A McREL Report Prepared for

Stupski Foundation's Learning System

Assessment

About McREL

Mid-continent Research for Education and Learning (McREL) is a nationally recognized, private, nonprofit organization dedicated to improving education for all students through applied research, product development, and service. Established in 1966, McREL now maintains a staff of around 110 in its Denver, Colorado, office.

This report is part of a larger set of reports prepared by McREL for the Stupski Foundation. The views, findings, conclusions, and recommendations expressed herein are those of the authors and do not necessarily express the viewpoint of the Foundation. Please e-mail any inquiries to Linda Brannan at info@mcrel.org.

Contact

McREL
4601 DTC Blvd., Ste. 500
Denver, CO 80237
Web site: www.mcrel.org • E-mail: info@mcrel.org
P: 303.337.0990 • F:303.337.3005

Acknowledgements

Many people contributed their time and expertise to this report. The authors express their gratitude to the following McREL staff members: Maura McGrath and Linda Brannan, who lent their library expertise to identifying and collecting the literature reviewed; Lisa Maxfield, who diligently supported our efforts to catalog and record the articles collected; Vicki Urquhart, who provided her keen editorial oversight; and Natalie Voltes, who is responsible for the design and layout of this report. In addition, we thank Jean Williams and Helen Apthorp for offering their expertise on research methodology and Jackie Kearns for her tireless project management efforts. We also acknowledge the support, advice, and constructive dialogue provided by staff members of the Stupski Foundation.

Authors

Tedra Clark, Ph.D.
Kerry Englert, Ph.D.
Dana Frazee, M.A.
Susan Shebby, Ed.M.
Bruce Randel, Ph.D.

20091030

Table of Contents

Executive Summary

This document is one of eight reports prepared to support the development of a new learning system, a development effort that is the first step in a major initiative undertaken by the Stupski Foundation. The Foundation endeavors to improve the life options of all students, especially disadvantaged urban youth of color, whom we refer to as "Our Kids," by fundamentally redesigning the education system.

This report was created collaboratively by researchers from McREL with guidance from officers of the Stupski Foundation. Its purpose is to provide members of a "Design Collaborative" team—consisting of practitioners, parents, students, and researchers—with a review of key findings from existing literature to support their efforts to develop the assessment component of the Stupski Foundation's Learning System.

Research methodology

McREL researchers, in collaboration with Stupski Foundation staff members, generated the following research questions to guide this review:

1. What is the gap between current assessment practice and ideal assessment practice?
2. What is the relationship between formative assessment and student learning and motivation?
3. What are the promising formative assessment practices for promoting students' progress toward becoming college ready?
4. How can we help teachers use formative assessments effectively?

While McREL researchers concentrated on these four questions to guide the assessment literature review, they did so always with an eye toward what benefits students of color and poverty. Thus, insights on how formative assessment might specifically address the unique needs of these underserved children are addressed throughout the report

The research questions served to focus an extensive review of scholarly (i.e., peer-reviewed publications) and "fugitive" literature (i.e., reports self-published by reputable foundations, associations, and other organizations). In all, the research team reviewed 116 articles and summarized 92 of these. Data and conclusions from these reports have been synthesized into several key findings.

Key findings

Findings presented in the report fall into three areas: 1) the relationship of formative assessment to student learning and motivation, 2) characteristics of effective formative assessment practice, and 3) professional development and support systems for improving and monitoring formative assessment practice.

Formative assessment for improving learning and motivation

Currently in the United States, educational assessment is highly focused on standardized state testing for monitoring school performance. However, research indicates that students benefit most when teachers use assessment to understand the extent to which students are *learning* and to make corresponding

changes in their instruction (Black & Wiliam, 1998ab). When assessments are used diagnostically to provide feedback to teachers and students during the course of instruction, it is known as *formative assessment.*

According to research, formative assessment practice has powerful effects on student learning and motivation (see Black & Wiliam, 1998b). Scholars in the area of educational assessment generally agree that when students are evaluated frequently for the purposes of monitoring learning and guiding instruction, they are more likely to be successful learners (Stiggins, 1998). The student who is aware of how he or she learns is better able to set goals, develop a variety of learning strategies, and control and evaluate his or her own learning process. Alternatively, summative assessments, which evaluate student performance at the conclusion of the instructional period, have little to no influence on student learning.

Characteristics of effective formative assessment

Assessment as instruction. Promising practice in formative assessment for improving student achievement involves the application of diverse evaluation practices to everyday classroom instruction to engage students in their own learning. Effective formative assessment involves real-time questioning and frequent classroom discussion to gain an understanding of what students know (and don't know) in order to make responsive changes in both teaching and learning (Black & Wiliam, 1998a).

Descriptive feedback. Feedback is an essential component of the formative assessment process and is widely recognized in the literature as a critical support mechanism for student learning (Callingham, 2008; Cauley, Pannozzo, Abrams, McMillan, & Camou-Linkroum, 2006; Center for Comprehensive School Reform and Improvement, 2006; Hattie & Timperley, 2007; Shepard, 2000; Stiggins, 2004). According to past research, effective feedback is specific, immediate, and focused on students' thought processes, and goes beyond merely directing the student to the "correct" answer.

Alignment. Formative assessments are more effective when they are aligned with learning objectives that 1) provide a trajectory of student learning at key points in the curriculum and 2) guide feedback to students about their performance (Ayala et al., 2008; Stiggins & Chappius, 2008; Valencia, 2008; Wiley, 2008). Furthermore, these learning objectives should be aligned with the unique learning styles, strengths, and developmental needs of individual students (Stiggins, 1998).

Authentic assessment. According to existing literature, high-quality formative assessment involves tasks that go beyond recall or recognition to include reasoning and justification of responses that teachers may or may not have anticipated prior to the assessment. More specifically, learning is enhanced when students are asked to formulate problems, organize their knowledge and experiences in new ways, test their ideas with other students, and express themselves orally and in writing (Newmann et al., 2001).

Technology. One of the challenges to implementing formative assessment is limited teacher time. Current advances in assessment technologies (e.g., computer-based software, Internet) afford teachers and students new ways to track learning progress in reference to automatic and objective feedback (Landauer, Lochbaum, & Dooley, 2009).

Professional development and support systems

Substantial gaps between contemporary ideas about assessment practice and teachers' knowledge about applying these ideas in the classroom are noted in the literature. Professional development can improve teachers' ability to draw clear connections between understanding what students know and how to use

that understanding to enhance learning through instruction (Even, 2005). However, it is unlikely that high-quality professional development will be widely practiced without adequate support and monitoring from various levels of the education system (Chappius, Chappuis, & Stiggins, 2009).

Recommendations

Based on the findings highlighted above, several recommendations are offered at three levels for how the Design Collaborative might proceed with its efforts to develop a K–12 assessment system for improving the educational outcomes of Our Kids.

Option 1: Classroom-level interventions

Interventions of formative assessment techniques at the classroom level should incorportate the following actions:

- Integrate assessment into daily classroom instruction, and more precisely, practice assessment *as* instruction.
- Provide students with frequent and descriptive feedback on their current level of understanding and next steps for enhancing future learning.
- Promote self-assessment.
- Align assessment with learning objectives and unique student needs.
- Utilize technology to improve assessment utility and efficiency.

Option 2: School-level curricular and assessment programs

At the school and/or district levels, the research team encourages the Design Collaborative to examine and potentially adopt a learning system that focuses on project-based learning paired with performance assessment. Several existing schools have demonstrated great success through their commitment to classroom evaluation aligned with intellectually demanding, authentic instruction, as well as an emphasis on transfer of learning to real-world problems or situations (Bass & Glaser, 2004; Crooks, 1988)

Option 3: Professional development and leadership support

A critical step to putting these recommendations into practice is ensuring that teachers receive the training necessary to engage in sound formative assessment practice. Specifically, teachers require high-quality professional development and support from school leaders, as well as district and state personnel, to maximize their effectiveness in assessment practice for improving the learning outcomes of their students.

A lofty but central goal of the Design Collaborative is the establishment of state and federally mandated assessment systems that strike an appropriate balance between accountability testing and classroom-based formative assessment. However, the Design Collaborative must weigh the benefits of this high-level reform with the immediate need to improve the learning outcomes of Our Kids in light of the many challenges (i.e., time, resources, and popular resistance) associated with change at the highest levels of the education system.

Final thoughts

Sound formative assessment in the classroom is one of the most potent factors for influencing student achievement (see Black & Wiliam, 1998ab). It guides students' judgments of what is important to learn, affects their motivation and self-perceptions of competence, structures their approaches to self-study, consolidates their learning, and facilitates the development of enduring learning strategies and skills (Crooks, 1988).

Assessment experts generally agree that a balanced assessment system that addresses both state accountability and assessment for learning is necessary to maximize student achievement (Stiggins, 2002; Valencia, 2008). While less frequent evaluations for summative purposes should focus on describing what students can and cannot do, ongoing evaluation activity in the classroom should be directed toward providing students with feedback to facilitate their learning (Crooks, 1988). Moreover, future explorations in assessment should consider how educators can take advantage of large-scale summative assessments to identify ways in which the results can help improve student performance (Boston, 2002).

Effective formative assessment practice cannot exist in isolation from other areas of the learning system. It must exist in a pedagogical environment wherein teachers practice intellectually demanding and authentic instruction for the purpose of transfer of knowledge to new, out-of-school learning situations. Furthermore, it must correspond with pedagogy and curricula that are sensitive to students' diverse cultures, languages, and backgrounds. Finally, there must be support from all levels of the education system—from school leaders to state and national policymakers—to ensure that teachers have the training and resources they need to improve student learning through assessment in the classroom.

Introduction

Purpose of this document

This document is one of eight reports prepared to support the development of a new learning system, a development effort that is the first step in a major initiative undertaken by the Stupski Foundation. The Foundation endeavors to improve the life options of all students, especially disadvantaged urban youth of color, whom we refer to as "Our Kids," by fundamentally redesigning the education system.

The report was created collaboratively by researchers from McREL and officers of the Stupski Foundation. Its purpose is to provide members of the Design Collaborative team with a review of key findings from the existing literature regarding critical research questions related to the assessment component of the Learning System and to offer recommendations for the development of this component. Together, the reports cover these topics:

- Assessment
- Curriculum
- Pedagogy
- Student Supports
- Systems Diagnostics
- Leadership
- College Readiness
- Our Kids

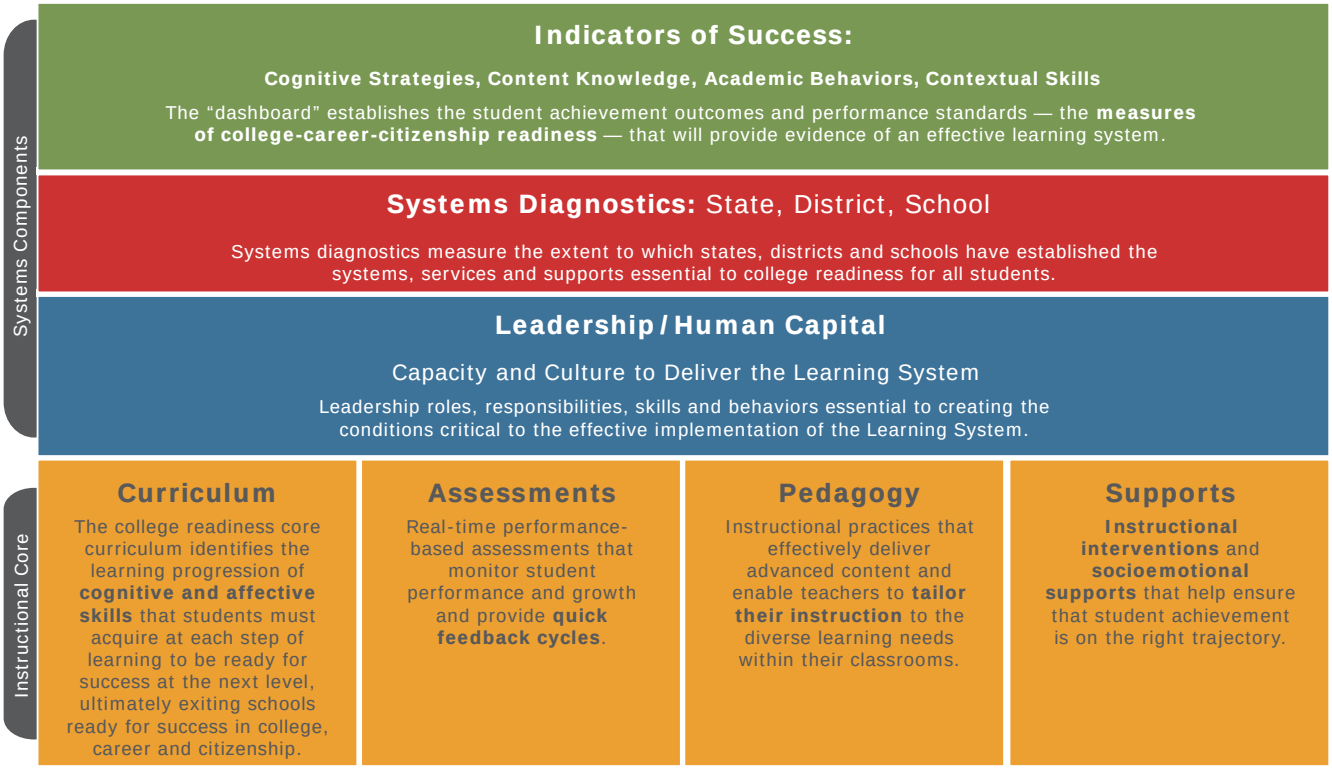
The first section of this report provides salient findings that emerged from the literature review. The second section offers a discussion of the findings along with several recommendations for how the Design Collaborative might proceed with developing a system of formative assessment that is effective in improving the educational outcomes of underserved children. A brief concluding discussion follows. Summaries of the studies and literature reviewed for this report are provided in a separate document.

About the Learning System

The Learning System is the product of the Stupski Foundation's extensive examination of research, best practices, and theories of action for improving education opportunities for all children. It is deeply rooted in the Foundation's mission to foster innovation in public school systems so that all students graduate ready for college, career, and success—as well as the notion that the United States' education system, in its current state, is incapable of accomplishing this goal. As stated on the Foundation's Web site, "The basic components of what public education systems need to teach all students to world-class standards, particularly those students for whom public schools are their only option, do not exist in any coherent, accessible or evidence-based way" (Stupski Foundation, n.d.).

Thus, the Foundation has focused its philanthropic efforts on supporting the "fundamental reinvention" of the American system of public education into one that prepares all children for the challenges of life, career, and citizenship in the 21st century. To accomplish this objective, the Foundation launched a multi-year, cross-sector collaboration among researchers and practitioners from inside and outside education to develop a new and comprehensive learning system. In its June 2008 *Strategy and Program Overview*, the Foundation posited that this system includes seven components, shown in Figure 1 (see p. 6). The indicators of success are dependent on a definition of college readiness, which is addressed in the respective report. Although Our Kids is not an explicit component of the Learning System, it is the basis for the work the Foundation is committed to in the education sector. As such, the populations of students of color and students of poverty warranted a separate report.

Figure 1: The Learning System



About "Our Kids"

The Stupski Foundation is committed to addressing the academic needs of underserved populations, in particular, students who are of color *and* in poverty (which comprises 42% of African American students and 37% of Hispanic students) (Duncan & Magnuson, 2005). Despite a dramatic rise in minorities enrolling in college (a 50% increase from 1995–2005), fewer minorities appear to be graduating. As shown in Figure 2 (see p. 7), in 2006, fewer minorities aged 25–29 reported having obtained an associate degree or higher than their older peers (aged 30 and over) (American Council on Education, 2008). This trend marks an important reversal in advances in educational opportunities for minorities and may mark the first time in history that a generation of students has demonstrated less educational attainment than its predecessors (American Council on Education, 2008).

Overview of methodology

McREL researchers followed a five-step process for translating findings into recommendations.

Step 1: Identification of key hypothesis

After conducting an initial survey of relevant literature, Stupski Foundation staff members identified the following hypothesis to guide the literature review for the assessment component:

The current system of assessment inadequately supports student learning and fails to help prepare Our Kids for college, the workforce, and life in general.

Step 2: Identification of research questions

McREL researchers, in collaboration with Stupski Foundation staff members, generated these questions:

1. What is the gap between current assessment practice and ideal assessment practice?
2. What is the relationship between formative assessment and student learning and motivation?
3. What are the promising formative assessment practices for promoting students' progress toward becoming college ready?
4. How can we help teachers use formative assessments effectively?

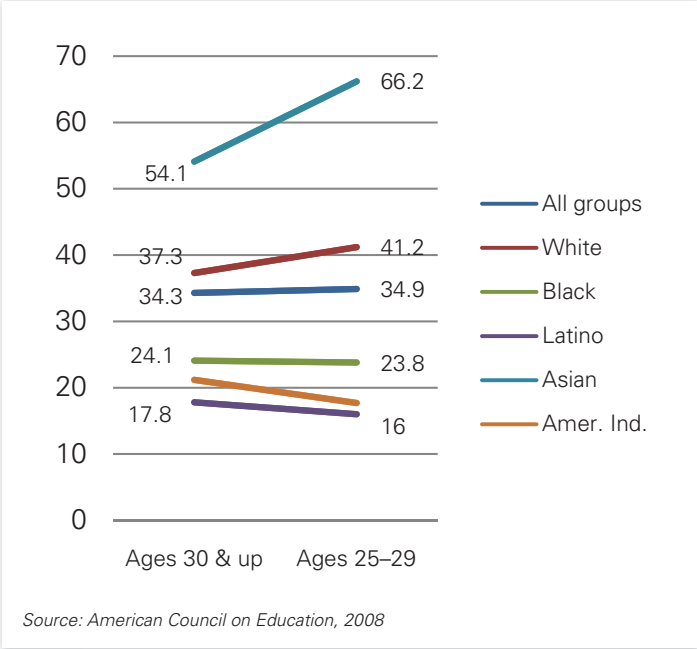
While McREL researchers concentrated on these four questions to guide the assessment literature review, they did so always with an eye toward benefits for students of color and poverty.

Step 3: Literature search

The four research questions guided a search for literature in academic databases (e.g., ERIC, Google scholar, JSTOR, sites at the U.S. Department of Education [e.g., IES databases], and prominent education journals). Sources were searched by several combinations of the following keywords:

- Achievement gap
- African American
- Assessment
- Assessment for learning
- At risk
- Effective
- Feedback
- Formative
- Hispanic
- Ideal
- K–12 education
- Low-income
- Methods
- Minority
- Poverty
- Strategies

Figure 2: Percentage of U.S. adults with associates degree or higher, 2006



Source: American Council on Education, 2008

Articles identified were retrieved and skimmed with particular attention to research methods, outcomes, and recommendations for future study due to gaps in knowledge. The research team then searched references to other studies in these sources, looking for potential consensus or debate. Meta-analyses of particular topics yielded a wealth of additional sources. This search was an iterative process influenced both by the results of research reviewed and by the changing focus of the search area. The retrieved articles were wide ranging in their methodologies, content, and intended audience and included meta-analyses as well as practitioner-focused journals.

A secondary search was conducted during the writing and quality assurance process. The intent of this search was to ensure the inclusion of influential authors in the area of assessment that might have been missed during the initial search. Twenty-four additional articles identified by the lead author in reference sections of salient articles were included in the present review. In total, 116 articles were identified.

Step 4: Identification and cataloging of findings

The research team cataloged findings from the summarized articles using the following identifications:

- Counterproductive *orthodoxies* (conventional ways of providing education which may be impeding student success)
- *Unmet needs* (areas where students are not yet well served by the current system of education)
- *Next practices* (a program or practice that needs to be developed, adapted, invented, and tested in response to an unmet need)
- *Promising practices* (practices based on research but not supported by rigorous efficacy data)
- Current *best practices* (practices demonstrated by research to be effective in improving outcomes for students)

Step 5: Generation of recommendations

In the final phase, research team members collectively reviewed key findings from the literature review in light of the following questions:

- What are the critical unmet needs related to this component of the Learning System?
- What is missing in current practices within this component of the Learning System?
- What is working and why?
- What is *not* working and why?

- What are the biggest misalignments between research and current practice?
- What things should educators do differently in light of the research findings?
- Where is the knowledge base too inconclusive to guide education innovation?
- Where is more research needed to advance practice?

Responses to these questions were synthesized into recommendations for further action. These recommendations include best or promising practices that should be *adopted* and scaled up or *adapted* to new settings or areas where there are gaps in practices that require new innovations to be *invented*.

Overview of the literature base examined

While summative assessment plays an important role in our educational system, the research and theory presented in this report focuses on formative assessment in the classroom for improving student learning. Creating a more balanced program of assessment that is aligned with instructionally sensitive learning goals will ultimately help develop and support the college, workplace, and life-readiness of Our Kids. A detailed discussion of large-scale accountability assessment is beyond the scope of this report; however, recommendations for improving those systems are discussed in the Systems Diagnostics report.

The literature search for this report focused on three major topics related to formative assessment: 1) the relationship between formative assessment and both student learning and motivation; 2) the key characteristics of high-quality formative assessment practice; and 3) professional development, support, and monitoring to increase the prevalence and improve the quality of formative assessment practice. Searches were conducted with a keen interest in formative assessment practices aimed specifically toward improving the outcomes of children of color and poverty. However, it was discovered that the majority of existing literature on formative assessment is more broadly focused on practices that improve learning for all children, regardless of race/ethnicity and socioeconomic status. Although few studies addressed formative assessment specifically for closing racial and/or socioeconomic achievement gaps, in general, the research indicates that all students benefit from high-quality formative assessment with larger learning gains demonstrated by lower achieving students.

Although the ideal works for inclusion in this review describe experimental research (either original studies or meta-analyses), these types of studies are few and far between. In fact, the majority of relevant works identified were theoretical or opinion papers written by authors with extensive expertise in the field of formative assessment. While these works have great value for the development of the Learning System, it should not be assumed that the recommended practices have a causal impact on student learning because, in many cases, they have not been subjected to rigorous testing. Thus, an important unmet need and vital first step in development of the Learning System is the use of rigorous evaluation techniques to identify "best practices" in formative assessment for improving student learning.

In addition, the experimental and correlational studies identified in this review sought only to determine whether a particular practice or program helps students demonstrate higher levels of performance on an outcome measure, usually a standardized achievement test, and not whether students have become more likely to succeed in college or the workplace. In short, these studies seek to answer questions about the effectiveness of particular practices or programs, not "big picture" questions about how assessments can be used to improve Our Kids' life options. Thus,

discussions of improvement in college, workplace, or life readiness in this report are generally only inferred, and not necessarily actual, consequences of increased motivation, learning, and/or achievement.

In summary, the literature reviewed, and the findings that follow, are derived from a variety of sources representing an array of research methodologies. In light of the difficulty of conducting experimental research on something as broad as K–12 formative assessment, the Design Collaborative will need to draw upon these data, but also professional wisdom—including a practical understanding of how to develop assessments aligned to critical learning pathways, insights into the increasing demands of college and workplace environments, and cross-disciplinary examinations of promising practices in other fields—when developing a college readiness assessment system for Our Kids.

Findings

Formative versus summative assessment

This section addresses the first research question, "What is the gap between current assessment practice and ideal assessment practice?"

The current assessment system is focused on standardized state tests administered to monitor school performance. However, research supports the notion that students will benefit most from an assessment system that is balanced, coherent, and comprehensive with regard to student learning goals. This balanced assessment system would include a strong focus on formative assessment, whereby frequent classroom evaluations are used diagnostically to provide feedback to teachers and students during the course of instruction.

Throughout this manuscript, we use the term formative assessment to describe ongoing classroom-based evaluations for improving student learning and informing instruction. Formative assessment activities may involve time spent inside or outside the classroom and include tasks such as teacher-made tests, curriculum-embedded tests, oral questioning, and a variety of other cognitive or psychomotor performance activities.[1] The current definition of formative assessment is in contrast to *summative assessment*, which is a process aimed at a final evaluation of a student's mastery or understanding of information, skills, concepts, or processes. Some common examples of summative assessments are state assessments, end-of-unit or chapter tests, end-of-term or semester exams, scores that are used for school accountability (AYP), and student reports (e.g., report card grades).

According to Black and Wiliam (1998b), assessment refers to "all those activities undertaken by teachers—and their students in assessing themselves—that provide information to be used as feedback to modify teaching and learning activities…[It is] formative assessment when the evidence is used to adapt the teaching to meet student needs" (p. 140). Likewise, Popham (2008) defines formative assessment as a series of evidence-collecting and decision-making events for both teachers and students in order to help students learn. Formative assessment is not merely a test; rather, it is rooted in classroom work. Stiggins (2001, 2004, 2007) takes a similar perspective, using the terminology "assessment *for* learning" to reflect the use of assessment for acquiring useful data to inform instructional practice. Conversely, "assessment *of* learning" reflects the use of assessment data to monitor students' and/or schools' progress (Stiggins & Chappuis, 2008; Stiggins, Arter, Chappuis & Chappuis, 2005).

Key finding

Formative assessment is associated with large gains in student achievement on a wide variety of conventional achievement measures across all ages and all subject disciplines (Black & Wiliam, 1998a).

Currently, assessment practice in the United States is largely summative. However, summative assessments, in particular large-scale accountability assessments, tend to be instructionally insensitive and may even undermine efforts to improve student learning (Popham, 2007; 2009). For example, a major literature review commissioned by The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) in London suggests that state

[1]Some authors use the term formative assessment to refer to benchmark or interim assessments (e.g., Starkman, 2006). For this review, benchmark or interim assessments fall under the category of summative assessments for progress monitoring.

Key finding

Effective formative assessment involves real-time questioning and frequent classroom discussion (Black & Wiliam 1998a).

accountability assessments have a negative impact on achievement of low-achieving students (Harlen & Deakin, 2002). In addition, a literature review by Solórzano (2008) suggests that high-stakes accountability tests do not accurately gauge achievement of English-language learners (ELLs) and may actually widen the achievement gap for ELLs because of punitive consequences, such as unequal retention and unequal graduation rates.

Moreover, research indicates that the effects of summative assessments may extend to other undesirable student outcomes, including reduced intrinsic motivation, increased assessment anxiety, internal attributions of failure, lowered self-efficacy, reduced use and effectiveness of teacher feedback, and poor social relationships among students (Crooks, 1988). In contrast, the literature provides strong evidence that high-quality formative assessment practice produces motivational and performance gains, which are amongst the largest resulting from educational intervention (see Black & Wiliam, 1998ab).

The benefits of formative assessment

This section addresses the second research question, "What is the relationship between formative assessment practice and student learning and motivation?"

Formative assessment and student learning. In 1998, Black and Wiliam published a seminal work on formative assessment titled *Assessment and Classroon Learning*. The manuscript was based on an extensive research review of 250 journal articles and reports to determine if classroom-based formative assessment increases academic achievement (Black & Wiliam, 1998a). The results showed that well-designed formative assessment is associated with major gains in student achievement on a wide variety of conventional achievement measures (standardized, accountability tests), across all ages and all subject disciplines. Effect sizes ranged from moderate to high, with formative assessment having the greatest impact on low-achieving students (Black & Wiliam, 1998b).

Following the initial large-scale review, Black and colleagues used classroom observations, inspection of teacher writings, and individual meetings with teachers and students in six elementary schools to understand the types of instructional techniques that are important in the link between classroom formative assessment and increased student learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004). They found that teachers improved their students' achievement by using instructional techniques in four categories: 1) questioning, 2) feedback through grading, 3) peer-and self-assessment, and 4) the formative use of summative assessments. Overall, the average effect size on student achievement was .3 standard deviations.

Several others have discussed the relationship between formative assessment and student learning (e.g., Boston, 2002; Chappuis & Stiggins, 2002; Crooks, 1988; Stiggins, 1998). In general, there is wide agreement among assessment experts that when teachers use formative assessment as part of their everyday classroom instruction, students are more likely to attain higher levels of achievement. When students are assessed frequently during the learning process, it allows teachers to adjust their instruction to address learning deficiencies and misconceptions before it is too late. Successful formative assessment informs students about their own learning and guides their decision-making so they can become more successful learners in the future (Stiggins, 1998). Students who are aware of how they learn are better able to set goals, develop a variety of learning strategies, and both control and evaluate their individual learning processes.

In contrast, summative assessment occurs at the conclusion of an instructional period, with results typically released months later, and therefore, has no influence on student learning. Thus, if the goal of educational reform is to improve student achievement rather than simply monitor it, then the students themselves (as opposed to parents, teachers, and states) must be the primary consumers of assessment-based data (Stiggins, 2007). This means involving students in the specification of learning objectives, assessing student performance often during the course of instruction, altering instruction based on the results, and providing descriptive feedback for students to improve their own learning.

Formative assessment and student motivation. Effective education requires the fusion of skill and will such that intrinsic interest and motivation are given at least as much attention as cognitive outcomes (Crooks, 1988). Research suggests that when students share in the assessment process, they perceive more control of, and more responsibility for, their own learning (Rieg, 2007). Allowing students to help determine the criteria by which their work is judged gives them a feeling of empowerment and makes evaluation of their work seem less punitive and more constructive (Brookhart, 1997; Rieg, 2007). In turn, the positive effects on self-efficacy and motivation are likely to promote learning and achievement. Consistent with this notion, Haydel & Roeser (2002) found that students who believe they can affect their learning through persistently engaging in the educational process score better on standardized tests.

Key finding

Students who believe they can affect their learning through persistently engaging in the educational process score better on standardized tests (Haydel & Roeser, 2002).

Cauley et al. (2006) performed a large-scale literature review to identify specific classroom strategies for capitalizing on the relationship between formative assessment and student motivation. In general, research indicates that in order to foster feelings of self-efficacy and improve student motivation, assessments must grant students regular opportunities to improve on their work, with errors and mistakes considered a natural part of learning. Furthermore, teacher feedback on student performance should focus on the student's effort and ability and should value the process or strategy toward producing an answer as opposed to the correctness of the answer itself. Lastly, students should be encouraged to use self-assessment strategies that will put them at the center of their own learning experience.

In sum, both research and theory support a strong relationship between classroom-based formative assessment and student achievement. When students are involved in the assessment of their own learning, they become more motivated to

Key finding

Positive learning outcomes are more likely when feedback focuses on how the student can improve his or her performance in relation to learning goals as opposed to non-specific praise or criticism (Bangert-Drowns et al., 1991).

learn, and when students want to learn, they learn better. The sound practice of formative assessment helps students understand their own strengths and weaknesses, provides them with a sense of control over their learning, and motivates them to obtain greater levels of achievement in the future.

Promising practices in formative assessment

This section addresses the third research question, "What are the promising formative assessment practices for promoting students' progress toward becoming college ready?"

Assessment as instruction. Formative assessment is a process that is deeply embedded in ongoing teaching and learning. It involves diverse evaluation practice for engaging the student in his or her own learning. Thus, promising practice in formative assessment for improving student outcomes goes far beyond traditional paper-and-pencil tests. According to Black & Wiliam's (1998a) large-scale review, effective formative assessment involves real-time questioning and frequent classroom discussion to gain an understanding of what students know (and don't know) in order to make responsive changes in both teaching and learning.

Descriptive feedback. Feedback has long been regarded as an essential component of the assessment process and is broadly recognized as a critical support mechanism for student learning (Callingham, 2008; Cauley et al., 2006; Center for Comprehensive School Reform and Improvement [CCSRI], 2006; Hattie & Timperley, 2007; Shepard, 2000; Stiggins, 2004). Feedback is generally defined as information that provides learners with an understanding of how they are doing or have done, as well as what they might do in the future to enhance their knowledge and performance (Callingham, 2008; Cowie, 2005). While few would dispute the notion that feedback is essential for learning, not all feedback is effective. In fact, two separate comprehensive reviews of research on feedback (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Kluger & DeNisi, 1998) found that a substantial number of studies showed negative effects—feedback about performance actually harmed learning outcomes.

There is an extensive body of literature on the nature and extent of feedback as related to its impact on student learning (e.g., Brookhart, 2008; Crooks, 1988; Kulhavy, 1977; Mory, 2004; Shute, 2008). Research suggests that positive learning outcomes are more likely when feedback focuses on features of the task, such as how the student can improve his or her performance in relation to standards and learning goals (Kluger & DeNisi, 1998; Hattie & Timperley, 2007). This task-oriented emphasis is advantageous over nonspecific evaluation (e.g., praise or criticism) or

normative comparisons (Tunstall & Gipps, 1996; for meta-analysis, see Bangert-Drowns et al., 1991). Specifically, it helps students become aware of misconceptions or gaps between desired goals and current knowledge, understanding, and skills, and then helps guide students through the process of obtaining those goals (Brookhart, 2008; Sadler, 1989). Research also suggests that effective feedback includes specific comments about errors and areas of improvement (Brookhart, 2008); however, too specific feedback compromises student exploration of his or her own learning (Goodman, Wood, & Hendrickx, 2004). Furthermore, immediate feedback is more effective than delayed feedback, and presearch availability (i.e., knowledge of correct responses prior to performing the learning activity) is counterproductive (Epstein et al., 2002; Kulhavey, 1977; Kulick & Kulick, 1988).

Self-assessment. Another important component of formative assessment is students' roles in evaluating their own learning (Shepard, 2000; Stiggins, 1998; Valencia, 2008). Self-assessment involves comparison of one's own work with established criteria, critiquing one's own work, or simply describing one's own performance by means of self-reflection. Research shows that children who have an opportunity to reflect on their own work, as related to understood learning objectives, show improved academic performance as compared to those without opportunities for self-evaluation (Fontana & Fernandes, 1994; Frederiksen & White, 1997; see also Stiggins & Chappuis, 2008).

Furthermore, students who receive training in self-assessment for monitoring their own understanding show significant learning gains as compared with those who do not receive training (McCurdy & Shapiro, 1992; McDonald & Boud, 2003). Research also indicates that students who view assessment as a means of personal accountability exhibit higher levels of academic achievement than those who view assessment as a tool for teacher and/or school accountability (Brown & Hirschfield, 2008). Self-assessment is linked to greater interest on the part of the students in evaluation criteria and substantive feedback than in the actual grades received (Klenowski cited in Shepard, 2000). Students who engage in self assessment also report being more honest about their own work, more equitable regarding the work of other students, and more prepared to defend their opinions with evidence (see Shepard, 2000).

Aligned assessment. A critical aspect of high-quality formative assessment is that it is well aligned with classroom-based learning objectives as well as the individual needs, performance levels, strengths, and weaknesses of the students in the class. Research indicates that formative assessments should be aligned with learning objectives that provide a trajectory of student learning, and ideally, teachers and students should work together to develop learning objectives (Ayala et al., 2008; Stiggins & Chappius, 2008; Valencia, 2008; Wiley, 2008).

In order for students to take an active role in the assessment process, they must first share teacher values and goals for learning and be willing and able to assess their current level of understanding and performance in relation to those goals (Cowie, 2005). As Sadler (1989) points out, an

Key finding

Children who reflect on their own work as related to understood learning objectives show improved academic performance as compared with those who are not engaged in self-evaluation (Stiggins & Chappuis, 2008).

indispensible condition for improvement in student learning is that "the student comes to hold a concept of quality roughly similar to that held by the teacher" (p. 121). Consistent with this notion, the results of two separate experimental studies indicate that students who understand their learning objectives, and are encouraged to reflect on their own work in reference to those objectives, demonstrate greater growth in academic achievement than those who do not (Fontana & Fernandes, 1994; Frederiksen & White, 1997).

Research has repeatedly demonstrated that student motivation and achievement are maximized when learning targets and standards are high but attainable (Crooks, 1988). In many instances, however, this is not possible if all students are working simultaneously on the same tasks and trying to meet the same targets (Crooks, 1988). Thus, it is also important that assessment be aligned to the needs of individual children in order that each child is appropriately challenged (Brimijoin et al., 2003). Consistent with this notion, studies have shown that student learning is enhanced when the material is taught at students' individual readiness levels, connected with their interests, and presented according to their strongest learning styles (Crooks, 1988).

Key finding

Instruction that demands higher order thinking and in-depth understanding promotes larger than average gains in standardized test scores (Newmann, Bryk, & Nagaoka, 2001).

Authentic assessment. Traditionally, instruction and assessment have taken the approach of presenting information to students and expecting them to reproduce that information in the form of short, unelaborated answers that conform to the teacher's predetermined notion of what is correct (Crooks, 1988). Research indicates that this type of practice encourages surface approaches to learning, where students quickly forget isolated details that have only temporary relevance to them (Crooks, 1988). In contrast, instruction that demands higher order thinking and in-depth understanding promotes larger than average gains in standardized test scores measuring basic knowledge and skills (Newmann, Bryk, & Nagaoka, 2001).

Key finding

Student learning is enhanced when the material is taught at students' individual readiness levels, connected with their interests, and presented according to their strongest learning styles (Crooks, 1988).

Research also suggests that students remember information best when it connects to their lives and when it is taught within a broader meaningful framework. This type of instruction encourages transfer of knowledge, where students extend their learning and apply it to new circumstances (Shepard, 2005). According to a study by Darling-Hammond, Rustique-Forrester, and Pecheone (2005), students in states currently using assessment systems that evaluate a full range of state standards, including higher order thinking and performance skills, show higher levels of achievement and lower dropout rates.

Technology. Current advances in assessment technology, specifically the use of computer-based software and the Internet, are beginning to afford teachers and students new ways to efficiently assess and track achievement and to do so more efficiently than ever before. While there are many quality high-tech programs available to facilitate classroom assessment, the research team has chosen just three to illustrate their collective benefits.

One example of a computerized assessment tool is The E-rater® system for automated essay scoring from the Educational Testing Service (ETS). This system includes measures of grammar, mechanics, style, organization, development, lexical complexity, and vocabulary usage—all aspects of writing routinely assessed by teachers—within an automatic and objective system. After several years of skepticism, there is now evidence of validity and reliability in the scores of this program (Attali & Burnstein, 2006). In a study conducted by ETS, the essays of students in 6th–12th grade were first scored using the E-rater system and again by two trained human readers using grade-specific rubrics. The results revealed a high correlation between the human scores and those of the E-rater system ($r = .97$).

The major benefit of using technology for formative assessment is that it provides immediate feedback to students regarding their performance, which can improve future learning (Epstein et al., 2002; Kulik & Kulik, 1988). Shute, Hansen, and Almond (2007) describe a computerized system called Adaptive Content with Evidence-Based Diagnosis (ACED) for delivering feedback in an instructional context. This system is capable of delivering elaborated feedback (i.e., feedback that provides an explanation of how one would determine the correct answer) as opposed to just simple verification ("correct" or "incorrect"). In a study by Shute and colleagues, the elaborated feedback had a significant positive impact on students' subsequent performance on the instructional tasks (Shute et al., 2007). Furthermore, the system has the added benefit of including accessibility features for visually disabled students, such as voicing of test content, talking tactile graphics, font enlargement, and screen magnification.

Lastly, a software program called WriteToLearn is a new Web-based tool that integrates practice and assessment in reading comprehension with student writing about what is learned (Landauer, Lochbaum, & Dooley, 2009). Based on the principle of immediate feedback, it uses a combination of summative and formative assessment tools to encourage, instruct, and reward progress while it is happening. For example, to foster reading comprehension and writing ability, the program evaluates students' summaries of reading passages for content knowledge by comparing them with the actual text. A summary scoreboard provides students immediate feedback on the content, spelling, redundancy, and irrelevancy of their summary. Students also receive suggestions on how to improve their summaries, after which they can revisit the reading passage, revise their summaries, and re-submit them for more feedback. Write-To-Learn also creates reports that allow teachers to monitor student progress individually and as a class: the teacher class summary provides the average class performance on a summary or essay writing activity, and individual student reports show particular students in need of improvement by specific performance areas (e.g., organization, sentence fluency, voice). In addition, the program allows teachers to review all student work in individual writing portfolios in order to evaluate progress and to reinforce discussions with individual students, other teachers, and parents. Evidence for construct validity of the system includes studies showing that older students, on average, write significantly better on the same readings than younger students. Furthermore, external validity is supported by results of randomized efficacy trials, indicating that students who used the system scored higher on some state reading comprehension items, with performance positively correlated with the frequency of using the tool (Landauer et al., 2009).

Professional development and support

This section addresses the forth, and final, research question: "How can we help teachers use formative assessments effectively?"

Key finding

Despite its strong relationship with improved student learning, teachers do not readily practice high-quality formative assessment (Popham, 2009).

Research suggests that teachers do not readily practice formative assessment, even when it is prescribed in an adopted curriculum. Furthermore, teachers and students alike tend to view assessment as a negative experience. Most teachers do not consistently share information about test content or discuss scores with their students, nor do they use assessments to inform their own practice (Popham, 2009). In addition, feedback is typically unfocused or of little use to students for improving their work (Weeden, Winter, & Broadfoot, 2002). Moreover, the most commonly used assessments reflect thinking skills that are not complex or rigorous (Popham, 2009).

In 2007, Colby-Kelly and Turner attempted to determine the extent and type of formative assessment used in ELL classrooms and found that the type of feedback prescribed by the curriculum did not align with the type of feedback teachers actually used. The authors also noted examples of teachers' hesitancy regarding formative assessment, including fears that public feedback would lead to student feelings of inadequacy.

In another study, Ross, Rolheiser, and Hogaboam-Gray (2002), found that teachers in a group of U.S. junior high schools were not using most of the assessments and assessment-related strategies that they perceived to be effective. For example, while the teachers seemed to understand the effectiveness of aligning assessments to diverse student needs, they were not actually providing students the opportunity to choose methods of assessment.

In a comprehensive overview of the role of classroom assessment in teaching and learning, Shepard (2001) highlights a substantial gap between current thinking about assessment practice and teachers' actual knowledge about applying these ideas. She states, "Although contemporary rhetoric implies that a shared understanding exists about what it means to use assessment data to improve instruction, examples offered suggest considerable ambiguity" (Shepard, 2001, p. 1093). According to the literature, a major reason for this gap between research and practice concerns the teachers' inability to make sense of students' understanding, knowledge, and learning. Similarly, Fuchs and colleagues stress that it is often difficult for teachers to objectively monitor their students' knowledge, which may lead to an overly positive judgment of the students' abilities (Fuchs, Deno, & Mirkin, 1984).

Another gap in teacher practice concerns the use of assessment data to adapt instruction. While there is extensive literature on the use of innovative assessment methods and tools (i.e., the use of open-ended tasks, journals, observations, etc.), there is little information available regarding how to use the data collected to inform next steps in instruction. If teachers are given new assessment methods and tools to integrate into their teaching, they also must be given ways of using the richer information acquired with those tools to guide instructional decisions and advance student learning.

Otherwise, the new tools will continue to serve the traditional assessment purpose of evaluating performance at the conclusion of an instructional period (Even, 2005).

On the positive side, research also suggests that there is hope that teachers can improve their understanding of what their students say, write, and do. Through high-quality professional development, teachers can learn to be more open to unexpected events, opinions, and answers in the classroom by learning students' common misperceptions. They also can learn to value students' original solutions by paying attention to students' problem-solving processes and by transforming their mode of listening from evaluative to interpretive (Even, 2005). Furthermore, professional development can also improve teachers' ability to draw clear connections between understanding what students know and how to use that understanding in instruction (Even, 2005). Specific recommendations for teacher professional development programs are discussed under Option 3 in the "Recommendations" section (see pp. 29–33).

Discussion & Recommendations

The recommendations presented in this section are derived from the research findings outlined in the previous section. In addition, they were shaped by the research team's understanding of the current "state of play" in assessment and in some cases, insights from other literature and knowledge within and outside the field. In addition to the questions described in Step 5 of the Overview of Methodology (see page 8), the following questions were used to guide the recommendations:

- What current practices have a strong enough evidence base that they should be *adopted* and scaled up?
- What current practices show enough promise in certain contexts that they might be *adapted* for use in settings for Our Kids?
- Where are there sufficient unmet needs and lack of promising practices to warrant the *invention* of new practices?

With the desire to have the greatest impact on as many students as possible in the shortest amount of time, the Design Collaborative might choose only a subset of recommendations on which to focus their efforts in the short term. However, it is important to stress that these recommendations are not mutually exclusive, and an ideal system of formative assessment would include all, if not more, of the practices, programs, and systems described below. Since pursuing some areas and not others would most certainly present opportunity costs, the benefits, costs, and potential road blocks associated with their practical implementation must be carefully weighed by the Design Collaborative.

An important issue to consider is the level of the education system at which reform will be initiated. Specifically, the Design Collaborative could implement classroom-level interventions in the form of formative assessment tools and techniques associated with positive student outcomes. Alternatively, it might tackle assessment reform at a higher level, for example, by developing and offering professional development for teachers and school leaders in sound formative assessment practice.

Intervention at any level of the education system comes with a unique set of benefits and challenges. Following, we present our recommendations as three hierarchical options: 1) classroom-level interventions, 2) school-level curricular and assessment programs, and 3) professional development and leadership support. It is important to note that the research team believes that in order to maximize the benefits of formative assessment for improving student learning, intervention must come at all three levels.

Option 1: Classroom-level intervention

Integration of classroom-based formative assessment techniques into classroom instruction does not have to be an elaborate, costly endeavor (Leahy, Lyon, Thompson, & Wiliam, 2005). For example, in working with the Educational Testing Service, teachers in nine states tried, adapted, and invested in at least 50 formative assessment techniques, many of which required only subtle changes in practice. Research on the underlying strategies of these techniques suggests that the small changes in practice can leverage large gains in student learning (see Black & Wiliam, 1998ab). Moreover, the teaching practices that support these strategies are, for the most part, low-tech, low-cost, and usually feasible for individual teachers to implement. Some of these assessment

techniques have proven just as effective, if not more effective, than comprehensive school reform models (Yeh, 2008). In addition, the notion that changes in assessment practice should focus on implementing low-cost, high-reward techniques is reflected in Dylan Wiliam's statement, "If we are to effect substantial change at scale, we need to focus on the changes that we can produce most easily" (Wiliam, 2006, p. 288). In light of these arguments, it seems reasonable that the Design Collaborative work to ensure the continued development and implementation of classroom-level instructional interventions.

The following recommendations for improving teacher practice were derived from key findings in the literature: 1) integrate assessment and instruction, 2) provide frequent descriptive feedback, 3) promote self-assessment, 4) align assessment with learning objectives and student needs, and 5) utilize assessment technology.

Integrate assessment and instruction

The literature on formative assessment stresses the importance of integrating assessment and instruction in order to address and support student progress toward specified learning goals. Teachers and students require ongoing and immediate feedback on how the students are performing relative to learning goals and to take action to close any gaps. Black and Wiliam (1998b) recommend the use of several classroom-based instructional techniques, which include, but are not limited to the following: 1) encouraging students to discuss their thinking about a question or topic in small groups or pairs and having a representative share what was discussed with the larger group (often called think-pair-share); 2) offering several possible answers to a question and asking students to vote on them; 3) asking students to write down an answer to a question and then reading a selected few aloud to the class; 4) asking students to express in writing their understanding of a concept both before and after instruction; 5) having students complete problems or answer questions at the end of a class period and check their answers; 6) interviewing students individually or in groups to understand their thinking processes as they solve problems; and 7) having students complete brief, in-class writing assignments to express their understanding of a topic.

Other examples of useful classroom strategies are presented by Crumrine and Demers (2007) in their description of a group of teachers who use real-time, verbal questioning to assess student learning and guide instruction. Specifically, while presenting a lesson, the teachers elicit responses from all students simultaneously by having them use gestures, flash cards, whiteboards, or clickers to indicate their level of understanding (see also Leahy, Lyon, Thompson, & Wiliam, 2005). The teachers also use index cards to obtain responses to questions at the end of class, as

well as longer journal writing to help gain an understanding of individual students' depth of knowledge and thought processes. Based on the student responses, they adjust their lesson for the next class.

Black and Wiliam (1998b) suggest that tests and homework assignments also can be used in a formative manner if the teachers provide specific and timely feedback to students regarding ways to improve their learning. Thus, they recommend frequent short tests as opposed to infrequent long ones, testing learning of newly presented information within a week of the first exposure, and ensuring the quality of test items by working with other teachers and outside sources.

In addition, activities such as producing a rough draft or completing a scientific experiment are good examples of formative assessment because they allow students to receive feedback and make adjustments to their work (Popham, 2009). Another recommendation is to create annotations of student work and compile them into portfolios. This practice provides an understanding of growth over time and highlights gaps in understanding (Duschl & Gitomer, 1997).

When more formal testing is warranted, it should incorporate immediate feedback and opportunities for students to modify their responses (e.g., the Feedback Assessment Technique, Epstein et al., 2002). A test or quiz should not be considered formative assessment unless feedback is delivered immediately and the student has the opportunity to take the same test, or some version of it, again.

Provide frequent descriptive feedback

As discussed in research findings, one of the key components to effective formative assessment is that teachers provide students with descriptive feedback as they learn. Several practical recommendations for enhancing the effectiveness of feedback emerge from the research (see Crooks, 1988 for partial review).

First, feedback should focus students' attention on their progress in mastering educational tasks. Focusing on personal progress encourages transfer of knowledge, fosters greater self-efficacy, encourages effort attributions, reduces social comparison, and improves student attitudes toward the subject matter (Bangert-Drowns et al., 1991; Crooks, 1988; see also Brunot, Huguet, & Monteil, 2000; Elawar & Corno, 1985). Second, feedback should be provided in a timely manner, when it is still relevant to the learning experience. This usually means that the feedback should be provided during or immediately after the task is completed; however, if feedback is available too early, students will not have the chance to thoroughly process the material. Third, feedback should be followed by subsequent opportunities to demonstrate improved performance as a result. Fourth, feedback should be specific and related to student needs. While basic knowledge of performance should be provided consistently, more lengthy and detailed feedback should be reserved for working through specific misconceptions or weaknesses in performance. Fifth, praise should be used sparingly and, when used, it should be task specific. Criticism is generally counterproductive for enhancing student learning and motivation. High-quality feedback enhances student motivation and learning by emphasizing student effort as opposed to a predetermined level of ability (Boston, 2002; Mory, 2004).

As highlighted by Kluger and DeNisi's (1996) review, certain types of feedback are ineffective or actually inhibit a student's future learning. For example, letter grades, check marks, nonspecific comments, such as "good job," and rigid feedback that simply directs students to a predetermined or expected answer do not help students identify strengths and weaknesses upon which to capitalize or improve, and may even undermine confidence, motivation, and performance in subsequent learning opportunities (Black & Wiliam, 1998b;

Chappuis & Stiggins, 2002). As expressed by Leahy et al. (2005), "To be effective, feedback needs to cause thinking. Grades don't do that. Scores don't do that. And comments like 'Good job' don't do that either" (p. 22).

Promote self-assessment

A vital characteristic of effective classroom-level intervention in formative assessment is the active involvement of the student in his or her own learning process (CCSRI, 2006). Teachers and students should work together to develop guidelines (e.g., rubrics, checklists) that clearly explain the standards against which their work will be evaluated. According to Stiggins (2004), well-designed assessment helps students understand what success looks like and how to improve in the future. Similarly, Black and Wiliam (1998a) argue that true success in student achievement requires that students "understand the main purposes of their learning and thereby grasp what they need to do to achieve" (p.10).

When integrating formative assessment into everyday instruction, teachers should actively encourage students to evaluate their own performance and level of understanding in a way that will guide future instruction and learning opportunities. In this type of learning environment, instruction and formative assessment are indivisible. Differentiated assessments are not only instructive, but serve to modify teaching and future learning activities. The teacher determines students' current level of understanding, tracks their responses to challenges, measures their performance outcomes against expected goals, and implements corrective instruction and provides additional opportunities to demonstrate learning (Brimijoin et al., 2003). As emphasized by Hargreaves, Earl, and Schmidt (2002), such assessment systems that use multiple measures and include student self-reflection serve the diverse range of skills evident in most classrooms.

Align assessment with learning objectives and student needs

In 1990, Gardner proposed a new model of education called *individually configured excellence* designed to support inclusiveness in education (cited in Glaser & Silver, 1994). According to this model, the primary focus of education is not to select or sort students into rigidly defined ability categories, but to identify and nurture individual sources of competency. Thus, schooling should provide for a wide range of opportunities individually tailored to the needs, backgrounds, talents, interests, and prior performance of students.

Classroom-level strategies for addressing individual differences and ensuring alignment of assessment to students' needs are to 1) set targets for individual students, 2) be flexible when setting targets, and 3) provide opportunities for cooperative learning in order to reduce pressure on individuals and compensate for specific areas of weakness (Crooks, 1988). Furthermore,

the use of rubrics allows an individual student to characterize his or her strengths and weaknesses with respect to instructional design (Stanford & Siders, 2001). Lower achieving students may benefit from identification of more attainable intermediate goals, thus allowing for a pattern of repeated success that will improve self-efficacy and motivation (Crooks, 1988). This differentiation is a key element for ensuring that all students meet learning targets and standards, especially in low-socioeconomic and high-diversity areas (Brimijoin et al., 2002).

Student-based alignment of assessment is particularly important for addressing the needs of culturally diverse learners. This point is nicely illustrated by Glaser and Silver (1994) in their description of a real classroom situation wherein teachers' consideration of students' cultural backgrounds led to a greater understanding of an unexpected answer to a seemingly simple mathematics problem. In preparing for a mathematics assessment in an urban middle school, a group of teachers asked their students the following open-ended question:

> *Yvonne is trying to decide whether she should buy a weekly bus pass. On Monday, Wednesday, and Friday, she rides the bus to and from work. On Tuesday and Thursday, she rides the bus to work, but gets a ride home with friends. Should Yvonne buy a weekly bus pass? Explain your answer. Bus Company Fares: One way = $1.00; Weekly Pass = $9.00*

At a subsequent meeting, the teachers met to discuss their students' performance, and to their surprise, many students indicated that Yvonne should purchase the weekly pass. Curious about this answer, the teachers discussed the question in class to gain a better understand of their students' thinking. During the discussion, they learned many students believed that purchasing the weekly pass would be a better decision because it would allow family members to use it, and it could also be used by Yvonne and her family on the weekends. The students' reasonable explanation, based in the context of urban living and cost-effective transportation, demonstrated to the teachers that there was more than one correct answer to the question.

This example highlights several important considerations regarding tailoring instruction and assessment to students. First, it highlights the notion that if a teacher's goal is to accurately assess what students know and are able to do, students must provide both their answers and their thinking and reasoning for deriving their answers. Second, effective instruction and assessment should consider and capitalize on the cultural backgrounds of the students as well as the array of problem-solving skills they have acquired from their home environments. Third, increasing the relevance of instruction to the lives of students involves more than presenting a real-world problem; real-world solutions must be considered as well. Fourth, standardized tests that do not offer opportunities for explanations of reasoning may be insensitive to individual differences among students of varying backgrounds, thereby introducing bias into the test results (Bentz & Pavri, 2000; see also Geva, 2000; Torres-Guzman, Abbate, Minaya-Rowe, & Brisk, 2002).

Based on the work of Kusimo and colleagues (2000) and Perrenoud (1998), recommendations for teachers in assessing children of culturally diverse backgrounds include 1) deepen your understanding of diversity and the cultures from which students come in order to better understand how culture and language influence students' responses to assessment (e.g., it might be difficult for students from certain cultures to participate in a whole class discussion), 2) always consider language and cultural differences as a potential reason for students not performing well on an assessment, 3) be a critical user of large-scale or externally created assessments, 4) prepare

students for the assessment experience, and 5) make test publishers aware of cultural and translational issues that might impede the validity of the test.

In addition, the following steps might help achieve equity among children of diverse backgrounds: 1) be vigilant about the interpretation of assessment outcomes and the consequences of those outcomes; 2) be flexible about the design of assessments; 3) increase your knowledge about how to instruct a diverse population; 4) work with communities to better understand students; 5) use multiple sources of information in assessment; for example, teacher observations or structured interviews to ascertain students' academic skills, interests, areas of strength and weaknesses, and background variables that may affect a child's academic progress; 6) focus on student motivation; 7) provide assessment accommodations; and 8) improve communication among teachers regarding how to better assess diverse students.

Utilize technology

The practice of formative assessment will likely benefit from the increased efficiency, automaticity, and objectivity inherent to technology. In the "Findings" section of this report (see pp. 11–19), positive associations between student learning and three separate computerized assessment tools—E-Rater, ACED, and WriteToLearn—were discussed. A word of caution, however, is that teachers should not allow increased reliance on technology to compromise their involvement in formative assessment. Teachers must preserve their involvement in developing learning targets, gauging student performance, and adapting instruction to meet the needs of students. The introduction of technology should facilitate the efficiency and effectiveness of formative assessment by increasing the differentiation of assessment tasks and allowing for simultaneous evaluation of multiple students.

Potential benefits of this option

- Several low-tech, low-cost techniques exist for teachers to integrate assessment and instruction that are extremely useful for eliciting information on how well students are conceptualizing information because they set the stage for providing immediate and individualized feedback for improving student performance.
- Descriptive feedback provides students with an understanding of what they are doing well, links to classroom learning, and gives specific input on how to reach the next step in a learning progression.
- When students take active roles in evaluating their learning, they become effective self-assessors for improving their own educational outcomes, and as a result, their levels of achievement improve (Shepard, 2000; Stiggins, 1998; Valencia, 2008). Moreover, by providing students with specific techniques for self-assessment, the teacher fosters in students autonomy and control over their own learning outcomes.

- Alignment of assessments with the learning styles, strengths, and weaknesses of individual students is another important aspect of quality formative assessment, especially for meeting the needs of underserved and culturally diverse students (Stiggins, 1998).
- Some research findings suggest there are benefits to using formative assessment that incorporates software tools for assisting students in evaluating their own learning and teachers in gaining and maintaining awareness of the learning progress of their students.

Potential drawbacks of this option

- Simply providing a teacher with a set of tools and techniques for implementing formative assessment in the classroom will not ensure their effective use for supporting student learning. Teachers must learn to appropriately align classroom assessment to well-specified learning objectives and to tailor their instruction based on student performance in order to have a positive influence on student achievement and motivation.
- An overreliance on technology for providing feedback to students could cause teachers to become *less* involved in formative assessment. Because teacher response to knowledge about students' strengths and weaknesses through tailoring instruction is a cornerstone of sound formative assessment, it is important that technology only be used to facilitate teacher involvement rather than replace it.

Option 2: School-level curricular and assessment programs

Adopt project-based curricula paired with performance assessment

Based upon the research findings, a central goal of assessment should be to foster deep and interactive learning through classroom evaluation aligned with intellectually demanding instruction and an emphasis on transfer of learning to novel problems or situations (Bass & Glaser, 2004; Crooks, 1988). Postholm (2006) suggests that use of project-based curriculum paired with performance assessment is an important next step for gauging intellectually demanding and real-world oriented thought processes.

Currently, there are a small number of schools across the United States combining project-based learning and performance assessment in an attempt to enhance the educational outcomes of their students through authentic instruction and assessment. For example, The New York Performance Standards Consortium utilizes a curriculum that requires students to apply what they learn to real-world tasks, such as designing a school building, improving the water quality of a nearby pond, arguing a case before a mock Supreme Court, or writing a play and having it performed. In these schools, performance assessment is an essential companion of project-based learning. As stated by Grant Wiggins, these students are "being tested the way historians, mathematicians, museum curators, scientists, and journalists are actually tested in the work place (cited in Furger, 2002). In fact, the consortium currently uses performance assessment as an alternative to the traditional state high school exit exam (Furger, 2002).

Project-based learning is also an integral component of the Key Learning Community, a K–12 school in Indianapolis. Student progress in this school is based on Gardener's (1983) theory of multiple intelligences and Boyer's (1995) theory of human commonalities. Before earning a diploma, students must demonstrate and document their real-world applied knowledge in what Boyer identified as "eight human commonalities." These include, but are not limited to, the shared use of symbols (through the creation of a multimedia presentation), shared production and consumptions (through a project on marketing and economics), and shared sense

of time and space (through a project on the history of Indianapolis or the contributions of an ethnic group to the development of the city).

The Key Learning and New York Consortium schools share a common commitment to developing a project-rich curriculum supported by an integrated system of assessment that encourages intellectually challenging, authentic learning experiences. These experiences are important components of formative assessment because they encourage greater student participation in the assessment process. More specifically, the process of elaborating responses and performing complex tasks facilitates deeper levels of understanding and shapes future learning opportunities.

A typical project at a performance-based learning school requires a significant amount of time, as it is subjected to multiple revisions involving many hours of discussions with teachers about key issues, questions, and concepts. However, many educators involved in this type of learning system would agree that it is time well spent. Linda Darling-Hammond, who has worked with the New York Consortium for more than a decade, suggests that the member schools' high college-acceptance rate of 91 percent (as opposed to 62% in all New York City schools) is a testament to their rigorous project-based curriculum and assessment system (Furger, 2002). She states, "The time [for performance assessments] is not lost to teaching and learning. The time *is* teaching and learning, because the actual conduct of the assessment is a learning experience for students as well as teachers" (cited in Furger, 2002).

Potential benefits of this option

- A project-based curriculum paired with performance assessment is useful for gauging whether students are developing intellectually demanding and real-world oriented thought processes.
- Project-based learning fosters the development of learning skills that will benefit students in a variety of contexts throughout their lives.

Potential drawbacks of this option

- Typically, performance-based projects require a lot of time for teachers to meet about key issues, questions, and concepts and to make multiple revisions.
- In many instances, a shift to this type of project-based learning will require a complete overhaul of school-wide curricular and assessment systems. This process will require more resources, time, and external support than is currently available to most schools.

Option 3: Professional development and leadership support

Develop and offer professional development for teachers

Many teachers do not practice research-based formative assessment, and in fact, many know little about it. In short, their level of assessment literacy is not adequate. According to Popham (2009), teachers who are assessment literate not only know how to create appropriate assessments, but they also know a variety of assessment options. If teachers are to effectively promote student learning and create supportive learning environments, they must develop a deeper understanding of the assessment-instructional cycle. From a motivational standpoint, Heritage (2007) believes that in order for formative assessment to not be viewed as "just one more thing to do," teachers must begin to view classroom assessment as a worthwhile process that yields valuable information about students' learning. This mentality begins with assessment literacy stemming from teacher training programs and high-quality professional development.

In a 2004 article, Stiggins states, "Teachers must possess and be ready to apply knowledge of sound classroom assessment practices" (p. 26). He asserts that this readiness is achieved when teachers are able to apply assessments that 1) arise from and reflect clear achievement targets, 2) arise from and promise to serve clearly articulated purposes, 3) accurately reflect the desired target, 4) lead to confident conclusions about student achievement, 5) control for relevant sources of bias, and 6) give rise to timely and understandable communications about student achievement (Stiggins, 1998)

Unfortunately, many teacher education programs do not require a course in educational assessment, or, if they offer one, it does not provide teachers with the knowledge and skills to implement a classroom assessment plan (Heritage & Bailey, 2006). Indeed, Stiggins (2002) asserts that the United States has a "national faculty unschooled in the principles of sound assessment" (p. 762). Experts agree that professional development is essential for teachers to change their practice from using assessments that are summative and evaluative to using assessments that are formative and reflective (Ayala et al., 2008; Herman & Choi, 2008).

Even (2005) asserts that it is not sufficient to merely prepare teacher candidates to plan and use various instructional strategies. Candidates must develop a keen understanding of the link between assessment and instruction and learn to use various alternative assessment techniques to understand the needs, interests, readiness, and learning profiles of students. Teachers require opportunities to reflect on their own assessment practices and benefit from observing and consulting with other professionals regarding advantageous changes (Boston, 2002). During sound professional training, teachers would have opportunities to learn how to embed assessment into the teaching process. This would help them abandon common concerns that assessment would interfere with their teaching or that they do not have the time to add assessments to a class period—concerns that highlight the widespread misconception that teaching and assessment are mutually exclusive activities (Heritage & Bailey, 2006).

Training and professional development should advise teachers how to motivate students (particularly those with low self-esteem) through assessment; communicate to students the learning requirements and expectations; develop opportunities for self- and peer assessments; and improve feedback quality (Weeden & Winter, 2002). In addition, Holloway (2003) argues for a closer alignment of professional development activities to student learning needs, as

demonstrated by both formative and summative assessment scores. Holloway advocates for adhering to Schmoker's (2002) guidelines for professional development activities "by having teachers work in teams that focus on assessment standards; review achievement data to target learning gaps; and regularly design and assess instructional strategies to target the specific standards that students are not meeting, according to the assessment data" (p. 2). He reports that, in several districts that have used student assessment data to target professional development activities, there is a general trend toward narrowing achievement gaps and increasing the percentage of students meeting proficiency standards in reading.

Indeed, teacher collaboration can be a powerful form of professional learning, and some think it is a necessary component of designing assessments for improving student learning (McTighe & Emberger, 2006). Specifically, teachers should work together to design tasks and assessments that are based on desired learning results, utilize a peer-review process for acquiring useful feedback on assessment designs, and conduct group evaluations of student work elicited through assessments (McTighe & Emberger, 2006).

Consistent with these principles, Dichter and Orlen (2005) utilize the principals of collaboration in their efforts to improve assessment practices in New York City schools and build professional communities for supporting the achievement of traditionally marginalized, underserved students. Their practice, adapted from National School Reform Faculty guidelines, encourages teachers to use protocols for "reflective practice, collaboration, and shared leadership" (p. 60). In this context, feedback from colleagues is an important component for improving teacher practice in "respectful, thoughtful, and safe conditions" (p. 60). The teachers use collaborative protocols for examining both student and teacher work, developing teacher portfolios, peer observation, and discussions of articles and books detailing research-based practice.

Over a decade ago, Black and Wiliam (1998b) suggested setting up local groups of teachers to offer guidance and support in integrating formative assessment into everyday practice. Today, collaboration among teachers is a major element of several training resources in formative assessment. For example, Classroom Assessment for Student Learning (CASL), a professional development program heavily focused on teacher collaboration (Stiggins, Arter, Chappuis, & Chappuis, 2005), is currently under rigorous evaluation by the Regional Education Laboratory–Central administered by McREL. Likewise, experts at the Educational Testing Service are involved in creating a set of tools and workshops to support teachers in developing deep and practical understanding of assessment for learning, particularly through school-based teacher learning communities.

In each of these programs, teachers are first introduced to the basic principles of formative assessment. Subsequently, they are encouraged to begin implementing these techniques in their classroom and meet with colleagues regularly to discuss experiences and learn what other teachers are doing. In this context, teachers are accountable for their learning and practice because they know they will have to share their experiences with colleagues (Leahy et al., 2005). Over time, the teacher learning community develops a shared language that enables teachers to easily discuss the techniques they are using. Teachers build both individual and collective skills and confidence, while colleagues help identify pitfalls and provide encouragement for taking on new challenges.

Involve school leaders

A shift away from traditional summative assessment practice to formative assessment practice requires multiple levels of support. Thus, the research team recommends the compilation of a support system, or network of people and resources interacting for mutual assistance in the design, implementation, and monitoring of classroom-based formative assessment to improve student learning. Support can come at every level of the public education system. Although teachers can provide support to one another in the form of learning communities to foster assessment literacy, they cannot do it by themselves.

School leaders are essential in the cultivation and implementation of a formative assessment initiative. Leaders can provide support by directing teachers to information on the correct usage of formative assessment, providing the training and tools for carrying out the practice, and monitoring both teacher and student progress (Chappuis, Chappuis, & Stiggins, 2009). However, they face many barriers in this endeavor. Some hindrances are the widespread belief that more frequent standardized tests will pave the way to school improvement, the shared fear of being held accountable for student learning, the collective lack of clarity about achievement targets, and the traditional misconception that grades and standardized test scores are what motivate students to want to learn (Stiggins, 2001).

The first step for school leaders in supporting the practice of formative assessment is to gain a clear understanding of how it can contribute to student success. This comes from engaging in a professional development process much like that of the teachers. Principals can use the professional development as an opportunity to model ongoing learning by leading a school or district learning community on formative assessment (Chappuis et al., 2009). The next step is to facilitate the consideration of a school-wide change in assessment practice to meet the needs of students (Stiggins, 2001). Thus, the leader's responsibility moves beyond managing an assessment system to facilitating improvement within it (Stiggins, 2001).

According to Stiggins (2001), in order to lead efforts to integrate assessment into their school's teaching and learning process, leaders must ensure two important conditions are met. First, they must ensure that all teachers specify and understand the learning targets that their students are to master. This might involve taking steps to specify what the school and community expect students to know and be able to do at certain ages and how those targets for success can be realized within the local curriculum (Stiggins, 2001). The second condition leaders must meet is to ensure that the school faculty is able to gather dependable information about student learning and use that information in instructional practice to improve student learning. Thus, school leaders must take steps to provide adequate time and resources for teachers and staff members to receive training and professional development about formative assessment.

Develop a balanced state-level assessment system

An important issue to consider is the extent to which state-level (or even national-level) support is

necessary for improving student learning through classroom-based formative assessment. One of the ways states might facilitate this process is to help ensure that formative assessments are well aligned to state academic content standards (Popham, 2006). Specifically, they could provide training for teachers to improve their understanding of the state standards tested on large-scale accountability assessments and to help them analyze the sub-skills and important knowledge within those standards that students are expected to master. According to Rabinowitz and colleagues, schools, local districts, and states should work together to ensure that the implementation of formative assessments is strategically planned (Rabinowitz, Roeber, Schroeder, & Sheinker, 2006).

Some states are currently developing and implementing a system whereby teachers have access to formative assessment items that closely correspond with the standards and the state's assessment (Rabinowitz et al., 2006). However, even with this support, teachers must make sound qualitative judgments about the achievements of students in relation to standards for the purpose of improving learning (Sadler, 1987). Thus, it is vital that classroom teachers be trained on the appropriate use of formative assessments and how to effectively deal with the data their states provide (Rabinowitz et al., 2006).

Some experts believe that the only way to ensure the widespread practice of sound formative assessment is to create balanced, state-level assessment systems that place as much emphasis on assessment *for* learning (formative assessment) as assessment *of* learning (summative assessment). For Stiggins (2002), this means matching every dollar invested in large-scale accountability testing and investing it in the support and development of classroom-based formative assessment. He also stresses the need for comprehensive, long-term professional development programs at the national, state, and local levels to foster literacy in classroom assessment, as well as a similar construct for developing assessment literacy in state, district, and school administrators. Lastly, he argues that preparation programs and licensing standards for teachers and administrators must have a component and expectation of competence in both summative and formative assessment.

With limited out-of-classroom time for training and professional development, it is likely that higher level initiatives from the state or federal government will take precedence over smaller-scale, teacher- or school-initiated programs, workshops, or learning teams, thereby maximizing compliance and implementation fidelity from teachers. However, with the No Child Left Behind Act (NCLB) and senate bills mandating ever-increasing development and attention to accountability assessment, development efforts for statewide balanced assessment systems are apt to be met with significant resistance. The Design Collaborative will be faced with the challenge of overcoming this resistance and shifting the focus of educational assessment from accountability to learning at all levels of the system.

Potential benefits of this option

- Findings suggest that using student assessment data to target professional development activities is one factor contributing to narrowing achievement gaps and increasing the percentage of students meeting proficiency standards in reading.
- Assessment reform at the state level is likely to be widespread, thereby affecting most, if not all, students in the state.

Potential drawbacks of this option

- Because many teachers lack assessment literacy, high-quality professional development in this area must be developed to correct teachers' misconceptions about teaching and assessment, which requires a substantial time and financial investment.
- Similarly, it will take time and a financial investment to develop needed support networks for school leaders who must be prepared to deal with entrenched belief systems and to lead a shift away from traditional summative assessment to more balanced assessment practice.
- In the current era of NCLB, there is likely to be resistance to significant change in the existing accountability-focused state assessment systems.

Final Thoughts

In summary, creating an assessment system wherein both summative and formative assessments are focused on learning as opposed to evaluation is likely to have important educational benefits for Our Kids. Black and Wiliam (1998a) describe it this way:

> Too often, assessments are viewed by students as something that is done to them. . . When assessments are an ongoing, clear reflection of the desired results of a unit of study, the potential exists for students to be more invested in the process. Assessment becomes something that is done with students. (p. 67)

Classroom-based formative assessment guides students' judgments of what is important to learn, affects their motivation and self-perceptions of competence, structures their approaches to self-study, consolidates their learning, and facilitates the development of enduring learning strategies and skills (Crooks, 1988). As stated earlier, a focus on sound formative assessment in the classroom appears to be one of the most potent factors for influencing student achievement (see Black & Wiliam, 1998ab). In contrast, summative assessments that are not linked to classroom learning objectives but instead are focused purely on evaluation of students or schools have been linked to negative student outcomes, including reduced academic performance of low-achieving students, decreased motivation and self-efficacy, increased test anxiety, and internal attributions of failure, among others.

Because high-quality formative assessment practice is focused on authentic, intellectually demanding tasks, the knowledge and skills gained from this type of education within the Learning System initiative are more likely to transfer to future learning episodes in college and beyond. Through high-quality formative assessment practice, students learn to incorporate new information into their existing knowledge base, a skill that will prove useful throughout their lifetimes. Furthermore, formative assessment encourages students to take responsibility for their own learning, providing a basis for understanding criteria for evaluation and the ability to reflect on their own performance in relation to those criteria.

Moreover, sound formative assessment practice is likely to improve the overall quality of classroom instruction. Specifically, teachers are able to gauge student performance in real time and adjust their instructional strategies, learning goals, and future assessments accordingly. In addition, students receive immediate feedback on their work in order to take steps to improve any weaknesses before it's too late. Consequently, the improved academic and social-emotional outcomes cultivated by a high-quality formative assessment system will provide Our Kids with more opportunities for success beyond high school.

It is important to note that by recommending more resources and attention be dedicated to improving the quality and effectiveness of formative assessment, we are not implying that large-scale summative assessments should be eliminated. Students can benefit from both classroom-based assessment and accountability assessments that are well aligned to well-specified learning goals (Stiggins, 2003; Valencia, 2008). In the era of NCLB, large-scale summative assessments will continue to be important for monitoring school progress and facilitating education reform. Crooks (1988) suggests that while less frequent evaluations for summative purposes should focus on describing what students can and cannot do, the majority of evaluation activity in education should be directed toward providing students with feedback to

facilitate their learning. Furthermore, future explorations in assessment should consider how educators can take advantage of external summative assessments to identify ways in which the results can help improve student performance (Boston, 2002).

Connections across the Learning System

The research findings and recommendations for assessment practice discussed in this review are intimately connected with other areas of the Learning System. Because high-quality formative assessment practice is essentially indistinguishable from pedagogy, any change in one will inevitably influence the other. Formative assessment practice is incompatible with instructional practice focused on "teaching to the test" and student reproduction of lower-level facts. Formative assessment practice must exist in a pedagogical environment wherein teachers practice intellectually demanding and authentic instruction for the purpose of transfer of knowledge to new, out-of-school learning situations. Furthermore, formative assessment practice must correspond with pedagogy that is sensitive to students' diverse cultures, languages, and backgrounds.

Because high-quality formative assessment is seamlessly integrated with instruction, assessment and curriculum go hand in hand. Formative assessment practice should be integrated with curricula that account for individual differences among students, allowing students to learn at their own pace with individualized lesson plans and assessments. Moreover, formative assessment should be integrated with curricula that have real-world applications and intellectually demanding content. If a curriculum has a heavy content focus, formative assessments must be aligned with and tailored to that focus.

Lastly, formative assessment cannot function without strong support from leadership. School leaders must have a high degree of assessment literacy in order to support and monitor teachers' practice. Teachers need the flexibility to focus on intellectually challenging and differentiated learning targets rather than practice for the state assessment. Thus, school leaders must balance classroom-based assessment for learning with their accountability goals and might consider ways of using the results of state tests formatively to improve learning outcomes.

References

American Council on Education. (2008). *Minorities in higher education 2008 twenty-third status report.* Washington, DC: Author.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with E-rater V.2. *Journal of Technology, Learning, and Assessment, 4*(3), 3–30.

Ayala, C. C., Shavelson, R. J., Ruiz-Primo, M. A., Brandon, P. R., Yin, Y., Furtak, E. M., et al. (2008). From formal embedded assessments to ref ective lessons: The development of formative assessment studies. *Applied Measurement in Education, 21*(4), 315–334.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*(2), 213–238.

Bass, K. M., & Glaser, R. (2004). *Developing assessments to inform teaching and learning.* (CSE Report 628). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Bentz, J., & Pavri, S. (2000). *Curriculum-based measurement in assessing bilingual students: A promising new direction*. Los Angeles: Sage Publications.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 9–21.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–144.

Boston, C. (2002). *The concept of formative assessment.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Boyer, E. (1995). *The basic school: A community for learning.* Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Brimijoin, K., Marquissee, E., & Tomlinson, C. A. (2003). Using data to differentiate instruction. *Educational Leadership, 60*(5), 70–73.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement. *Applied Measurement in Education, 10*, 161–180.

Brookhart, S. M. (2008). *How to give effective feedback to your students.* Alexandria, VA: ASCD.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. Assessment in Education: *Principles, Policy & Practice, 15*(1), 3–17.

Brunot, S., Huguet, P., & Monteil, J. (2000). Performance feedback and self-focused attention in the classroom: When past and present interact. *Social Psychology of Education, 3*, 277–293.

Callingham, R. (2008). Dialogue and feedback: Assessment in the primary mathematics classroom. *Australian Primary Mathematics Classroom, 13*(3), 18–21.

Cauley, K. M., Pannozzo, G., Abrams, L., McMillan, J., & Camou-Linkroum, S. (2006). *The relationship between classroom assessment practices and student motivation and engagement*. Richmond, VA: Virginia Commonwealth University.

Center for Comprehensive School Reform and Improvement (CCSRI). (2006). U*sing classroom assessment to improve teaching.* Washington, DC: Author.

Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership, 60*(1), 40–43.

Chappuis, S., Chappuis, J. & Stiggins, R. (2009). Supporting teacher learning teams. *Educational Leadership, 66*(5), 56–60.

Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review, 64*(1), 9–37.

Cowie, B. (2005). Student commentary on classroom assessment in science: A socio-cultural interpretation. *International Journal of Science Education, 27*(2), 199–214.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 5*8(4), 438–481.

Crumrine, T., & Demers, C. (2007). Formative assessment: Redirecting the plan. *Science Teacher, 74*(6), 28–32.

Darling-Hammond, L., Rustique-Forrester, E., & Pecheone, R. L. (2005). *Multiple measures approaches to high school graduation*. Stanford, CA: School Redesign Network.

Dichter, A., & Orlen, V. (2005). NSRF-New York. *Educational Horizons, 84*, 60–63.

Duncan, G. J., & Magnuson, K. A. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15*(1), 35–54.

Duschl, R. A., & Gitomer, D. H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment, 4*(1), 37–73.

Elawar, M.C., and Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behaviour a little rather than a lot. *Journal of Educational Psychology, 77* (2), 162–173.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., et al. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record, 52*, 187–201.

Even, R. (2005). Using assessment to inform instructional decisions: How hard can it be? *Mathematics Education Research Journal, 17*(3), 45–61.

Fontana, D., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology, 64* (3), 407–417.

Frederiksen, J. R., & White, B. J. (1997). *Ref ective assessment of students' research within an inquiry-based middle school science curriculum.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*(2), 449–460.

Furger, R. (2002). Assessment for understanding: Taking a deeper look. *Edutopia.* Retrieved from http://www.edutopia.org/assessment-for-understanding-taking-deeper-look

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Geva, E. (2000). Issues in the assessment of reading disabilities in L2 children-beliefs and research evidence. *Dyslexia, 6*, 13–28.

Glaser, R., & Silver, E. (1994). *Assessment, testing, and instruction: Retrospect and prospect* (CSE Technical Report 379). Pittsburgh, PA: National Center for Research on Evaluation, Standards, and Student Testing.

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied School Psychology, 89*(2), 248–262.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69–95.

Harlen, W. & Deakin, C. R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. In *Research Evidence in Education Library.* London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Haydel, A. M., & Roeser, R. W. (2002). *On the links between students' motivational patterns and their perceptions of, beliefs about, and performance on different types of science assessments: A multidimensional approach to achievement validation.* (CSE Technical Report). Los Angeles: UCLA/Center for the Study of Evaluation.

Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 8*9(2), 140–145.

Heritage, M., & Bailey, A. (2006). Educational assessment: Special issue. *Educational Assessment, 11*(3 & 4).

Herman, J. L., & Choi, K. (2008). *Formative assessment and the improvement of middle school science learning: The role of teacher accuracy.* (CRESST Report 740). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Holloway, J. H. (2003). Linking professional development to student learning. *The Challenges of Accountability, 61*(3), 85–87.

Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current DIrections in Psychological Science, 7*(3), 67–72.

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(2), 211–232.

Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79–97.

Kusimo, P., Ritter, M. G., Busick, K., Ferguson, C., Trumbull, E., & Solano-Flores, G. (2000). *Making assessment work for everyone: How to build on student strengths.* San Francisco: WestEd.

Landauer, T. K., Lochbaum, K. E., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice, 48*(1). 44–52.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3), 18–24.

McCurdy, B. L., and Shapiro, E. S. (1992). A comparison of teacher monitoring, peer monitoring, and self-monitoring with curriculum-based measurement in reading among students with learning disabilities. *Journal of Special Education, 26*(2), 162–180.

McDonald, B. & Boud, D. (2003). The effects of self assessment training on performance in external examinations. *Assessment in Education, 10*(2), 210–220.

McTighe, J., & Emberger, M. (2006). Teamwork on assessments creates powerful professional development. *Journal of Staff Development, 27*(1), 38–44.

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.). *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Lawrence Erlbaum Associates.

Newmann, F., Bryk, A., & Nagaoka, J. (2001). *Authentic intellectual work and standardized tests: Conf ict or coexistence?* Chicago: Consortium on Chicago School Research.

No Child Left Behind Act of 2001, Pub. L. No. 107–110, §115 Stat. 1425 (2002).

Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education, 5*(1), 85–102.

Popham, W. J. (2006). Assessment for learning: An endangered species? *Educational Leadership, 63*(5), 82–83.

Popham, W. J. (2007) Instructional sensitivity: Educational accountability's dire deficit. *Phi Delta Kappan, 89* (2), 149–155.

Popham, W. J. (2008). Formative assessment: Seven stepping-stones to success. *Principal Leadership, 9*(1), 16–20.

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice, 48*(1), 4–11.

Postholm, M. B. (2006). Assessment during project work. *Teaching & Teacher Education: An International Journal of Research and Studies, 22*(2), 150–163.

Rabinowitz, S., Roeber, E., Schroeder, C., & Sheinker, J. (2006). *Creating aligned standards and assessment systems.* Washington, DC: Council of Chief State School Officers.

Rieg, S. A. (2007). Classroom assessment strategies: What do students at-risk and teachers perceive as effective and useful? *Journal of Instructional Psychology, 34*(4), 214–225.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Inf uences on student cognitions about evaluation. *Assessment in Education, 9*(1), 81–95.

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*(2), 191–209.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.

Schmoker, M. (2002). Up and away. *Journal of the National Staff Development Council, 23*(4), 10–13.

Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning.* (CSE Technical Report). Los Angeles: National Center for Research and Evaluation, Standards, and Student Testing.

Shepard, L. A. (2001). *The role of classroom assessment in teaching and learning.* In V. Richardson (Ed.), *The Handbook of Research on Teaching,* (4th ed.). Washington, DC: American Educational Research Association.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66–70.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Shute, V. J., Hansen, E. G., & Almond, R. G. (2007). *An assessment for learning system called ACED: Designing for learning effectiveness and accessibility.* Princeton, NJ: Educational Testing Service.

Solórzano, R. W. (2008). High stakes testing: Issues, implications, and remedies for English Language Learners. *Review of Educational Research, 78*(2), 260–329.

Stanford, P., & Siders, J. A. (2001). Authentic assessment for intervention. *Intervention in School and Clinic, 36*(3), 163–167.

Starkman, N. (2006). Formative assessment: Building a better student. *T.H.E. Journal, 33*(14), 41–46.

Stiggins, R. J. (1998). *Classroom assessment for student success. Student Assessment Series.* Annapolis Junction, MD: NEA Professional Library.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5–14.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*(10), 758–765.

Stiggins, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*(1), 22–27.

Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership, 64*(8), 22–26.

Stiggins, R. J., Arter, J. A., Chappuis, J., & Chappuis, S. (2005). *Classroom assessment for student learning: Doing it right—using it well.* Portland, OR: Assessment Training Institute.

Stiggins, R., & Chappuis, J. (2008). Enhancing student learning. *District Administration, 44*(1), 42–44.

Stupski Foundation. (n.d.). R & D for innovation in public education. Retrieved from http://www.stupski.org/reforinnovation.htm

Torres-Guzman, M. E., Abbate, J., Minaya-Rowe, L., & Brisk, M. E. (2002). Defining and documenting success for bilingual learners: A collective case study. *Bilingual Research Journal, 26*(1), 23–44.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal, 22*(4), 389–404.

Valencia, S. W. (2008). *Understanding assessment: Putting together the puzzle. Beyond the Book.* Retrieved from http://www.beyond-the-book.com/strategies/strategies_100307.html

Weeden, P., Winter, J., & Broadfoot, P. (2002). *Assessment for learning: What's in it for schools?* London: Routledge.

Wiley, C. R. H. (2008). Traditional teacher tests. In T. L. Good (Ed.), *21st century education: A reference handbook: Vol. 1.* (pp. 431–442). Los Angeles: SAGE.

Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment,11*(3 & 4), 283–289.

Yeh, S. S. (2008). The cost-effectiveness of comprehensive school reform and rapid assessment. *Education Policy Analysis Archives, 16*(13), 1–32.

References Reviewed but Not Cited in This Report

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). London: Sage Publications.

Brookhart, S. M. (2008). Feedback that fits. *Educational Leadership, 65*(4), 54–59.

Clarke, B., Baker, S., Smolkowski, K., & Chard, D. J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education, 29*(1), 46–57.

Ellis, K. (2005). Learning one on one: Teaching kids to pursue their dreams—and more. *Edutopia.* Retrieved from http://www.edutopia.org/learning-one-one

Fuchs, L. S., & Fuchs, D. (2001). *What is scientifically-based research on progress monitoring?* Washington, DC: National Center on Student Progress Monitoring.

Fuchs, L. S., Fuchs, D., & Hamlett, C. L. (2007). Using curriculum-based measurement to inform reading instruction. *Reading and Writing: An Interdisciplinary Journal, 20*(6), 553–567.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Martinez, E. (2002). Preliminary evidence on the social standing of students with learning disabilities in PALS and No-PALS classrooms. *Learning Disabilities Research and Practice, 17*(4), 205–215.

Galvin, M., & Parsley, D. (2005). Turning failure into opportunity. *Educational Leadership, 62*(9).

Gulek, J. C., & Demirtas, H. (2005). Learning with technology: The impact of laptop use on student achievement. *ERS Spectrum, 23*(4), 4–20.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education, 35*(2), 213–224.

Hartl, S., Berger, R., Seidel, S., & Dawson, K. (2009). *Digital student work archive project.* Phase I report. Brookline, MA: The Nellie Mae Education Foundation.

Herman, J. L., Yamashiro, K., Lefkowitz, S., & Trusela, L. A. (2008). *Exploring data use and school performance in an urban public school district: Evaluation of Seattle public schools' comprehensive value-added assessment system.* (CRESST Report 742). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Hickey, D. T., Kindfield, A. C. H., Horwitz, P., & Christie, M. A. T. (2003). Integrating curriculum, instruction, assessment, and evaluation in a technology-supported genetics learning environment. *American Educational Research Journal, 40*(2), 495–538.

Howe, M. J. A. (1987). Using cognitive psychology to help students learn how to learn. In J. T. E. Richardson, M. W. Eysenck, & D. W. Piper (Eds.), *Student learning: Research in education and cognitive psychology.* Milton Keynes, England: Open University Press & Society for Research into Higher Education.

Johnson, D. W., Johnson, R., & Maruyama, G. (1983). Interdependence and interpersonal attraction among heterogeneous and homogeneous individuals: A theoretical formation and a meta-analysis of the research. *Review of Educational Research, 53,* 5–54.

Matsumura, L. C., Patthey-Chavez, G. G., Valdes, R., & Garnier, H. (2002). Teacher feedback, writing assignment quality, and third-grade students' revision in lower- and higher-achieving urban schools. *Elementary School Journal, 103*(1), 3–25.

Niemi, D., Baker, E. L., & Sylvester, R. M. (2007). Scaling up, scaling down: Seven years of performance assessment development in the nation's second largest school district. *Educational Assessment, 12*(3), 195–214

Parsley, D., & Galvin, M. (2008). Think systemically, act systematically. *AASA Journal of Scholarship and Practice, 4*(4), 4–10.

Pijanowski, L. (2008). Striking a balance: Georgia district adds assessments and transforms classroom practice. *Journal of Staff Development, 29*(4), 43–46.

Rubenstein, G. (2006). Vital signs: Learning is alive and well at Faubion Elementary. *Edutopia.* Retrieved from http://www.edutopia.org/vital-signs

Rubenstein, G. (2007). *Northern lights: These schools literally leave no child behind.* Retrieved from http://www.edutopia.org/chugach-school-district-reform

Ruiz-Primo, M. A., & Furtak, E. M. (2004). *Informal formative assessment of students' understanding of scientific inquiry*. (CSE Report 639). Los Angeles: Center for the Study of Evaluation/ National Center for Research on Evaluation, Standards, and Student Testing.

Sadler, D. R. (1983). Evaluation and the improvement of academic learning. *Journal of Higher Education, 54*(1), 60–79.

Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of national board certification. *American Educational Research Journal, 45*(3), 669–700.

Schimmel, B. J. (1983). *A meta-analysis of feedback to learners in computerized and programmed instruction.* Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Stecher, B. M., Rahn, M. L., Ruby, A., Alt, M. N., & Robyn, A. (1996). *Using alternative assessments in vocational education.* Washington, DC: U.S. Department of Education, the Office of Vocational and Adult Education.

Stiggins, R., & Chappuis, S. (2005). Putting testing in perspective: It's for learning. *Principal Leadership, 6*(2), 16–20.

Tomlinson, C. A. (2008). Learning to love assessment. *Educational Leadership, 65*(4), 8–13.

Wentzel, K. R., & Watkins, D. (2002). Peer relationships and collaborative learning as contexts for academic enablers. *School Psychology Review, 31*, 366–377.

Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., et al. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education, 21*(4), 335–359.

Appendix

Literature review method

In June 2008, the Stupski Foundation created a conceptual framework for the reinvention of American education. The framework identified seven essential components and focused on delivering 21st century college readiness for all students, but especially for "Our Kids," children of color and poverty. The Foundation explained that "graduating all students from high school with the knowledge and skills that qualify them as 'college ready' is the most meaningful and measurable way to increase life choices and options for all children, but most especially children of color and poverty" (About the Foundation, para. 3).

The Learning System includes four core teaching and learning components: Curriculum, Assessments, Pedagogy, and Supports. Surrounding these components, are three organizational components necessary to support the core: Leadership/Human Capital, Systems Diagnostics, and a Dashboard of College Readiness Indicators (College Readiness Learning System, n.d.).

The Foundation envisions convening a Design Collaborative, a cross-sector group of researchers, practitioners, and designers from inside and outside education, to "define, develop and continually improve" (Design Collaborative, n.d.) all of the components. To orient Design Collaborative members to the accumulated and maturing knowledge base related to each of the components and to children of color and poverty, the Foundation contracted with Mid-continent Research for Education and Learning (McREL). McREL conducted eight literature reviews—one on each of the components plus one on Our Kids—to identify and integrate theories and philosophical perspectives, issues, scientifically based research practices, unmet needs, and innovations relevant to designing one or more of the system components to accelerate learning for Our Kids.

This Appendix contains a description of the review method, including a general explanation of McREL's approach and descriptions of the particular procedures used for each phase of the review: identification of key hypotheses and research questions, literature search, identification and cataloguing of finds, and generating and communicating recommendations.

McREL's overall approach

Since the primary users of the reviews are the members of the Design Collaborative, the qualitative, iterative approach taken for the literature reviews sought to achieve the multiple goals of identifying emerging ideas, counterproductive orthodoxies, and promising practices relevant to the reinvention of the Learning System. Thus, eight research teams were assembled, each with one or more researchers familiar with the respective topic areas.

Qualitative approach. A *qualitative approach* shares several practices with those of *systematic reviews,* including comprehensive searches and transparency to reduce bias, but it differs with respect to inclusion/exclusion criteria. Systematic reviews emphasize explicit and a priori inclusion/exclusion criteria and criteria for evaluating the methodological quality of individual studies, carefully limiting the sources of evidence to support inferences about cause and effect relationships (Cooper, Hedges, & Valentine, 2009). The qualitative approach emphasizes diverse sources and types of evidence and knowledge to support a broader base of inferences (Pope, Mays, & Popay, 2007; Suri & Clarke, 2009).

The qualitative approach is particularly well-suited to the review's purpose and audience because the Design Collaborative needs both empirical studies and other literature to identify possible innovations for the current education system. An assumption underlying the Foundation's work to fundamentally reinvent American education is that the current system fails to deliver college readiness for all students, especially Our Kids. This assumption is supported by research indicating that students of color and in poverty have low high school and college graduation rates, and research from the last two years shows that college graduation rates for minority and poor students have further declined (American Council on Education, 2008). Therefore, a priority for the Foundation's work is to identify innovations that have not yet been studied, with the intent to evaluate their effectiveness. Literature specific to innovations is found outside the traditional scientific or academic journals.

Inclusive approach. McREL researchers adopted an inclusive approach, searching for and including phenomenological reports describing the experiences of Our Kids in and out of school and documenting the challenges and successes of their teachers and educational leaders. The researchers included literature on innovative, emerging models and untested ideas, as well as reports on mature, well-specified models with experimental evidence of effectiveness. Relevant quantitative research literature included correlational and experimental studies and meta-analytic reviews. Narrative reviews of research were included, as were policy briefs and position papers produced by opinion leaders and professional organizations. Literature sources included the World Wide Web, peer-reviewed journals, and practitioner magazines. Each document was identified by type of literature and evaluated in terms of the quality of the supporting evidence. Care was taken to draw only those inferences appropriate to the quality of the evidence.

McREL researchers judged the quality of the evidence in the context of the type of literature or study design and in relation to its relevance to answering particular questions. Guidance from Pope, Mays, and Popay (2007) on conducting reviews in the field of health research supports this approach:

> The inclusion of diverse sources of evidence in a review does not mean abandoning the rigor of a systematic review, but it does mean judging the quality of evidence in context and defining the relevance of evidence to answering specific questions, rather than defining some forms of evidence as intrinsically, and universally, of lower quality than others. (p. 1)

Each research team followed the five or six phases of any review process relevant to a quality knowledge synthesis (Cooper, Hedges & Valentine, 2009; Suri & Clarke, 2009). Table 1 (see p. 47) provides a side-by-side comparison of the phases of a systematic review of research (Cooper, Hedges & Valentine, 2009), a qualitative review (Suri & Clarke, 2009), and McREL's approach to this review.

Each team began by drawing from pertinent philosophical and theoretical literature and preliminary discussions with the Foundation to formulate hypotheses and research questions. Each team conducted extensive searches to find as much relevant literature as possible in order to include literature from the scientific and academic journals as well as literature from harder-to-find, cutting edge innovators. Additionally, teams revisited databases and alternative sources to purposefully search for additional literature written by authors identified by one or more stakeholders or to fill conceptual gaps that became apparent during the identification and cataloguing of findings and generating and communicating recommendations phases.

The phased process was iterative (Cooper, 2009) ref ecting new understanding and insights as the search, analysis, interpretation, and discussions between component teams and between the Foundation and McREL progressed toward conceptual clarity and the exhaustion of new search hits.

Table 1: Phases of a literature review

Phase	Cooper, Hedges & Valentine (2009, p. 8)	Suri & Clarke (2009, p. 414)	McREL 's approach
1	Problem formulation	Drawing from pertinent philosophical and theoretical discussions	Identification of key hypotheses
2		Identifying an appropriate purpose	Identification of research questions
3	Data collection	Searching for relevant evidence	Literature search
4	Data evaluation	Evaluating, interpreting, and distilling evidence	Identification and cataloguing of findings
5	Analysis and interpretation	Constructing connected understanding	Generating and communicating recommendations
6	Public presentation	Communicating with an audience	

The number of documents included in each team's review was extensive, and the types of literature varied representing the experiential knowledge of a diverse group of stakeholders, including researchers, teachers, administrators, program developers, and leaders and scholars at the local and national levels.

Team approach. Teams were composed of researchers and practitioners with different areas of expertise. Teams met weekly, and team leaders from across teams met biweekly. Meetings were used to update other individuals and teams and share resources, pose and address questions, challenge assumptions, provide guidance on interpretation of evidence, open up new areas of consideration, clarify boundaries and overlap between system components, consider alternative perspectives, and develop connected understanding.

Identification of key hypotheses and research questions

McREL teams began by clarifying terms, relationships, and the conceptual scope of each review. Teams read and discussed a document produced during the Foundation's strategy definition process, *Research Guide for CRLS: Outline of Research Questions for Each Component of the CRLS* (n.d.). Included in this Guide were preliminary questions for each literature review. Teams previewed relevant literature, confirmed that the questions could be answered by the extant knowledge base, and posed additional questions when important issues related to accelerating learning for students of color and poverty were identified in the literature but missing in the Guide. The revised set of questions for each system component and Our Kids was reviewed and refined during ongoing dialogue between the Foundation and McREL.

Literature search

Multiple searches were conducted in a phased approach to identify as much literature as possible related to each system component and Our Kids. Teams conducted searches using multiple

bibliographic databases: Academic Onefile, Academic Search Premier, Educators Reference Complete, ERIC, JSTOR, Proquest, and PsychInfo. Teams also conducted manual searches of journal and book tables of contents and reference lists of articles. Additional searches were conducted specifically to identify recent experimental and other research and reviews on the efficacy of interventions for accelerating learning of students of color and poverty. These searches were conducted by visiting the U.S. Department of Education What Works Clearinghouse Web site (http://ies.ed.gov/ncee/wwc/reports/) and the Campbell Collaboration Library of Systematic Reviews Web site (http://www.campbellcollaboration.org/library.php). Relevant documents were identified on state education agency (SEA) Web sites, and SEA officials were interviewed or named as seminal authors or sources of models that had been developed and implemented to monitor and accelerate learning of Our Kids.

Each team identified and used key terms and synonyms relevant to the topic for searching. Searches were conducted for literature published in the most recent 10 years (1998–2008); however, works by seminal authors and other recommended literature were included from outside these years. The search landscape varied for each team based on the topic and relevant sources; for example, while What Works Clearinghouse was a relevant source for the Pedagogy team, it was not a relevant source for the Leadership/Human Capital team. Internal review of search records and results led to additional leads on sources. Searching continued until all recommendations had been implemented and/or few new hits were identified.

Identification and cataloguing of findings

A coding protocol was developed and implemented to categorize the literature. Each team used the same protocol, adding categories and decision rules, as needed to organize the particular literature relevant to their topic. Each team leader and one or more members of each team were trained on the decision rules in the coding protocol and provided follow-up support to resolve uncertainties in its application. Team leaders periodically conducted quality assurance reviews of completed coding sheets and updated the protocol as needed during weekly team leader meetings or discussions with the Foundation. The coding protocol included identifying the following information:

- Full APA reference citation
- Category of literature (i.e., primary and secondary relevance)
- Type of literature (e.g., quantitative study, policy brief, program description)
- Locale
- Outcome
- Grade level
- Program or innovation name and description
- Main findings or points
- A recommendation for or against summarizing and including the selection in an annotated bibliography.

In addition, component teams added to the protocol by categorizing relevance to particular parts of their conceptual model or concept map.

Guidelines were developed and used by teams to identify counterproductive orthodoxies, unmet needs, next practices, promising practices, and best practices based on type of literature and quality of evidence. These were defined in the following ways:

- *Counterproductive orthodoxies:* Conventional ways of providing education which may be impeding success of Our Kids
- *Unmet needs:* Areas where Our Kids are not yet well served by the current system of education
- *Next practices:* A program or practice that needs to developed, adapted, invented, and tested in response to an unmet need related to accelerating learning for Our Kids
- *Promising practices:* Practices based on research but not supported by rigorous efficacy data from randomized controlled trials
- *Best practices:* Practices demonstrated by one or more randomized controlled trials to be effective in improving outcomes for Our Kids

The research team reviewing the college readiness component of the Learning System employed a slightly different process. Rather than using the categories above, this team reviewed literature on college readiness and categorized findings into four essential areas as defined by the Foundation and Conley (2007): cognitive strategies, content knowledge, academic behaviors, and contextual skills.

Component teams met weekly to discuss and categorize findings and to develop a conceptual map of the insights gained from the literature summaries and review. Teams used different conceptual mapping tools (e.g., SmartArt) to organize the insights (findings) and presented and discussed their respective maps at cross-team meetings. Features common across teams' concept maps were identified and a standard framework developed. Teams arranged findings onto the concept maps, identifying conceptual gaps and conf icting or discrepant findings, and returned to searching and reviewing to fill in the gaps and resolve or explain discrepant findings. The conceptual maps served as an organizing framework for report construction.

Generating and communicating recommendations

Working collaboratively, component teams drew conclusions from the insights (findings) derived from the review and identified potential options and recommendations for each component of the system. Teams used an iterative process of identification, reviewing for validity against the knowledge base, and further refinement until they determined they had identified the most promising options and that each was informed by the existing knowledge base.

Team leaders used the outcomes of team discussions and cross-team discussions, literature summaries, and the researcher's own review and integration of the literature to write a draft report of the findings. Draft reports were reviewed by knowledgeable internal experts and revisions in search strategies, interpretations of findings, and/or conclusions were made. Revised reports were reviewed by the Foundation and other outside reviewers prior to final revisions and production.

Although the wide-ranging literature searches produced reports on extensive baseline information related to Our Kids and each system component, the reports are living documents. As living documents, they bridge the creative and scientific enterprises of the past and present, and we envision the need to return to some of them for updating, extending, and drilling-down in the future.

References

About the Foundation. (n.d.). Retrieved from http://www.stupski.org/about_the_foundation.htm

American Council on Education. (2008). *Minorities in higher education 2008 twenty-third status report.* Washington, DC: Author.

College Readiness Learning System. (n.d.). Retrieved from http://www.stupski.org/college_readiness.htm

Conley, D.T. (2007). *Toward a more comprehensive conception of college readiness.* Eugene, OR: Educational Policy Improvement Center.

Cooper, H. (2009). Hypotheses and problems in research syntheses. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis (pp. 19–35).* New York: Russell Sage Foundation.

Cooper, H., Hedges, L., & Valentine, J. (2009). *The handbook of research synthesis and meta-analysis.* New York: Russell Sage Foundation.

Design collaborative. (n.d.). Retrieved from http://www.stupski.org/design_collaborative.htm

Pope, C., Mays, & Popay, J. (2007). *Synthesizing qualitative and quantitative health research: A guide to methods.* New York: McGraw Hill Education.

Research guide for CRLS: Outline of research questions for each component of the CRLS. (n.d.). San Francisco: Stupski Foundation.

Suri, H. & Clarke, D. (2009). Advancements in research synthesis methods: From a methodologically inclusive perspective. *Review of Educational Research, 79*(1), 395–430.

Stupski Foundation

101 2nd Street, Suite 1100
San Francisco, CA 94105
Tel: 415.644.4800
Fax: 415.644.4801