice variables were identified to be significantly and positively associated with growth in student achievement. These variables included professional development on how to teach the CPMP course, more group and pair work, less presentation and whole group discussions, less time spent on non-academic activities, using a variety of assessment strategies, using the instructional and assessment materials provided by the curriculum, and teacher collaboration. The materials provided to teachers as part of CPMP helped teachers implement standards-based teaching and assessment in the classroom. None of the variance in student achievement was explained by demographic variables (FRL, minority, class size, and minutes of class).
Schoen & Hirsh (2003) Book chapter National standards: NCTM	SB Curriculum: *Core-Plus Mathematics Project* (CPMP)– a high school program Study examined the math achievement of students who participated in the Core-Plus Mathematics Project (courses 1-4) compared to students in traditional math courses and in the nationally representative norm group. Method: Quantitative quasi-experimental Measures: The Ability to Do Quantitative Thinking (ITED-Q), a subtest of the Iowa Tests of Educational Development (ITED) was administered as a pretest and posttest. 25 items from the 1990 and 1992 National Assessment of Educational Progress (NAEP) were administered after course 3. CPMP developed performance assessments were given after course 1 and course 2. Sample: Course 1-3 were field tested in 36 sites and course 4 was field tested in 28 sites across 11	T-tests were conducted to analyze differences between the CPMP group and comparison group. Results from the ITED-Q showed that the CPMP had positive effects, especially in the first year. Data for students that was available for all three courses showed that these positive effects were maintained. Data after course 1 showed that CPMP students scored significantly higher on the overall ITED-Q test, Interpreting Information subtest, and Solving Problems subtest than the pre-algebra comparison group. CPMP students significantly higher on the overall ITED-Q test and the Interpreting Information subtest than the algebra comparison group scored. Two year trends across course 1 and course 2 showed that there was an increase of 10

	states. Sites were representative of urban, suburban, and rural locales and ethnic diversity. Sample included 11field test schools who volunteered to pretest and posttest students in traditional, comparison classes. By the end of course 2, five of these schools agreed to test students at the end of year 2. There was complete data for 186 comparison students and 287 CPMP students. All field test CMP students (1,457) were compared to norm groups (i.e., NAEP).	percentile points for CPMP students compared to a 2 percentile point increase for comparison students. Results from the CPMP performance tests showed that students, after course 1 and course 2, were better at reasoning and applying algebraic and geometric methods than comparison students. CPMP students performed better on contextual problems than comparison students. Comparison students performed better at the end of course 1 on algebraic symbolic manipulation with no context but at the end of course 2 there was no difference between groups on these types of problems. CPMP students, at the end of year 3, had higher means than the national sample on all 5 content and 3 all process categories that the 25 item NAEP based assessment tested.
Smith (2003) Conference presentation State standards: VA	SB Curriculum: Virginia's Standards of Learning (S.O.L.) history curriculum Ethnographic study of teachers' instructional methods for teaching history in the context of a new state curriculum and the implementation of the S.O.L. tests. Method: Qualitative Measures: teacher interviews, classroom observations, observations of social studies department meetings, and school and S.O.L. documents. Sample: Five high school social studies teachers from one suburban high school.	Teachers used other sources to represent the views of women and minorities, not covered by the S.O.L. curriculum. Teachers reported that their teaching methods had not changed in response to S.O.Ls, but observations indicated that teachers resorted to lecture formats and "cramming" to cover the material on the S.O.L. test. Teachers of poor readers emphasized reading comprehension strategies to students so they would perform better on the S.O.L. test.
Speas (2003) Year 1 Technical report State standards: NC	SB Curriculum: *Project Achieve* (modified from Brazosport, TX to meet NC Standards Course of Study as well as local needs) First year evaluation of SB instructional initiative designed to increase student performance Method: Mixed methods Measures: Student test scores in 2001 and 2002 in reading and math, WCPSS Effectiveness Index (created using student prior achievement, special education status, and 2 measures of SES), teacher interviews, document analysis Sample: 6 elementary schools, 2 middle schools All in Wake County Public School System	All elementary and middle schools met state ABC High Growth standards for 2002. A higher percentage of students were at or above grade level for both elementary and middle schools compared to percentages in the previous year. One elementary school was named one of NC "Top 25 most improved K-8 school". In general, the overall WCPSS Effectiveness Index scores were higher in 2002 than 2001 in elementary and middle schools.

	(WCPSS)	
Speas (2004) Year 2 Technical report State standards: NC	SB Curriculum: *Project Achieve* (modified from Brazosport, TX to meet NC Standards Course of Study as well as local needs) Second year evaluation of SB instructional initiative designed to increase student performance Method: quantitative quasi-experimental Measures: student test scores in 2001, 2002, and 2003 in reading and math, AYP measurement report, least squares regression analysis to create End of grade (EOG) scale scores Sample: 10 elementary schools and 3 middle schools - all in Wake County Public School System (WCPSS)	All elementary and middle schools met state ABC High Growth standard. A higher percentage of students were at or above grade level for all elementary schools and two middle schools compared to percentages in the previous year. The middle school that was new to the program remained the same. Project Achieve schools, as a group, showed higher growth than Non-Project Achieve schools in reading (grades 4 &5) and in math (grades 3-5) Overall the Project Achieve schools had more FRL, low achieving and special education students than the district average so gains in these schools were interpreted as encouraging.
Thompson, Zeuli, & Borman (1997) Conference presentation National standards: AAAS, NSES State standards: MI	SB Curriculum: Michigan Statewide Systemic Initiative (MSSI) science reform (funded by NSF) Study examined the impact of the MSSI on classroom instruction in science. Method: Qualitative Measures: Teachers were observed and interviewed twice. Sample: To select the sample, researchers used the reform scale of the teacher questionnaire from the Third International Mathematics and Science Study (TIMSS) sent to all 640 elementary and middle school math and science teachers in Michigan. A stratified sample was chosen and 6 elementary (3rd and 4th grades) science teachers and 6 middle school (7th and 8th grades) science teachers were identified. The average years of teaching experience was 15 years and 13 teachers had a master's degree.	Many teachers reported that their teaching reflected reform practices but observations showed that teachers were not reflective of reform practices. Although five teachers taught science using ideas and used an explanatory approach, linking concepts to explain patterns, researchers' determined only two teachers truly incorporated all the components of reform based instruction. Although some teachers incorporated hands-on activities, art, and lively narratives, they often did so in a manner to make the lesson more fun as opposed to giving students an opportunity to generate additional data to examine. Student-student discussions were rare. About half of the teachers emphasized science as facts, described as a more surface level understanding of science. Other non-reform practices observed included: focusing their teaching on procedures to the exclusion of content, failing to move students from their "impressions" of an experiment to any "crystallization of an idea", and giving students the concept rather than going through the steps of developing the idea with them.
VanSledright (2002) Journal article National standards: National Center for History in the	SD Curriculum: pedagogical approach following recommendations of history reform Study assessed a researcher-practitioner's investigative approach to teaching 5th grade history on students' ability to analyze historical documents and images.	Performance tasks were analyzed based on four levels: level 1 – makes sense of text or image using comprehension strategies, level 2 – evaluates text or image and makes judgments about sources, level 3 – corroborates details from different sources, check sources, and draw from growing event knowledge, and level 4 – checks inter-textual

Schools	Method: Qualitative Measures: Pre and post performance tasks were given to students to analyze, videotaped classroom observations, field notes, teacher journal, lesson plans examined Sample: 8 students from the class of 23 were chosen (students were representative of gender, ethnicity, and reading levels).	reliability and constructs an evidence-based interpretation of the event. Majority of students' responses to the first performance task were at level 1 and level 2, over half were at level 1 only. After the investigative approach to learning history was implemented, students used more level 3 and level 4 strategies in the second performance assessment. Monitoring strategies were replaced with using sources in a more systematic way. Students identified primary and secondary sources, corroborated details and made judgments regarding reliability, validity, and point of view of sources. Gains from the first performance task to the second performance task varied among students. Low level readers did not progress to the same point as high level readers. An ESL student struggled with analyzing the texts and images. *Note: The researcher was the teacher of the class.*
Weiss, Pasley, Smith, Banilower, & Heck (2003) Technical report National, state, and district standards – in general	SB Curriculum: influence of national, state, and district standards Descriptive study of what math and science instruction looks like in classroom across the nation and the factors that influence teachers' instruction. Method: Mixed methods (non-experimental) Measures: Data were collected from classroom observations and teacher interviews. Sample: Drawn from a nationally representative sample used for the 2000 National Survey of Science and Mathematics Education using stratification sampling. A subset of 40 middle schools was selected. An elementary school and high school were randomly selected from the same feeder system, so each site included three schools. Teachers were randomly selected for classroom observations in each school – two science and two math teachers. 31 schools and 364 teachers participated in the study. Of the schools, 22% were urban, 61% were suburban, and 17% were rural. The average school size was 797 students with 40% FRL and 39% minority.	Teachers indicated that the content of what they taught in a lesson was influenced by state/district curriculum but not national standards. Teachers also indicated that content was influenced by textbooks/ programs for class, state/district accountability tests, knowledge, experience and beliefs, student characteristics, collegiality, and to a lesser extent, the building/district administrator and principal. Factors that influenced teachers' choice of the instructional strategies used in a lesson were identified as teacher knowledge, beliefs, and experience, textbooks/program for class, student characteristics, district provided professional development collegiality, the principal, and to a lesser extent, state/district standards/frameworks and accountability tests. No teachers reported that their instruction was influenced by national standards. Observers rated few of the teacher lessons as high quality. Lessons of higher quality were from teachers who engaged students' learning and helped students make sense of the content. Rural schools and classes with a high minority had lower quality lessons.

Woodward, Monroe, & Baxter (2001) Journal article National standards: 1989 NCTM	SB Curriculum: *Everyday Mathematics* Study examined students with learning disabilities in math and the impact of two interventions on their problem solving: practice with performance assessments and ad hoc tutoring. Method: Quantitative quasi-experimental Measures: Performance assessments from Everyday Mathematics were modified to resemble extended response problems from the state test. Sample: Seven 4th grade teachers and 182 students from 3 schools in the Pacific Northwest. Students were identified as at risk or learning disabled (LD) if they had an IEP or their ITBS math scores were at or below the 34th percentile. Scale scores from the math problem solving subtest of the Iowa Test of Basic Skills (ITBS) was used to determine comparability of the intervention and comparison classrooms. Intervention group: 4 classes, 102 students- 17 at risk students and 6 learning disabled Comparison group: 3 classes, 79 students- 8 at risk and 5 learning disabled	LD students in the intervention group made larger gains in their performance from pretest to posttest compared to students of average ability and at risk students. LD students were able to complete the first steps of a problem at the posttest and were using more systematic problem solving strategies while the comparison students showed no change in their problem solving abilities from pretest to posttest.
Woodward & Baxter (1997) Journal article National standards: 1989 NCTM	SB Curriculum: *Everyday Mathematics* Study examined low achieving students in math and the impact of a reform based curriculum on their achievement. Method: Quantitative quasi-experimental Measures: All students were given the math portion of the Iowa Test of Basic Skills (ITBS) in September and in April of the school year. The Informal Mathematics Assessment (IMA) was given to assess problem solving strategies. Sample: Three schools in the Pacific Northwest agreed to participate. Schools were suburban, middle class schools with low rates of FRL. Two schools were using EM (five teachers) and the third school acted as a comparison (four teachers). Students identified as at risk or learning disabled if they had an IEP or their ITBS math scores were at or below the 34th percentile. Intervention and comparison classes were comparable along SES and demographics. Intervention group: 104 students- 9 at risk students and 7 learning disabled Comparison group: 101students- 17 at risk and 5	Pretest scores on the ITBS served as a covariate in ANCOVAs. Students in the intervention group scored significantly higher on the concepts subtest of the ITBS than comparison students who declined slightly over the year. Analyses by ability group showed no significant differences for the low ability students between the intervention group and the comparison group. Average ability students in the intervention group scored significantly higher on concepts than those in the comparison group. High ability students in the intervention group scored significantly higher on concepts and problem solving than high ability students in the comparison group. An ANCOVA on the IMA showed that students in the intervention group scored significantly higher than the comparison group. An analysis of the IMA showed that all students were using manipulatives to solve problems. High ability students used calculators twice as much as low ability students. Low achieving students guessed

	learning disabled	more often, said “I don’t know”, or just repeated back the numbers in the problem while average or high ability students were able to conceptualize problems and used more “if-then” logic. Overall, low achieving students in the comparison group stayed at 40% correct on the IMA while low achieving students in the intervention group showed modest improvement.

AAAS = American Association for the Advancement of Science
ELL = English language learners
FRL = Free or reduced lunch
NCTM = National Council of Teachers of Mathematics
NRC = National Research Council
NSES = National Science Education Standards
SES = Socioeconomic status

APPENDIX D
REVIEWED STUDIES OF STANDARDS-BASED INSTRUCTIONAL GUIDELINES (CHAPTER 4)

Study Source Standards	Standards-based (SB) Variable Methods Relevant to Research Synthesis Questions Measures & Samples	Findings
Adams (1999) National standards: NCTM (1989 and 1991)	SB Instructional Guidelines: District curriculum with NCTM-influenced content and processes. Study compared math achievement of students whose teachers implemented the curriculum and those who received the traditional curriculum Method: Quantitative quasi-experimental Measures: Student math ITBS scores Sample: Study participants came from one MS school districts; treatment teachers volunteered Sample students: 220 in treatment group, grades 1-7 454 in comparison group, grades 1-7	The analysis of covariance (ANCOVA) indicated that students who used the district-developed curriculum had a significantly higher average ITBS score than students who received the traditional curriculum.
Barth, Haycock, Jackson, Mora, Ruiz, Robinson, & Wilkins (1999) Technical report Standards: general and as expressed in state assessments	SB Instructional Guidelines: Use of standards and general instructional practice. The study described standards influences and instruction in high-performing, high-needs schools. Method: Quantitative non-experimental Measures: Principal survey Sample: The sample was drawn from nominated schools with at least 50% Free/Reduced Lunch (F/RL) enrollment that were in the top ten-ranked schools in achievement (absolute or improving) on the state math and/or reading assessment. Principals from 366 schools responded to the survey.	These schools use standards a great deal in general, 80% of principals report that they're used "extensively" to design curriculum and instruction and 94% say standards are used to assess student progress in their schools. Principals report that certain practices have changed over the past few years, although attribution to standards is not made explicit. 63% report a decrease in the use of ditto sheets, 83% report an increase in students discussing their work with others, 86% report an increased use of technology and most report an increase in time allocated to reading (86%) and math (66% - 77%) *Note: The study is not designed to establish whether these practices relate systematically to school achievement.*.

Blank, Kim, & Smithson (2000) Evaluation Technical report National standards: NCTM, NSES,	SB Instructional Guidelines: Participation in NSF's Urban Systemic Initiatives (USIs). The study describes math and science instruction across 4 USI sites with school-level variations in the level of reform implementation. Method: Quantitative non-experimental Measures: Teacher surveys Sample: Responses were gathered from 144 elementary and middle school teachers in 58 schools. 71 math teachers and 73 science teachers	In higher implementation schools, elementary students spend less time completing computational problems or solving word problems from a text or worksheet than their counterparts in lower-implementation sites. At both levels, math students also spend more time analyzing data than at lower-implementing schools. They are also more likely to be asked to justify or explain their answers and to do performance tasks significantly more than their counterparts. In science, the frequencies are similar across high and low-implementation schools. Middle school science students in low-implementation schools experienced more emphasis on memorization and analyzing information than in high-implementation schools.
Brenner, Mayer, Moseley, Brar, Duran, Reed & Webb (1997) Journal article National standards: NCTM	SB Instructional Guidelines: 20 day instructional unit based on NCTM recommendations for multiple representations of concepts, meaningful problem-solving and guided discovery. Statistical analysis of participation in the unit and pre-algebra achievement (sub-analysis conducted for English-Language Learners (ELLs)) Method: Quantitative quasi-experimental Measures: Researcher-developed assessment of functions and relationships with sub-domains: function word problems, word problem representations, word problem-solving, and equation-solving Sample: Sampling was based on three teachers in three junior high schools. Each taught two sections of pre-algebra. One section for each teacher was randomly assigned to the treatment group; the other used the traditional textbook. Participating students included 128 seventh and eighth graders (72 in the treatment group, 56 in the comparison group). Of the 128 students, 35 were second-language learners (19 in the treatment group, 16 in the comparison group).	ANCOVAs indicated that students in the treatment group did significantly better on representing and solving a function word problem than the comparison group., They were also better at problem representation tasks like translating word problems into tables and graphs than comparison group students. The achievement patterns were similar for students who spoke English as a Second Language. On the equation-solving section of the text, the treatment group did significantly worse than the comparison group, indicating the differences in learning outcomes produced by the treatment. If accuracy in solving equations is the goal of instruction, then conventional methods of instruction appear to be more effective than methods emphasizing multiple representations.
Bridge, Compton-Hall, & Cantrell (1997)	SB Instructional Guidelines: Kentucky Educational Reform Act (KERA) of 1990 1995 replication of 1982 study of writing	In 1995, teachers dedicated much more total time to writing than in 1982. They also focused on much higher-level activities. Teachers described KERA as the impetus for

Journal article State standards: KERA – Kentucky Learning Goals and Academic Expectations	instruction in Kentucky. Method: Mixed methods (Quantitative non-experimental) Measures: teacher surveys, teacher interviews, classroom observations Sample: The 1982 study had focused on two elementary schools in one KY district; in 1995, researchers observed 6 classrooms in each school at grades 1, 3, and 4. Classrooms were identified as typical by the principal. 206 teacher surveys gathered from a mailing to all K-5 teachers in the district	the change in instruction, with third and fourth grade teachers indicating the most changes. Fourth-grade teachers reported mandated writing portfolios, assessment requirements for open-ended writing responses and assessments focusing on on-demand-compositions as the greatest influences on their writing instruction.
Burian-Fitzgerald, McGrath & Plisko (2003) Journal Article National standards: NCTM State: assessment and curriculum guidelines District: curriculum guidelines	SB Instructional Guidelines: NCTM, state education department guidelines, assessment specifications and district and school curriculum guidelines. Secondary analysis of TIMSS and TIMSS-R teacher survey data; examination of changes between 1995 and 1999 in instruction Method: Quantitative non-experimental Measures: Teacher surveys Sample: The study draws on data from the 1995 and 1999 TIMSS surveys which uses a national probability sample designed to generalize to the entire United States.	Awareness grew about state assessment specifications and curriculum guides; more students in 1999 had teachers who were familiar with them than in 1995. Teachers who had more familiarity with standards were more likely to ask their students to do problem solving activities. There were no differences in the frequency of practicing computation between teachers who were more familiar with state and national standards and teachers who were not. Overall, teachers who were very familiar with standards were more likely than their peers to use practices consistent with current state curriculum guides and state assessments.
Butty (2001) Journal Article National standards: NCTM	SB Instructional Guidelines: Reform-oriented instruction Secondary analysis of NELS 88 data on student attitudes toward math, exposure to reform-oriented instruction and achievement. Method: Quantitative non-experimental Measures: Student IRT math scores in grades 10 and 12, teacher instruction relative to reform and traditional practice from NELS teacher survey Sample: A sub-sample of 190 African-American and 174 Hispanic American students in the NELS 88 database was used.	One-way ANOVAs indicated that 12^{th} grade students who received reform instruction had significantly higher achievement scores than those who received traditional instruction. This was not the case for 10^{th} grade students. 10^{th} grade students who had better attitudes toward math had significantly higher achievement scores than those who didn't. They also had significantly higher achievement in 12^{th} grade. There were no differences in attitude and achievement for 12^{th} grade students. Student attitudes toward math were unaffected by whether they received traditional or reform-oriented instructional practices.
Erickson & Niess (1996) Journal article	SB Instructional Guidelines: Teacher instructional strategies consistent with NCTM recommendations	Teachers reporting more types of student activities had classrooms with significantly higher overall math achievement and higher achievement on the application and problem-

National standards: NCTM	The study analyzed correlations between instruction and student achievement. Method: Quantitative non-experimental Measures: Researcher-developed math assessment with computation, application and problem-solving subscales; teacher surveys and lesson planning books Sample: 17 volunteer middle school math teachers who had participated in NCTM-oriented professional development prior to the school year. Each teacher randomly chose one 7th grade math class to focus on for planning books and questionnaires.	solving subscales. Instructional activities that correlated significantly with achievement on the application and problem-solving subscales were instruction to foreshadow coming developments and maintain student skills and organizing students in small groups, particularly in pairs. Use of manipulatives significantly and positively correlated with application and problem-solving scores, as well as the total math achievement score. The same was true for the weeks per year that teachers reported spending on geometry. Computation and problem-solving had an inverse relationship. Classrooms that spent more time per week on problem solving had a significant negative correlation with computation achievement. Similarly, teachers who spent more weeks per year on numbers and computation saw a significant negative correlation with student problem-solving scores.
Flexer, Cumbo, Borko, Mayfield, & Marion (1995) Technical report National standards: NCTM	SB Instructional Guidelines: NCTM standards, professional development in instruction-embedded assessment Description of changes in teachers' behaviors and beliefs over the course of a one-year professional development workshop Method: Qualitative Measures: Teacher interviews and data from biweekly workshops Sample: A purposive sample of 6 third grade teachers from 3 suburban elementary schools, representing a range of assessment and instructional practice and comfort with math and math teaching. Teachers were selected from a large sample of professional development participants.	Over the course of the project, teachers began to see changes in their practice and assessment more in line with the influence of the NCTM standards. Although the process was difficult for them, the changes they saw in their students reinforced their support for the reforms and they felt they knew more about their students at the end of the year. Researchers noted changes in teachers' beliefs about students' capabilities for learning math to a more developmental approach. They also noted changes in beliefs about important skills to teach, with problem-solving gaining more prominence over the year. Instructional practice also changed to a more student-centered approach over the course of the project. The researchers noted the difficulty of replicating such extensive changes across a large number of teachers due to the intensive needs for professional development.
Ginsburg-Block & Fantuzzo (1998) National standards: NCTM	SB Instructional Guidelines: Math instruction focusing on problem-solving (PS), peer collaboration (PC), or a combination of problem-solving and peer collaboration strategies	Two-way ANCOVAs showed that students who received instruction in problem-solving (PS) correctly computed a significantly higher rate of computations per minute than did students in the non-PS conditions. Similarly,

	Study compared achievement of students in 3 different treatment conditions with that of students who received traditional individual math instruction. (Also examined the effects on motivation and perceived math competence) Method: Quantitative experimental Measures: Rate of correct computations and rate of correct word problems completed by students on researcher/teacher-developed assessments. Sample: The sample consisted of 104 low-achieving 3rd and 4th grade children in an urban elementary school. Students were ages 8 – 12 and were all performing in the lower half of their grade on prior computation tests. Student ethnicity was 68% black, 18% Asian, 11% white, 3% Hispanic	students who participated in the peer-collaboration (PC) conditions also correctly computed a higher rate of computations per minute than students in the non-PC conditions. Significant effects were also found for the PS and PC groups on the rate of accurate word problems. These treatments also showed positive main effects on academic motivation and perceived competence in math. There were no significant effects for the combined condition on math achievement or affective variables.
Grant & Kline (2000) Conference presentation District standards, based on NCTM	SB Instructional Guidelines: Standards-based K-2 math curriculum (*Investigations in Number, Data and Space)*, based on NCTM Case studies of three teachers implementing the curriculum Method: Qualitative Measures: Reviews of teacher written reflections, classroom observations, teacher interviews Sample: The study examined three volunteer teachers (two at grade 1, one at grade 2), selected from professional development participants.	At the beginning of the year, teacher comments tended to focus more on student skills, moving toward "big ideas" by the end of the year. Specific challenges for all teachers included how to elicit and engage with student ideas and how to use incorrect answers as opportunities to learn for students. All teachers became better at dealing with these challenges, but progressed at different rates, depending on where they started and the synchronicity of their own beliefs and goals with the reform.
Hamilton, McCaffrey, Stecher, Klein, Robyn & Bugliari (2003) Journal article National standards: NCTM, NSES	SB Instructional Guidelines: Teacher use of reform and traditional teaching practices Large-scale statistical analysis of math and science instruction and student achievement Method: Quantitative non-experimental Measures: Student math and science scores on a variety of standardized and researcher-developed measures; teacher surveys Sample: Students and teachers in elementary and middle schools selected from 11 NSF-funded systemic change sites. Exemplary sites and comparison sites matched on demographics were chosen.	The pooled regression analyses indicate considerable in-school variation in the extent to which teachers use reform or traditional instruction. Use of reform strategies has a small positive effect on achievement across content areas and different types of test format, but this effect is not often statistically significant. The relationship is much smaller than any student background characteristic effects on achievement. The small size of the relationship varied from site to site, leading the researchers to note that longitudinal studies may identify larger instructional effects on achievement. The

	Teacher data was collected from 492 math teachers and 434 science teachers Student achievement data Math: 13,122 – 13,341 students (grades 3-7) Science: 14,164 – 14,249 students (grades 4-7).	study also notes pressing issues to improve the quality of large-scale education evaluations, including the need to validate surveys and to develop accurate appropriate measures of instructional practice, the lack of appropriate achievement measures (which can be aggregated across sites and which are also in line with NCTM standards) and the need to study within-school variation (rather than school-level data). The difficulties of linking individual student data with that of teachers over time were highlighted. Note: This study is an extension of the 2000 article by Klein, et al, using a larger sample of sites and including the data cited in the earlier report.
Hickey, Moore, & Pellegrino (2001) Journal article Evaluation National standards: NCTM	SB Instructional Guidelines: Participation in a videodisc-based math problem solving series (*JASPER*). The study analyzed the relationship between participation in JASPER, teacher orientation to math reform, SES and student achievement. Method: Quantitative quasi-experimental Measures: Student math ITBS scores at grades 3 and 5; district math coordinator/site resource teachers' identification of classroom reform orientation. Sample: The sample was drawn from one school district and consisted of 293 fifth-grade students in 19 classrooms in four schools (Two schools participated in JASPER and two did not, with one high-SES and one low-SES school in each treatment group). Classrooms were also rated as high and low reform orientation.	Four-way analyses of variance (ANOVAs) indicated that participants in the JASPER program had significantly higher increases in math achievement on problem-solving and data interpretation (PS/DI) tasks. A similar pattern was observed on the concepts/estimation subscale (CI), indicating the measure had an impact on student ability to solve moderately complex math problems – although math computation (C) declined slightly across groups. High-SES classrooms experienced positive CI achievement results from JASPER, but not low-SES classrooms, although there were no negative effects. PS/DI and CI achievement increased more in reform-oriented classrooms. There was an opposite pattern for C. Also, for low-SES classrooms, achievement on the CI scale decreased less when teachers were more reform-oriented. Students in JASPER classrooms more consistent with reform showed the largest gains in PS/DI, while students in less consistent, non-JASPER classrooms showed the only decline. A similar pattern occurred for CE, but none of these are statistically significant. However, for C, scores declined significantly in JASPER classrooms that were consistent with reform and increased in less consistent JASPER classrooms.
Ivey (1996)	SB Instructional Guidelines: NCTM standards	The class was divided into two distinct

Journal Article National standards: NCTM	Description of two different "cultures" operating simultaneously in one math class; one teacher-centered and the other student-driven. Method: Qualitative Measures: extensive classroom observations, document review and teacher and student interviews Sample: Sample consists of one teacher, teaching an 8th grade beginning algebra class. Data were also collected from 10 students who served as informants.	subcultures, depending on the type of task. The teacher interspersed some activities that conformed to the NCTM standards into his usual classroom format. He taught using a combination of direct instruction requiring book work where teacher/student roles were clear (mechanistic); however, the course also required attention to application problems where much of the work was student-driven and required a stretch. Students worked in groups and determined their own work processes (organic). Teacher and student responsibilities varied in the two cultures, with different acceptable behaviors. The study examined underlying philosophical and cultural changes required for implementing the standards in classrooms.
Klein, Hamilton, McCaffrey, Stecher, Robyn, & Burroughs (2000) Technical report National standards: NCTM, NSES	SB Instructional Guidelines: Teacher use of reform and traditional teaching practices Large-scale statistical analysis of math and science instruction and student achievement Method: Quantitative non-experimental Measures: Student math and science scores on a variety of standardized and researcher-developed measures; teacher surveys Sample: Students and teachers in elementary and middle schools selected from 6 NSF-funded systemic change sites. Exemplary sites and comparison sites matched on demographics were chosen. Teacher data was collected from 324 math teachers at 97 schools and 303 science teachers at 103 schools. Student achievement data Math: 9,995 – 10,196 students (grades 3-7) Science: 9,241 – 9,390 students (grades 5-7).	Pooled analyses based on localized regression coefficients indicated that teacher use of reform-oriented instructional strategies have weak positive effects associated with student achievement across content areas. These are rarely statistically significant. Similarly, traditional practices have a weak negative relationship to achievement. Both relationships are much smaller than any student background characteristic effects on achievement. *There are some indications that the relationship varies by test format (e.g., multiple-choice or open-response).*
Lane, Silver & Wang (1995) Conference presentation National standards: NCTM	SB Instructional Guidelines: Participation of schools in the QUASAR project, NCTM Quantitative analysis of student achievement gains over three years, disaggregated by ethnicity and student status as language learners. Method: Quantitative non-experimental Measure: QUASAR Assessment Instrument	At school A, white students made consistent gains from 1990 – 1994. Black students improved parallel to white students except for in the final year, which saw a dramatic drop-off. The researchers' explanation was that a new challenging algebra class was started that year, and that black students had unequal access to it (only 30% of black students got in, compared to 68% of the white students

	(QCAI) a set of open-ended tasks assessing mathematical understand, problem solving, reasoning and communication Sample: The sample was drawn from two urban middle schools and used 6th, 7th, and 8th grade math achievement data of all students. In the first school, data from 56 –67 students were used (30-33 students were Caucasian and 26-34 students were black) and at the second school 75-119 students provided data (42-56 English speakers and 33-63 bilingual students).	participating in previous years). At school B, bilingual students made parallel gains with English-speaking classes.
Lubienski (2002) Paper presentation National standards: NCTM	SB Instructional Guidelines: Teacher use of technology consistent with NCTM recommendations Study was a secondary analysis describing instruction and achievement trends over time, disaggregated by student ethnicity. Method: Quantitative non-experimental Measures: Student math scores on the National Assessment of Educational Progress (NAEP) (descriptive frequencies); teacher surveys Sample: NAEP national probability sample of grade 4, 8, and 12 students and their teachers: Sample students: 1990 – 8,072 students 1992, 1996 – approx. 21,000 students per year 2000 – more than 42,0000 students	Math achievement improved for all students between 1990 and 2000, but white students still outscore black and Hispanic students across grade levels. White 8th graders are allowed more access to calculators almost daily and are allowed more access for calculator use in tests than Hispanic and black students. Teachers of black students report computer use more for drill and practice than do teachers of Hispanic or White students. *Note: There were no analyses of whether achievement was directly linked to teacher instructional practice.*
Mayer (1997) Conference presentation Standards: NCTM-like	SB Instructional Guidelines: Teacher use of an active approach, as endorsed by NCTM, (teacher-led discussions and small group work) or a more passive approach (lecture and seatwork). Secondary analysis of data from the Longitudinal Survey of American Youth (LSAY) about teacher math instruction and student achievement Method: Quantitative non-experimental Measures: Unweighted NAEP gain scores on math items, developed in 1986, teacher surveys Sample: Sample used was a subsample of the Longitudinal Survey of American Youth (LSAY) national probability sample (1988-1989), focusing on junior high schools and	The HLM analysis indicated that teacher background and teaching style variables accounted for 42% of the residual variance in student achievement gains. The number of years of teaching experience was significantly related to gain scores in algebra and more active practices had a negative relationship with achievement. The more emphasis teachers placed on class discussion and small group work, relative to lecture and seatwork, the less students gained on the NAEP exams administered in the fall of their 8th grade year and the fall of their 9th grade years ($p<.10$) *Note: Because the study examines eighth grade students in algebra, findings are generalizeable primarily to high-ability, fast-track students.*

	particularly 8th grade students who are enrolled in algebra. The data used for the study is from 37 teachers in 34 schools, with student data from 325 students.	
Mayer (1998) Journal Article National standards: NCTM; filtered through district policy	SB Instructional Guidelines: District policies to adopt NCTM; teacher participation in professional development on using NCTM professional standards in the classroom. Statistical analysis of math instruction relative to NCTM recommendations and student algebra achievement Method: Quantitative non-experimental Measures: Student gain scores on 3 criterion-referenced, multiple-choice algebra tests, teacher surveys Sample: The study used a sample of all black and white 8th and 9th grade algebra students in a large suburban district who had the same math teacher through the 1995-1996 school year. The analytic sample used data from 2369 students (or 67% of the population) and 94 teachers in 41 middle schools and high schools.	The HLM analysis indicated that, at the middle school level, teachers spend much more time on NCTM tasks than HS teachers. Middle school students are also more advanced than high school students (taking algebra in 8th grade). At high school, there is no significant role between NCTM and student achievement At the middle school there's a positive effect of NCTM on achievement, but higher-achieving students are more likely to experience NCTM practices. The most talented students get the most NCTM practice. The effect is more pronounced for higher achieving students. More use of NCTM practices is not negatively related to student achievement on traditional measures.
Mayrowetz, (1999) Conference presentation State standards: New Jersey Core Content Standards National standards: NCTM, NRC Science standards on inclusion	SB Instructional Guidelines: Standards guidelines about inclusion of special needs students and recommendations about modification of instruction relative to standards. Descriptive study of teachers' math instruction practices, particularly related to the inclusion of special education students in standards-based instruction and extent to which standards-based tasks were modified for special needs students. Method: Qualitative Measures: Teacher interviews and classroom observations Sample: Purposive sample of 6 4th grade teachers of math to special needs students. Teacher responses on initial surveys suggested their teaching reflected NCTM recommendations.	About half of all math tasks observed were of brief duration and called on student procedural knowledge (e.g., memorization of formulas) rather than principled knowledge (e.g., understanding of underlying concepts). In most of those tasks, teachers did not modify instruction for special education students. For longer-term, more challenging tasks, teachers are more likely to intervene but task modification was not a common strategy and separate instruction is also rare. Interventions varied by whether the problem-solving strategy was taught implicitly or explicitly and also by whether the teacher or the student implemented the strategy – although there were no instances of an implicit intervention that led to student performance of the task.
McCaffrey, Hamilton, Stecher, Klein, Bugliari, &	SB Instructional Guidelines: Teacher self-reported use of strategies from *Interactive mathematics* program and *College*	Regression analysis revealed a significant positive relationship between use of standards-based or reform practices and

Robyn (2001) Journal National standards: NCTM professional standards	*preparatory Mathematics* (both are SB integrated mathematics curricula funded by NSF's Urban Systemic Initiative program) Study compares classroom type (integrated vs. traditional course sequence) and teacher self - reported use of reform-based instructional practices on student achievement Method: Quantitative non-experimental. Measure: Stanford-9 achievement test multiple choice and open ended questions in math used as student achievement measure. Teacher survey used to measure use of SB instruction. Sample: Students self-selected into either traditional math course sequence or one of the SB integrated math curricula in 26 high schools in a large urban school district: Surveyed 182 teachers Multiple choice data from 4,799 10th grade students and open ended data from 4,709 students.	achievement on student achievement for students taking integrated math courses, but not for students taking more traditional algebra and geometry courses. The authors suggest that changes in instructional practice may need to be coupled to changes in curriculum/course type (integrated vs. traditional course sequence) to impact student achievement. *Note: Possible differences could exist between the two groups that could pose alternate explanations for the results.*
McGinnis & Parker (2000) Conference presentation National standards: NCTM and AAAS Benchmarks	SB Instructional Guidelines: Teacher education program (MCTP) supported by NSF and based on math and science standards. Descriptive study of teacher survey data from new teachers, compared to national pool of teachers and combined with follow-up data after two years of teaching Method: Mixed methods (non-experimental) Measures: Teacher surveys, teacher interviews, observations Sample: 57 new teacher graduates for survey data, compared with NSF national sample; follow-up interview and observation data from 3 volunteer teachers.	Initially, new teachers expressed beliefs and actions aligned with their preparation program's emphasis on standards. Their responses indicated more likelihood than the comparison group of NSF teachers to: assist all students to achieve high standards, provide examples of high-standard work, use authentic assessments, use standards-aligned materials and make connections between math and science. Teachers' school environments provided varying levels of support for these practices and beliefs. In instances where the school culture provided support, teachers flourished. In others, they found various coping or exit strategies.
Norman, Stein, Moussiaux, & Clay-Chambers (1998) Conference presentation National standards: NCTM, NSES	SB Instructional Guidelines: School participation in the Urban Systemic Initiative (USI) reforms Longitudinal examination of reform-oriented math instruction in schools participating in the Detroit USI, by tier (year) of implementation. Method: Quantitative non-experimental Measures: Teacher surveys, student surveys	Substantive changes in instruction take time. Teachers who participated longer in the USI reported using constructivist practices more frequently than teachers who were newer (Tier 1 teachers, compared with Tiers 2 and 3). *Note: The lack of a comparison group precludes attribution of changes to the USI.*

	Sample: At the school level, 18 schools (10 elementary, 5 middle level, and 3 high schools) were randomly selected for each of three tiers of USI implementation. All math and science teachers in those schools were administered surveys. Data were gathered from 570 teachers and from 1080 students (randomly sampled from two classrooms per school). Student surveys were used to validate teacher surveys.	
Rizor (2000) Doctoral dissertation State and National standards: Wyoming, based on NCTM	SB Instructional Guidelines: WY and NCTM standards, teacher reports on lesson influences, teacher math instruction The study gathered data about math teaching in WY and correlated reform practice with achievement. Method: Quantitative non-experimental Measures: Mean school Wyoming Comprehensive Assessment System (WyCAS) scores, teacher surveys Sample: The study used a random sample of 36 WY elementary schools, stratified by size (large, medium, or small). From those schools, all 250 teachers assigned to grades 2 –4 were sent surveys. 177 teachers returned usable surveys to determine math practices.	Adopted textbooks and standards most often influenced the content of lessons taught. The WyCAS had more influence on content taught for teachers at grade 4 than at grades 2 and 3. The overall instructional reform scale developed did not significantly correlate with WyCAS scores, although particular items did. Items that correlated significantly with WyCAS mean scaled scores were: total hours of inservice provided for 4th grade teachers to implement math standards, hours per week spent on writing, hours per week spent teaching science, and other influences on content (such as district testing or professional decision making) *Note: It is unclear whether the WyCAS scores used in this study were for math or whether they were composite math/reading scores.*
Saxe, Gearhart, Selzer, (2000) Journal National standards: NCTM.	SB Instructional Guidelines: NCTM-oriented instruction Subject teachers were either engaged in math reform curriculums (*Seeing Fractions* or *My Travels with Gulliver)* or not engaged in reform. Their instructional practices were rated on the extent to which they aligned with NCTM principles, and the relationship of their instruction to students learning of fractions was examined. The impact of student's initial proficiency in fractions was also examined. Method: Quantitative non-experimental Measure: Teacher instruction was measured through classroom observation. Student achievement was measured through a pre-post assessment. Sample: 19 upper elementary classroom teachers and their 481 students	Hierarchical Linear Modeling (HLM) suggests that alignment of classroom practices with NCTM reform principles was related to student achievement in problem solving but not computation. For the 313 students who began with rudimentary pre-test understanding of fractions this relation was linear. For the 168 students without a basic pre-understanding of fractions: performance was very low when alignment of teaching to NCTM principles was below the mean. However when the level of alignment of classroom practice with NCTM principles passed the mean their performance became linear.

Shymansky, Yore, Anderson (2000) Conference Presentation National standards: NRC, NBPTS, NSES	SB Instructional Guidelines: ratings of teachers use of strategies from *Science: Parents, Activities, and Literature (PALS):* Local districts' systemic reform The study examined the correlation between the science coordinator's ratings of science teachers' use of the SB Curriculum's strategies and student achievement. Method: Mixed methods (non- experimental) Measures: Ratings of teacher pedagogy by science coordinator; Student achievement assessed with multiple choice items taken from TIMSS; student attitudes towards science Sample: 16 elementary schools participating in the PALS program for 3 years. 235 3rd and 4th grade science teachers	In an ANOVA, the ratings of 3rd and 4th grade students' past three science teacher's use of PALS strategies did not show a significant relationship with their achievement or attitudes towards science. *Note: it was not specified how often the science coordinator observed the classrooms or what he/she based the teacher ratings on.*
Silver & Lane (1995) Journal Article National standards: NCTM	SB Instructional Guidelines: Participation of schools in the QUASAR project, NCTM Item-by-item comparison of student achievement between students in QUASAR schools and the NAEP national and disadvantaged urban samples Method: Quantitative non-experimental Measures: 1992 NAEP released items in math Sample: Between 275-314 eighth-grade students in 5 QUASAR schools. QUASAR schools were described as urban and diverse. NAEP samples included 1759-1763 students for the nationally representative sample and 191-204 students in the nationally representative urban disadvantaged sample.	QUASAR students were found to do significantly better on some math items than the NAEP disadvantaged urban sample and as well on the remainder of the items. The QUASAR sample did not do as well on most items as the national NAEP sample. Sub-analyses indicated that these patterns continued across math content areas and ability categories (e.g., procedural knowledge, conceptual understanding and problem-solving_ *Note: The article lacks a comprehensive description of the QUASAR sample to establish its comparability with the NAEP urban disadvantaged sample. It also lacks clear descriptions of the statistical analyses used to establish significance.*
Smith (2000) Journal article National standards: NCTM	SB Instructional Guidelines: A middle school math curriculum *(Visual Mathematics)* that emphasizes problem solving, reasoning, communication, connections and constructivism. Case study of teacher reflections and changing math practice over the course of a year of professional development related to NCTM standards Method: Qualitative Measures: teacher journal entries, teacher interviews, videotapes and summaries of	The changes inherent in the standards led to cognitive dissonance for the teacher, presenting her with new ideas about what students needed to know and be able to do in math. These experienced posed dilemmas for her in terms of how to resolve them with her long-held beliefs. For instance, her new conception of teaching – that students should become competent problem-solvers – conflicted with a long-held belief that students need to feel successful. Confronting these dilemmas allowed the teacher to learn and her practice to change substantively. Implications for teacher preparation include finding ways

	teacher reflections on teaching and change Sample: One middle school math teacher	to help teachers build their capacity for supporting students to struggle without reducing the complexity of the tasks students are asked to do.
Snow-Renner (2000) Dissertation National standards: NCTM	SB Instructional Guidelines: NCTM standards Secondary analysis of Colorado TIMMS data, examining relations between elementary teacher instruction in math and student math achievement. Method: Quantitative non-experimental Measures: Teacher surveys, student math achievement on TIMSS Sample: 104 third and fourth grade Colorado teachers and 2,163 students. Data were from the statewide representative sample of TIMSS.	Elementary teachers predominantly focused on whole numbers and measurement topics, spending shorter amounts of time on higher-order content. Instructionally, teachers used a combined repertoire of reform and traditional practice. Correlations between reform-oriented instruction and student achievement were inconsistent and small in size, dependent on the particular domain of math being assessed and the student's grade level. The researcher highlighted analysis issues around the nature of the achievement measure and the level at which analyses are to be conducted.
Spillane & Jennings (1997) Journal article District standards, focusing on reading comprehension	SB Instructional Guidelines: District presence of aligned policies and a curriculum for ambitious reading/language arts learning. Case study describing teacher instruction in response to standards-based policies Method: Qualitative Measures: Teacher interviews, classroom observations, document review. Sample: Nine teachers were sampled from three schools in a MI district participating in a national reform study. Schools were identified by local reformers as representing a range of changes in instruction – two that had managed to significantly change instruction and one that showed only modest change. Teachers were nominated by district administrators as also representing the range of teaching in the school.	In interviews, teachers reported more emphasis on literature and the writing process than in the past, rather than using basals and lower-level instruction. They also talked about reading and standards in relation to their instruction. However, observations indicated considerable variation in instruction beneath the similarities – both in the nature of classroom tasks and in the level of discourse engaged in by students. Some classrooms emphasized more substantive comments from students, more in line with standards reforms, which others provided surface alignment, but little substantive change from traditional teacher-centered instruction. Alignment of policies was insufficient to help teachers understand their roles, particularly as teachers bring their own experience as learners to the role. The authors highlight the importance of policy as an educative tool for teachers, with examples of instructional practice.
Spillane & Zeuli (1999) Journal Article National standards:	SB Instructional Guidelines: NCTM standards Analysis and description of variations in math instruction by a group of teachers identified through survey responses as reform-oriented teachers.	All teachers reported knowledge and implementation of standards on the survey measure but their actual classroom practices looked very different from one another. Of the 25 teachers who indicated reform practice, only 4 taught in ways consistent with the

NCTM	Method: Qualitative Measures: teacher surveys to determine initial sample, classroom observations Sample: Teachers were selected for observation based on their responses on TIMSS instructional practice items relative to reform. The sample consisted of 25 teachers of 3rd/4th grade and 7th/8th grade students in MI districts, stratified across district type and geography, who were highly oriented to math reform teaching.	intent of the reforms, when observed. The following three main patterns of instruction were observed; conceptually grounded tasks and conceptually centered discourse (closest to NCTM) – which focused on students justifying answers and explaining their reasoning conceptually-oriented tasks and procedure-bounded discourse. – where questions focused on students getting a right answer, and peripheral changes, with continuity at the substantive core of instruction – in which, for example, manipulatives were used, but primarily to uncover procedural understanding.
Swanson & Stevenson (2002) Journal Article National standards: NCTM	SB Instructional Guidelines: State standards activism and coherence of policy Secondary analysis of NAEP data to develop a model of state, school, and classroom level predictors of instructional practice Method: Quantitative non-experimental Measures: Chief State School Officer (CCSSO) indicators of state policy, NAEP teacher surveys of instruction Sample: Two samples were drawn from states that administered the NAEP public school 8th grade math assessment in 1992 and 1996; a panel sample (of 30 states) and a longitudinal subsample of 20 states.	A preliminary HLM model estimating the amount of variation in outcomes by classroom, school, and state level, indicates that 3/4 of the variance in instructional practice is attributable to the classroom level and less than 3% is explained by state-level variables. Using the longitudinal sample, the researchers found that state level of standards-based activism had a significant positive effect on the use of SB instructional practices (fully standardized beta effect of .09) This is the case even after controlling for prior instructional norms. When teacher knowledge and attitudes were included (receptivity indicators) in the model, state level policy influence decreases. Training, mediated through teacher receptivity factors, shows the greatest amount of change on instruction. Standards policy is more likely to promote change by promoting greater teacher knowledge about and receptivity towards SB education – primarily through professional development. There's a modest but robust policy effect for state policy.
Swierzbin, Liu & Thurlow (2000) Technical report State standards: MN High	SB Instructional Guidelines: MN High Standards (across content domains) Descriptive analysis of data from MN teachers of Limited-English Proficient (LEP) students about their participation in teaching MN High Standards in their schools.	MN High Standards are not being implemented for LEP students in their ESL and Bilingual Education classes. Although about 1/5 of the teachers reported having their students do content-based work toward completing a specific standard, ESL teachers more often act as a resource for general

Standards	Method: Mixed methods (non-experimental) Measures: Teacher surveys, interview data Sample: 57 ESL/Bilingual Education teachers were surveyed from 6 high ESL population districts in MN ; 22 surveys were returned from 10 teachers at elementary, 3 at middle school, 6 at high school and 3 across levels. Interviewees were from a volunteer sample of the same set, with 2 elementary teachers, 1 middle school teacher and 2 high school teachers.	education teachers, teaching in the general education classrooms. Many teachers commented on the lack of coordination between general education teachers and ESL/Bilingual Education teachers. ESL/Bilingual teachers are not included in the standards conversations. Teachers indicate that few, if any, of their LEP students participate in High Standards work in their own languages. *Note: Very small sample size and low response rate in study*
Turner (1998) Conference presentation National standards: NCTM	SB Instructional Guidelines: NCTM, Pacesetter Mathematics course and 8-day summer professional development Descriptive analysis of instruction relative to 4 NCTM practices Method: Quantitative non-experimental Measure: teacher surveys Sample: A convenience sample was drawn from 10 – 12 teachers participating in the Pacesetter Math program.	Respondents reported a variety of math practices and strategies, although they did not align with the goals of NCTM. In general, it did not appear that teachers were implementing the program.
U.S. Department of Education (2001) (Volume 1) (Volume II) Technical report National standards: NCTM	SB Instructional Guidelines: Teacher instruction in math and reading designed to engage students in advanced tasks, standards-based policy frameworks in the school. Analysis of the relationship between teacher instruction in Title I Schools and student achievement, 1996-1999 Method: Quantitative non-experimental Measures: SAT-9 reading and math scores (grades 3, 4, and 5) Sample: Students and teachers in 71 elementary schools receiving Title I funding; 56% of schools had more than 75% of students qualifying for F/RL. No school had fewer than 35% of its students qualifying for F/RL.	One factor indicated by the HLM analysis that affected math achievement gains was teacher use of more exploration activities at grade 5 (e.g., use of manipulatives, discussion of multiple problem-solving approaches, small-group work, whole-class discussions, student-led whole-group discussions, use of tables and graphs, writing assignments of at least a paragraph, and assignments taking a week or more to complete). In reading, students whose fifth grade teachers reported spending more time on basic instruction (use of worksheets, reading aloud, completing workbooks, practicing phonics and word attack) gained 1.9 points less on average than those whose teachers reported an average amount of time working on basic skills. Teacher reports about the use of standards and assessments have inconsistent effects on student achievement. Other findings: Achievement in math and reading improved faster when teachers gave high ratings to their professional development

		in that particular content.
Watson (1996) Doctoral dissertation National standards: NCTM	SB Instructional Guidelines: Participation in SCRAP ((Southern California Regional Algebra Project) – an NCTM-influenced alternative algebra class The study compared the math achievement of SCRAP participants and students who took a regular algebra course. Method: Quantitative experimental Measures: Grades in math classes after 9th grade, PSAT math and verbal scores Sample: The sample was drawn from 464 high school students in one relatively high-achieving high school. 195 students were randomly assigned to the SCRAP section of algebra in grade 9, with 269 assigned to regular algebra. School ethnicity: 76.5% white 9.4% Latino 8.7% Asian, 3% black 1.5% Filipino, 0.7% Pacific Islander 2% Native American	Participation in the SCRAP program was not substantially better for student math achievement, and in some cases, the SCRAP group had lower achievement scores. Members of the comparison group had significantly higher grades in algebra II, enrolled in more semesters of higher math and had more successful semesters of higher math. Scores disaggregated by ethnicity indicated that black students in SCRAP got significantly higher grades in geometry than did their counterparts in the comparison group. However, white students in the comparison group earned significantly higher grades in geometry and algebra II than their SCRAP counterparts. *Note: Indications of initial differences between SCRAP/comparison groups not explored.*

AAAS = American Association for the Advancement of Science
ELL = English language learners
FRL = Free or reduced lunch
NCTM = National Council of Teachers of Mathematics
NRC = National Research Council
NSES = National Science Education Standards
SES = Socioeconomic status

APPENDIX E
REVIEWED STUDIES OF STANDARDS-BASED ASSESSMENT (CHAPTER 5)

Study Source Standards	Standards-based (SB) Variable(s) Methods Relevant to Research Synthesis Question Measures & Sample	Relevant Findings
Avery, Beach, & Coler (2002) Technical report State standards: MN	SB Assessment: The two components of the MN state test (one of basic skills and the other of interdisciplinary, higher-order skills) Study examined teachers' perceptions of the effects of state mandated testing on instruction. Method: Mixed methods (non-experimental) Measures: Teacher surveys (frequency analyses) and interviews (transcribed responses) Sample: Study participants were chosen to represent English/language arts and social studies teachers; 45% rural, 44% suburban, 11% urban Surveys: 171 English/language arts teachers 487 secondary social studies teachers Follow-up interview: 51 English teachers and 89 social studies teachers	Surveys: In response to SB testing, a strong majority of responding teachers reported an increased need for planning time, about half reported no change in their use of nontraditional assessments, most reported no change in their use of computer technology, and about half reported no change in their use of different teaching approaches. Interviews: Selected responses were used to support survey results.
Barksdale-Ladd & Thomas (2000) Journal article State standards: two unidentified states	SB Assessment: State-level (unidentified) Study examined teachers' perceptions of the effects of state mandated testing on instruction. Method: Qualitative Measures: Teacher interviews and focus groups Sample: 35 teachers from a large southern state 24 teachers from a large central/northern state Study participants were chosen to represent reading, writing, and language arts teachers	A majority of those interviewed indicated that they perceived test preparation activities that as being unpleasant for students, that they believed these activities would not encourage depth of understanding, and that they believed neglected student collaboration and the development of social skills. Classroom activities that were employed less frequently in response to SB testing included silent reading, partner reading and writing, science experiments, field trips, cooking, drama, and other creative activities.

	for grades 1 through 8.	
Beran (2003) Dissertation State standards: NE	SB Assessment: Nebraska Leading Educational Achievement through Rigorous Standards (LEARNS) Study examined teachers' perceptions of the effects of state standards (with emphasis on mandated testing) on instruction. Method: Quantitative non-experimental Measures: Survey of 4th-grade teachers (4th grade is one of the three mandated testing years) Sample: Study participants were chosen to represent teachers statewide; 257 4th grade teachers surveyed, roughly representative of state in experience, gender and school SES percentages.	Teachers reported that LEARNS resulted in an increased workload. The authors provided survey evidence that this increased workload had increased levels of self-reported teacher stress and decreased teacher morale. The surveyed teachers also reported that the standards process had resulted in a narrowing of the 4th-grade curriculum. *Note: State standards are minimal competency standards while local standards that exceed them are encouraged (this plan is referred to as STARS - School-based Teacher-led Assessment and Reporting System)*
Clarke, Shore, Rhodes, Abrams, Miao, & Li (2003) Technical report State standards: KS, MI, MA	SB Assessment: State assessments in KS, MI, MA The study examined the effects of SB reform on teaching and learning with emphasis on the effects produced by SB testing. Methodss: Qualitative Measures: Interviews Sample: 120 school employees were interviewed in each state (sample chosen to represent variety in each state with respect to district setting, grade level, and tested and non-tested subject area). The states were chosen to represent variety in the levels of stakes attached to test results: KS: low stakes for students, high stakes for teachers MI: moderate stakes for students, high stakes for teachers MA: high stakes for both students and teachers	Most of the responding teachers reported that they had changed and altered curricular planning to align with state tests. No significant difference between states was reported. Most of the responding teachers reported changing instructional and classroom assessment approaches to align with state tests (e.g. emphasizing writing activities, encouraging critical-thinking skills, and expecting students to explain processes and results). These same teachers also reported perceived negative effects of the state tests (e.g. fewer activities that promoted depth and creativity). Elementary teachers reported significant changes in their own practice to a greater degree and frequency than did their middle- and high school counterparts. Teachers in rural schools reported more curricular changes made in response to state tests.
Daniels (1995) Conference presentation State standards: VT	SB Assessment: VAP, which assesses writing skills through two components: a pencil-and-paper test and student portfolios Study examined how portfolios are used in instruction and classroom assessment and what effects the Vermont Assessment Program (VAP) has on instruction in the first	The VAP reform encouraged teachers to plan for increased student writing time to better align their instruction with VAP expectations. The teachers also experienced a greater amount of paperwork as a result of the VAP in efforts to document student progress and provide helpful feedback. Much of the classroom adaptation taking

	year of implementation Method: Qualitative Measures: Classroom observations, case interviews, principal interview, colleague interview, and district teaching support staff interview Sample: The study focused on one 5th grade teacher over a seven month period in one school year.	place was a response by teachers to the expectations of the VAP portfolio component.
Din (1996) Conference presentation State standards: KY	SB Assessment: The Kentucky Educational Reform Act (KERA) included mandated assessments that were aligned with state reform. Study examined teachers' perceptions of the effects of state mandated testing on instruction. Method: Quantitative non-experimental Measures: Teacher survey Sample: 350 teachers interviewed; 43% from elementary, 29% middle school, and 29% high school; schools were 57% non-urban, 43% urban	Surveys: A majority of the teachers surveyed indicated that KERA substantially influenced choices of curricular content and objectives, instructional approaches, and classroom assessment practices.
Doran (2001) Dissertation State standards: AZ	SB Assessment: Arizona Instrument to Measure Standards (AIMS) The study examined teachers' perceptions of the effects of state mandated testing on instruction in tested and non-tested grades. Method: Quantitative quasi-experimental Measures: Teacher survey grades 1 through 5 (experimental group comprised of teachers of grades 3 and 5 which are mandated testing grades) Sample: Sample drawn from one metropolitan district (chosen due to diversity - 5 of the 7 target elementary schools received Title I support - and large population); 153 teachers surveyed (evenly distributed among grade levels)	Based on an ANOVA of Likert-scale responses, the teachers in tested grades indicated that they used test scores to guide their curricular choices in significantly greater numbers and/or to a significantly greater degree than did teachers in non-tested grades. This trend was found in both Title I and non-Title I subgroups.
Falk & Drayton (2004) Journal article National	SB Assessment: MA The study examined science teachers' changes in practice under a statewide systemic initiative (SSI) to encourage higher test	The between-district responses to state testing suggested that the tests themselves did not encourage specific changes in practice. Instead, the responses were linked to the district interpretation of the state tests (for example, whether the district perceived the

standards: NRC State standards: MA	scores. <u>Method:</u> Qualitative <u>Measures:</u> Interviews, observation, and document review of science education in six schools, from six districts <u>Sample:</u> 40 teachers were interviewed and observed in six schools from 6 different districts. The school cases represented a variety of low- and middle-SES populations in urban and suburban settings.	tests dictated a significant change in pedagogy). District interpretation appeared to have a direct effect on the changes seen in classroom teaching practice. A consistent difference between middle- and low-SES classroom responses was observed. The middle-SES school teachers were more likely to increase the number of topics covered in a school year, while their low-SES school counterparts avoided this action for fear of compromising the quality of the learning that was taking place.
Fernandez (2004) Dissertation State standards: FL	<u>SB Assessment:</u> Florida Comprehensive Assessment Test (FCAT) The study examined the effect of state testing on teaching practice. <u>Method:</u> Qualitative <u>Measures:</u> Classroom observation, teacher interviews, and document review <u>Sample:</u> 5 4th grade teachers on a team in case study school, an urban, low-performing school with a 50% population of ELL students. Teachers were observed during 2 or 3 teaching sessions, before and after FCAT testing.	In the weeks prior to testing, the teachers chose to direct sessions in which students practiced FCAT-like prompts instead of what they viewed as more creative approaches to instruction (e.g. having students summarize readings or designing interdisciplinary projects). The teachers also worked to cover greater numbers of topics, an act which they perceived to be ineffective.
Firestone, Mayrowetz, & Fairman (1998) Journal article State standards: ME, MD	SB assessments: the Maine Educational Assessment (MEA) and the Maryland School Performance Assessment Program (MSPAP) The study examined the effect of the MEA and the MSPAP on middle-school math teaching practice (both testing programs include a test for 8th-grade math). <u>Method:</u> Qualitative <u>Measures:</u> Teacher and administrator interviews <u>Sample:</u> 14 8th grade math teachers in 6 schools in 3 ME school districts (FRL of 20%, 35%, and 40%) 11 8th grade math teachers in 4 schools in 2 MD districts (FRL of 20% and 45%)	Despite MEA and MSPAP influence that intends to encourage higher-order skills, the teachers in these states emphasized drill of relatively small problems and covered topics at superficial levels. The teachers made changes to curricular content and sequence, but did not change their classroom methodologies significantly.
Grant (2000) Journal article	<u>SB Assessment:</u> New York State tests The study examined how state-level tests and the results are influencing NY state teachers'	State testing influenced teachers to align their teaching to what they saw as test-practicing activities. For low-performing students, these

State standards: NY	practice (in English, math, science, and social studies). Method: Qualitative Measures: Focus group data from 4 groups (2 groups in year 1 and 2 groups in year 2) Sample: Year 1: 1 group of 7 elementary teachers and 1 group of 12 high school teachers (chosen to represent the state) Year 2: 1 group of 5 elementary teachers (3 were in year 1 sample) and 1 group of 8 high school teachers (5 were in year 1 sample)	approaches took the form of remedial drilling.
Grant (2001) Journal article State standards: NY	SB Assessment: New York State tests This study examined the influence of state-level testing on teaching practice. Method: Qualitative Measures: Classroom observation and teacher interviews Sample: The two case teachers were chosen because they were both identified as exemplary teachers and because their instructional approaches were markedly different.	State testing had very little impact on teaching practice, at least the practice of these two exemplary teachers. An effort to cover mandated curricula was an apparent effect of the advent of the NY testing, but the testing had little effect on the classroom approaches employed. This conclusion was supported by the widely diverse approaches to teaching employed by the two cases in this study.
Janson (2002) Dissertation State standards: OH	SB Assessment: Ohio Proficiency Tests (OPT) This study sought to determine the extent to which teachers had aligned classroom assessments with mandatory state tests. Method: Mixed methods (non-experimental) Measures: Interviews, document review, and surveys Sample: 5 3rd-grade science teachers and 5 5th-grade science teachers in one school district.	Document review: Classroom assessments were not aligned with state tests in terms of format, style, scientific process targeted, and curricular content. In general, classroom assessments did not match format and style characteristics of the OPT, and the items targeted lower-level thinking skills. Interviews and surveys: Teachers expressed their dependence on publisher-produced assessments for use in the classroom.
Khattri, Reeve, Kane, & Adamson (1995) Book Standards: various national, state, district, local	SB Assessments: various (AZ, KY, MD, NY, OR, and VE) This study examined the impact of mandated assessments on teaching and learning nationwide. Method: Qualitative Measures: Interviews and classroom	Teacher instruction: The authors reported that teachers responded to mandated testing by changing classroom curricula, and by instructing students to conduct research projects, practice writing skills, and to work in groups with increasing frequency. The authors also noted that these activities often fell short of the depth of learning that was expected as a product of the various state standards.

	observations Sample: 16 sites were visited in a first round, then 7 were revisited (sites chosen to represent schools nationwide with respect to source and type of mandatory testing, stage of testing program development, testing content areas, grade level(s), and geographical area)	Student achievement: Teachers reported that students were demonstrating improvements in the newly emphasized skills (math, writing, and reading).
Koretz, Mitchell, Barron, & Keith (1996) Technical report State standards: MD	SB Assessment: Maryland School Performance Assessment Program (MSPAP) The study examined the effect of state testing on teaching practice. Method: Mixed methods (non-experimental) Measures: Mail and telephone surveys of 5th-grade teachers and 8th-grade math teachers (both grades are testing grades) Sample: 226 teachers surveyed - approximately half were 5th grade teachers and the other half were 8th-grade math teachers (chosen to represent the grade/subject teachers statewide)	Teachers reported greater focus on improving practice in response to mandatory testing. For both groups of teachers this focus was manifested in curricular content changes that were seen as moving toward alignment with MSPAP. 5th-grade teachers devoted more classroom time to writing instruction while deemphasizing spelling, punctuation, and grammar. These teachers also reported an increased emphasis on higher-order skills such as problem solving, practice within applications, and communicating concepts in math. 8th-grade math teachers also reported an increased emphasis on higher-order skills.
Louisville University, School of Education. (1995). Technical report State standards: KY	SB Assessment: The Kentucky Educational Reform Act (KERA) mandated performance assessment Analysis of teacher survey data about the use of performance assessments in KY classrooms Method: Mixed methods (non-experimental) Measures: A researcher-developed teacher survey. Interview data gathered background information. Sample: A stratified random sample of 32 schools was drawn each of KY's 8 regions (2 elementary schools, 1 middle school and 1 high school from each region). All teachers of language arts, math, science, and social studies in these schools were surveyed and 6 teachers were randomly selected for interviews at each school, for a total of 192 teacher interviews. 500 surveys were returned.	Performance assessment is happening in KY classrooms, but there was considerable variation between those studied. KY teachers were primarily using performance assessment to prepare for the KERA tests rather than as an integral part of instruction. Teachers in tested grades made more use of performance assessments as an integrated part of their instruction.
McDonnell & Choisser (1997) Technical report	SB Assessments: Kentucky and North Carolina state tests (KY having a high-stakes test and NC having low stakes) The study examined the impact of mandated	Teachers demonstrated instructional approaches that mirrored the two different state tests. Teachers in North Carolina, for example, were more likely to employ multiple-choice items in their classroom

State standards: KY, NC	assessments on teaching practice. Method: Qualitative Measures: Teacher interviews and document review (lesson plans) Sample: 139 teachers interviewed in 1993 (chosen to represent the states in terms of geography, SES, and community setting) - 48 of the teachers were interviewed again in 1994 - KY teachers from elementary, middle, and high schools - NC teachers from grades 3 through 8	assessments (consistent with the state test format). The Kentucky teachers, on the other hand, were more likely to require extended written responses, an activity that was consistent with their state test format.
McMillan, Myran, & Workman (1999) Conference presentation State standards: VA	SB Assessment: Virginia Standards of Learning (SOL) The study examined the impact of mandated assessments on teaching practice. Method: Mixed methods (non-experimental) Measures: Survey (administered to each teacher twice: before and after the implementation of the state testing program); open-ended questions Sample: - 152 elementary teachers (grades 3 through 5) - 570 secondary teachers (in all tested subjects)	Quantitative analysis: Based on *t*-tests, elementary teachers reported that the SOL influenced them to rely less on lecture, whole-class discussion, and seatwork in teaching math and language arts lessons. What they chose instead was not reported. Secondary teachers reported less frequent use of lecture and small-group instruction. Qualitative analysis: Both groups of teachers reported altering the content coverage to mirror the curricular profile of the SOL. Elementary teachers reported an increased breadth of topical coverage and an increased use of multiple-choice assessments in the classroom also in response to the SOL. Secondary teachers also reported an increase in topical breadth.
Moon, Brighton, & Callahan (2002) Journal article State standards: states nationwide	SB Assessment for survey: state tests nationwide SB assessment for focus groups: mandatory state testing in California, Texas, and Virginia The study examined the impact of mandated assessments on teaching practice (with emphasis on the teaching of gifted students). Method: Mixed methods (non-experimental) Measures: Survey of teachers, focus groups (3 to 5 teachers guided by script), and classroom observations (of focus group teachers' classrooms) Sample: 1289 kindergarten through 5th-grade teachers were surveyed (chosen to represent teachers in these grades nationwide based on student abilities, community setting, and SES)	Survey: Teachers perceived their role in SB instruction was the delivery of the state's standards curricula, and they believed the best way to accomplish this was through recitation. Despite their familiarity with other approaches such as small group instruction, projects, and manipulative activities, the teachers expressed a need to conform to tested standards by reverting to a whole-class lecture format. Focus groups: reinforced survey findings

Pedroza (1998) Conference presentation State standards: TX	SB Assessment: Texas Assessment of Academic Skills (TAAS) The study examined the effects of TAAS on the students (mostly English Language Acquisition — or ELA) of a district on the Mexican border. Method: Qualitative Measures: Document review (test scores) and interviews with district teachers (and counselors, administrators, community members, and board members) Sample: There were 31 teachers (and 34 other stakeholders) interviewed over a 9-week period. In 1 rural district with, 99% Hispanic students, 92% FRL, and 82% limited English proficient students.	Student achievement: The TAAS had no direct effect (and no indirect positive effects) on the students of the district due to district characteristics (ELA, rural, and staffing challenges).
Roderick, Jacob, & Bryk (2002) Journal article Local standards: Chicago	SB Assessment: Chicago Public Schools (CPS) mandated testing The study examines the effect of Chicago's mandated testing of student achievement in high-stake grades. Method: Quantitative quasi-experimental Measures: Comparison of student test scores before and after taking mandated tests (high-stake testing grades 3, 6, and 8) Sample: CPS students in grades 3, 6, and 8 with exemptions for retained, special education, and ELL students. 3rd-grade sample size: 200,000 6th-grade sample size: 200,000 8th-grade sample size: 170,000 (reading and math test score data was collected)	Student achievement: Based on HLM analyses, the advent of high-stakes testing in Chicago was followed by a substantial increase in student test scores. Low-performing schools demonstrated greater gains than did higher-performing schools. The lowest-performing reading students and the highest-performing math students showed the largest achievement gains under the testing policy.
Schorr, Firestone, & Monfils (2003) Journal article National standards: NCTM State standards: NJ	SB Assessment: New Jersey Elementary School Performance Assessment (ESPA) The study examined the activities that were characteristic of math teachers' classrooms and the degree to which these activities were influenced by state testing. Method: Qualitative Measures: Classroom observations and interviews of 4th-grade (mandated testing year) teachers	Interviews: The teachers indicated that they had changed their classroom practice to include higher-order problems and activities in response to ESPA. Observations: The choices described by teachers in their interviews were observed in the teachers' practice, but the authors also observed that these practices were superficial relative to the overall classroom emphases and dialogue. Teachers demonstrated an unchanged foundation in their practice evidenced by these deeper characteristics, and

	Sample: The sample was drawn from the respondents of a previous survey and from participants of a SSI (described as being representative of the state in terms of both school SES and geography). 22 4th-grade teachers were chosen because they were identified to be exceptional in terms of their high or low level of direct or inquiry-oriented instruction. 31 4th-grade teachers were chosen due to participation in professional development opportunities offered through the SSI. The 63 teachers were questioned regarding class activity choices and the influence of state testing on these choices.	revealed that their overall approaches were not influenced by the state test.
Schulte, Villwock, Whichard, & Stallings (2001) Journal article State standards: NC	SB Assessment: North Carolina End-of-Grade tests. The study examined the progress and achievement levels of Special Education students under the North Carolina mandated testing system. (The study also examined levels of inclusion.) Method: Quantitative non-experimental Measures: Student test score data were subjected to frequency analyses Sample: The test scores of 461 students with learning disabilities in grades 3, 4, and 5 were analyzed. District: suburban, 19% FRL	Student achievement: The state mandated testing was associated with higher reading achievement levels in learning disabled students in grades 3 through 5. *Note: Direct attribution of results to the testing policy was not possible.*
Smith (1997) Technical report State standard: AZ	SB Assessment: Arizona Student Assessment Program (ASAP) The study examined the effect of mandated state testing on teaching practice. Method: Qualitative Measures: Classroom observations, teacher interviews, and document review Sample: 4 elementary schools with emphasis on grades 3 and 4 (cases chosen to represent state with respect to setting and school SES).	After 4 years under the state standards program, teachers in Arizona made small if any adjustments to their practice in response to ASAP. Only a small minority of responding teachers reported making significant changes in practice. In response to ASAP teachers were observed training students for the tests (e.g. drilling or breaking integrated units into components to encourage mastery). These observations were characterized as activities that are not in alignment with test program higher-order expectations. The lack of conceptual depth of the ASAP had the effect of hampering the efforts of the more advanced classrooms/students in the cases. Teacher choice of conceptual activity at any

		depth greater than ASAP was programmatically discouraged.
Stecher & Barron (1999) Stecher, Barron, Kaganoff, & Goodwin (1998) Koretz, Barron, Mitchell, & Stecher (1996) Technical reports State standards: KY *Note: These 3 titles are treated as 1 study in the Overview of Studies.*	SB Assessment: Kentucky assessment system The multi-year study examined the influence of the Kentucky mandated testing on teaching practice. Method: Quantitative non-experimental (pre- and post-sample design) Measures: Survey of teachers (once in 1994-95, again in 1996-97, and again in 1997-98) Sample: 1999 study: 365 writing and math teachers from grades 4 through 7 (chosen to represent the state) were surveyed. Although writing and math were both tested subjects, the sampled grades provided both tested and non-tested grades for each subject. 1998 study: teachers surveyed (chosen to represent state): 136 4th-grade writing 83 5th-grade math 95 7th-grade writing 77 8th grade math These are all testing grades for subjects. 1996 study: 94 4th-grade teachers and 84 8th-grade teachers surveyed (these teachers were in grades/subjects in preparation for initial mandatory testing cycle)	1999 study: Based on chi-square tests, teachers in testing grades spent a significantly greater amount of time on the tested subject and topics than did their non-testing grade counterparts. Also, teachers in testing grades were significantly more likely to engage on reform-specific practices. 1998 study: Based on chi-square tests and *t*-tests, mandated state testing had a significant effect on teaching practice by increasing reform-oriented approaches to both math and writing instruction such as open-ended questioning, extended written responses to prompts, and the use of math manipulatives. The results also revealed teacher responses to the state tests through classroom time spent on tested subjects, topics, and approaches to classroom assessment. Survey responses also indicated that the teachers in the lower grades were more likely to employ reform oriented approaches such as open-ended questioning, interdisciplinary activity designs, and requiring written responses to math problems. 1996 study: teachers reported aligning classroom curricula with state testing curricula. Specifically, the teachers employed approaches to enhance more creative skills such as problem solving and writing while deemphasizing instruction on untested content. The teachers also reported a decrease in the classroom use of multiple-choice assessment format.
Stecher & Chun (2001) Technical report State standards: WA	SB Assessment: Washington Assessment of Student Learning (WASL) The study examined the influence of the Washington State mandated testing on teaching practice. Method: Quantitative non-experimental (pre- and post-sample design) Measures: Teacher survey Sample: 400 4th- and 7th-grade teachers (chosen to represent the state) were surveyed once in 1999 and again in 2000.	Teacher instruction: Based on chi-square tests, teachers were found to be narrowing the curriculum to mirror the WASL content as well as spending time familiarizing students with test format and test-taking skills. There was also evidence that teachers were allotting more time to tested subjects and therefore borrowing time from instruction on non-tested topics. This approach was of particular concern since Washington was adding new tested topics in gradual expansion of WASL and, therefore, overemphasizing a small number of subjects would not prove to be a viable approach in the future.

		Student achievement: Student demographics were the only significant predictor of student test score differences.
Stone & Lane (2000) Conference presentation State standards: MD	SB Assessment: Maryland School Performance Assessment Program (MSPAP) This study examines the changes in classroom instructional and assessment practices encouraged by state-level testing. Method: Quantitative non-experimental Measures: Survey of teachers and document review (survey responses used to explain differences in student test scores) Sample: Math and English (1996-97): 59 elementary schools, 31 middle schools Science and Social Studies (1998-99): 103 elementary schools, 58 middle schools The schools were chosen to represent the state in terms of school SES and level of change in MSPAP scores reported between 1993 and 1995.	Teacher instruction: Structural equation modeling revealed that instruction-related variables (such as aligning instruction and classroom assessments with the state test) were related to the differences between the school MSPAP scores. Thus, aligning instruction and other reform-oriented approaches were seen as improving student achievement. This trend carried over the different subject area results. In English subject sub-scales, the instruction-related variables also explained the 4-year change of scores leading into the study. Thus, reform-oriented practice led to score gains. There was also correlational evidence suggesting that test training (such as familiarization with test format) was an effective predictor of low-SES school performance levels. Thus, higher levels of test training led to higher achievement scores. Student achievement: Students scored higher on standards-based assessments after being exposed to test-preparation activities.
Taylor, Shepard, Kinner, & Rosenthal (2003) Technical report State standards: CO	SB Assessment: Colorado Student Assessment Program (CSAP) mandatory state tests The study examined the effect of CSAP on teaching practice. (The study also examined the effect of CSAP on teachers' impressions of school and classroom environments.) Method: Mixed methods (non-experimental) Measures: Surveys and interviews Sample: 357 teachers were surveyed by mail 161 teachers were interviewed by phone (The sample was chosen to represent the state in terms of size of district and overall district ability level.)	Teachers have adopted writing programs and given greater emphasis to writing instruction in response to CSAP (both the content standards and the mandated testing), although they also report less attention to social studies and science and increased instructional time spent on test format practice. The authors note other teaching effects , but attribute these changes in practice to the state standards framework and not the state mandated testing. In particular, the standards were perceived to have a greater impact on improving instruction than testing.
Wong, Anagnostopoulos, Rutledge, & Edwards (2001).	SB Assessment: Chicago Academic Standards Examination (CASE) based on the Chicago Academic Standards (CAS) Study examined teachers' perceptions of the	Document review: Frameworks and standards documents were used primarily as accountability rather than professional development in the district, with some minimal test prep materials provided for the

Technical report Local standards: Chicago	effects of state mandated testing on instruction. <u>Method:</u> Qualitative <u>Measures:</u> Classroom observations, interviews and document review (standards, curricula, testing materials, and accountability policies) Interviews of 9th and 10th grade math and English teachers in selected schools. <u>Sample:</u> 4 high schools representing varying degrees of district intervention, all with at least 77% low-income students (other demographics varied); 71 9th and 10th grade teacher surveys of math and science teachers in schools, 8 10th grade classrooms for observation.	CASE. The CAS and CASE had dissimilar instructional emphases. The CASE did not challenge teachers to emphasize the interpretive or implied reasoning skills addressed in the CAS. Interviews: Teachers are responsive to the pressures of the CASE in changing instruction, rather than CAS. Effects include topics chosen, sequence taught, skills addressed, and test preparation activities. Observed teachers focused on coverage of the book and themes they think will be covered in the CASE. Observations: Teachers spent no time engaging students in discussion. The questions asked of students elicited literal answers such as facts cited in the book.

AAAS = American Association for the Advancement of Science
ELL = English language learners
FRL = Free or reduced lunch
NCTM = National Council of Teachers of Mathematics
NRC = National Research Council
NSES = National Science Education Standards
SES = Socioeconomic status