Toward Balanced Mathematics Assessment: Some Findings from the Field

Pamela L. Paek, Ph.D.

Charles A. Dana Center, The University of Texas at Austin, Austin, Texas, United States

Paper for presentation in the Research and Development in Assessment and Testing in Mathematics Education Topic Study Group (TSG 36) at the eleventh International Congress on Mathematical Education, Monterrey, Mexico, July 6-13, 2008

Introduction

This paper discusses what the role of assessment and testing in mathematics education is, what that role could be, and how understanding the uses of assessment in the field can contribute to effective change in mathematics assessment practice. To begin, the paper gives an overview of the landscape of assessment and testing in the United States, where currently, a great deal of effort is placed on assessment *for* learning, while assessment *of* learning is required as part of the American educational accountability system. Next, the paper examines how school districts are using assessments to improve instruction and thus attend to student specific learning needs, as well as to tailor professional development opportunities for teachers.

The school districts profiled in this study are spending time on defining assessment *for* learning and redefining how formative, benchmark, and interim assessments can guide teachers in modifying their instruction. This redefinition is based on the concept of using evidence of student learning to adjust instruction as it is taking place, so that teachers can immediately address students' learning needs. The introduction and use of assessment *for* learning in these districts is intended to build balanced assessment systems that allow teachers to use the many layers of assessment to best teach their students. In addition, teachers in these districts are regularly involved in professional learning communities that reinforce and continue to expand on their learning of the best uses of assessment. Results show positive changes in teachers' attitudes about the use of assessments and in one case, a direct improvement in student learning.

Theoretical Framework

While the current trend in assessment is moving toward assessment *for* learning (Black, Harrison, Lee, & Marshall, 2003; Black & Wiliam, 1998; Brookhart; 2005; Leahy, Lyon, Thompson, & Wiliam, 2005), the United States still has a strong tradition of assessment *of* learning, particularly given that such assessment is required by the federal *No Child Left Behind Act*. Given that no single type of assessment provides all the answers to understand and improve student learning (National Education Association, 2003), we need to find a way to incorporate the types of assessments currently in use in more balanced and coherent ways. Balanced assessment systems—that use both assessment *for* learning and assessment *of* learning—maximize the role that assessment can play in improving mathematics teaching and learning (Stiggins, 2002) and minimize the role of polarizing rhetoric that demonizes one or another kind of assessment. It is a *both/and* versus *either/or* instructional design paradigm. The availability and transparency of a balanced assessment system enable teachers to learn what their students know, assess student achievement, and use these assessment results as part of ongoing efforts to improve instruction and consequently improve student learning (Fullan, Hill, & Crévola, 2006; National Education Association, 2003). This paper examines a handful of U.S. school districts that are creating balanced assessment systems in mathematics and describes some of the ways that these districts think about mathematics assessment and work with teachers to use these ideas in their instruction.

Research on the effects of a balanced assessment systems shows that they correlate with impressive gains in student learning. Black and Wiliam (1998) showed 0.5 to 1.0 standard deviation gains, while Meisels and colleagues (2003) showed gains of 0.75 to 1.5 standard deviations, and Rodriguez (2004) showed 0.5 to 1.8 standard deviation gains. To put these gains in

standard deviations in context, a 1.0 standard deviation gain is equivalent to 35 percentile points on a standardized test.

One challenge for practitioners is to learn how to create such balanced assessment systems. Aligning standards, assessments, and instruction is a crucial piece in standards-based reform (Ainsworth & Viegut, 2006; Webb, 1997; 1999). Webb's process (1997; 1999) was used by the districts profiled in this study to align standards, assessments, and instruction. Webb's process helps to inform both teachers' judgments about what content to focus on and district articulation of effective instructional strategies across grades. When applying Webb's system, teachers learn about the depth of knowledge required for a student to successfully answer each content item. Webb also provides a framework for balancing the representation of content across an assessment, helping teachers think about the assessment's emphasis on different topics, instructional activities, and tasks that should be represented in instruction. Using frameworks such as Webb's provides teachers with ways of becoming better informed about assessments *of* learning and using that knowledge in improving their assessments *for* learning in their classrooms.

Cobb and colleagues (2003) argued that professional development is more effective when it is tailored to teachers' specific needs and constraints, and other researchers (Cohen & Hill, 2000; Ingvarson, Meiers, & Beavis, 2005) have argued that one-day professional development meetings are not as effective as a series of meetings that span the school year. Professional development that occurs throughout the school year provides teachers with opportunities to form and engage in professional learning communities, which in turn enable them to work continuously toward a common understanding about student learning and to establish a set of concrete and measurable instructional goals (DuFour, DuFour, Eaker, & Many, 2006). Professional learning communities enable teachers to be involved as participants in their own learning. As they make changes in their professional practice, they move from being *reactive* learners who are told about certain strategies but do not have opportunities to discuss or practice them during their professional development meetings (Garet, Birman, Porter, Desimone, & Herman, 1999) to being *active* learners who engage in ongoing discussion and collaboration around improving instruction and student learning. It is important, then, that when changing the role of assessment, strategies such as professional learning communities to help teachers become enculturated in the new thinking and processes are built in.

Method

This study profiles three U.S. districts (two urban and one rural with students spanning grades K–12) and their use of assessments in mathematics. District mathematics and assessment staff were interviewed about their approaches to assessment, in light of accountability and how they were helping teachers use assessments effectively for improving their mathematics teaching.

Results

Lamoille South Supervisory Union (LSSU)

LSSU comprises three districts around Morrisville, Vermont. Even with three districts, the population of students is only around 3,800. Students are predominantly white and do not have

limited proficiency in English. Approximately 19% of LSSU students are classified as economically disadvantaged.

LSSU's main goal is to create a balanced assessment system in mathematics that will coexist with and complement the overall local assessment system. At the core of this work is a focus on helping teachers develop "assessment literacy," which LSSU defines as knowing how to gather dependable information about student achievement and using that information to inform instruction. LSSU wants teachers to think of assessment as part of instruction rather than as a separate activity. LSSU believes that professional development is a leverage point for systemic change in the district, since it enriches teachers' understanding of the interrelationship between content knowledge and assessment. LSSU professional development workshops are intentional and specific to Vermont mathematics standards, focusing on the content knowledge needed in mathematics. However, 75% of the professional development focused on balanced assessment work, specifically assessment literacy, while 25% is specific to mathematics content.

The district mathematics assessment system is being developed in three phases.

Phase 1. In spring 2006, phase I began with the development of four separate grade-span assessments that are summative in nature: K–2, 3–5, 6–8, and 9–10. The K–2 assessments measure additive reasoning; grades 3–5 assessments measure multiplicative reasoning; grades 6–8 assessments measure proportional reasoning; and grades 9–10 assessments measure algebraic reasoning. These grade-span assessments are administered in the second semester of the last level of each grade span (i.e., grades 2, 5, 8, and 10).

Professional development for Phase 1 began with a work session followed by a two-day workshop on assessment literacy. LSSU reported that the impact of the workshop was monumental—even a 16-year veteran teacher said he had to rethink the way he assesses students, realizing he had spent years doing it "wrong." LSSU then held a three-day mathematics summit in August 2006. Teachers spent part of each day either learning or reviewing content alignment using Webb's (1997, 1999) concepts of *depth of knowledge* and *balance of representation* and then aligning assessments in these two areas.

In Webb's system, teachers learn about the depth of student knowledge required in each assessment content item. Webb identifies four levels of depth of knowledge: At Level 1, students are expected simply to recall; at Level 2, students must use skills and knowledge of some procedures; Level 3 requires students to engage in strategic thinking, since the assessment items have some complexity, generally taking more than 10 minutes to solve; and Level 4 items involve nonroutine thinking and solving with multiple steps. Webb also provides a framework for balance of representation of content across a test, which helps teachers think about the emphasis of different topics, instructional activities, and tasks that should be represented in instruction as well as assessment.

In teams, teachers used the depth of knowledge paradigm to review and rate the assessment items they aligned, and spent time drafting assessments using balance of representation to ensure that a variety of content and item difficulties were represented. After this training, teachers can build better assessments, since they learned to choose items based on reliability issues,

psychometric properties, and relation of the items to Vermont's mathematics standards. In short, teachers have learned how to choose items that measure what they purport to measure.

Phase 2. In Phase 2, which began in summer 2007, LSSU mathematics teachers started to develop grade-level screening assessments for all grades from Grade 1 through Algebra II. These assessments were to be administered as pre- and posttests at the beginning and end of the 2007–2008 academic year.

LSSU wants to use Levels 1 and 2 complexity assessments (basic skill type items) as pre- and postmeasures of mathematics for the Phase 2 work. Items from Levels 1 and 2 represent the skills students need to successfully solve more complex mathematics item types. Items from Levels 3 and 4 would be used for diagnostic assessment, as an assessment for learning tool that breaks down concepts so that teachers can learn from what students conceptually understand. These items can be used in the grade-span assessments as well as in the common assessments to be developed and used districtwide, starting in Phase 3 (summer 2008).

Phase 3. In Phase 3 (summer 2008), mathematics teachers will develop common grade-level assessments for all students using similar processes for depth of knowledge and levels of complexity. Items on these assessments will include both assessment of learning items (e.g., levels 1 and 2 items) for teachers to see what concepts students mastered and assessment for learning items (e.g., levels 3 and 4) for teachers to see where students need more support and instruction.

Learning Communities. LSSU's mathematics assessment system incorporates the use of ongoing and embedded professional development structures, including professional learning communities, to help increase mathematics teachers' professional capacity. The teachers are learning more about the development and use of assessments, and will learn how to create and analyze assessments across all phases of the system's development. Teachers are involved in writing assessment items because LSSU leaders believe that teachers need to understand what is expected at the district level in order to affect instruction at the classroom level. They also believe that teachers need to be involved in conversations that help them reflect on their practice. In developing items, teachers are talking about the purposes of formative assessments and summative assessments and learning how to make judgments about student learning depending on the type of student work or data they have available.

Summary: LSSU believes that the development and implementation of a balanced assessment system requires targeted professional development and active participation of teachers to secure buy-in and successful implementation. By design, teachers are developing all of the local assessments for all three phases of the mathematics work. LSSU believes that when more than one teacher uses an assessment, teachers can collaborate to analyze the results and then plan interventions and modifications. Having such information and practices will allow teachers to better serve their students in improving their understanding and performance in mathematics from grades K–12.

Cleveland Municipal School District (CMSD)

CMSD serves grades K–12 with about 58,000 students in Cleveland, Ohio. The majority of students in Cleveland Municipal School District are black (~70%), followed by white (~17%) and

Hispanic (~10%). Since 2002, at least 98.8% of students has been classified as economically disadvantaged. Only a small percentage of Cleveland's students is classified as having limited proficiency in English (4%).

During 2005–2006, CMSD piloted Keeping Learning on Track (KLT), a formative assessment program developed by Educational Testing Service (ETS) with teachers from the 10 lowest performing K–8 schools on the state's NCLB assessment, the Ohio Mathematics Achievement Test. The focus of KLT is on using assessment for learning—that is, assessing student learning continually, without tests. The practice of using assessment for learning is based on the concept of using evidence of learning to adjust instruction while it is taking place, so that teachers can immediately address students' learning needs. Regardless of what techniques teachers decide to use, these types of formative assessment checks can provide teachers with the information they need to change their daily practice, which may result in large changes in teacher pedagogy, the classroom culture, and student learning.

Since teachers' instructional styles vary, KLT provides a variety of ways for teachers to measure student learning on the fly. One example is the thumb technique: The teacher asks the students how well they understand a concept, and students can put their thumbs up to represent good understanding of the concept, down to indicate confusion, or somewhere in the middle to indicate that they understood it but still need further instruction or practice. Teachers can then use these student responses to make instructional adaptations right at that moment.

Learning Communities. The framework of KLT is to have teachers meet regularly in teacher learning community (TLC) meetings to reinforce these ideas and build upon the techniques, strategies, and ideas behind the KLT program. In summer 2005, participating teachers from these 10 schools were given a one- to three-day introduction to assessment for learning by ETS KLT developers. The introduction included modeling of several different assessment techniques teachers could use during lessons.

At the end of the introduction, teachers wrote individual action plans outlining the specific techniques within each strategy they would like to implement in their classrooms and identifying what current practices they would relinquish in order to make time for the new techniques. The teachers were asked to complete journal entries describing their experiences as they implemented the new techniques. KLT is not designed specifically for any one content area, focusing instead on improving the pedagogical skills of teachers; however, Cleveland decided to focus their use of KLT strategies on mathematics, since the district mathematics coordinator also helped incorporate mathematics content in the monthly TLC meetings.

In the monthly meetings, teachers discussed the implementation of assessment for learning in their classrooms and refined their understanding of KLT techniques. One of the goals of the TLC meetings was to create a comfortable community in teachers could learn from one another and standardize the ways they viewed teaching and learning.

At the start of each meeting, teachers would gather in small groups for a "How's It Going?" check-in to discuss the techniques they had been using and whether the techniques were successful. ETS and district staff would ask teachers about certain techniques they may not have tried and what supports they would need to use KLT more in their classrooms. Teachers shared

their successes with the entire group and received feedback on what they were doing; then they spent the end of every meeting revising the action plans they had developed in the summer and documenting their intent to try new techniques.

The TLCs included mathematics teachers from a range of grade levels in K–8. This allowed for diverse discussions across topics and helped create a support network comprising teachers in similar situations with similar students. Through regular sharing of how they were implementing and refining KLT practices, as well as presentations of new techniques and ideas relevant specifically to mathematics content, the TLCs helped teachers absorb these new ideas into their own practices.

Each monthly meeting also focused on a different aspect of assessment for learning, such as how to look at student grades and make inferences, how to assess students' mathematics work, or how to gather evidence that could demonstrate student learning. Every meeting reinforced the idea that increasing teacher capacity is important for improving student performance, thereby assuring teachers that learning the new techniques and reflecting on their own practices were worthy enterprises.

Summary. Overall, schools in Cleveland that participated in the KLT program showed substantially greater gains on the mathematics summative assessment than those that did not. In March 2005, the mean score of the ten participating schools on the Ohio Mathematics Achievement Test was 379.85; in March 2006, it was 388.29. This is a mean gain of 8.44 scale score points from one year to the next. For all the other K–8 schools in the district, the mean Ohio Mathematics Achievement Test scaled score was 388.94 in March 2005 and 391.16 in March 2006, which translates to a mean gain of 2.22 scale score points. The effect size for the KLT schools in grade 7 was 0.11; in grade 8, it was 0.18. To put these effect sizes in context, it is useful to consider the fact that the renorming of the SAT-9 achievement test showed the differences between the two grades to be around 0.5 standard deviations in mathematics. Thus, the effect size for a seventh grader might be seen as equivalent to a fifth of a year's growth, and for an eighth-grader, it would be closer to over a third of a year's growth, which certainly reflects a substantial gain associated with schools participating in KLT compared with other schools in these grades.

Anchorage School District (ASD)

ASD serves grades K–12 with enrollment of around 50,000 students in Anchorage, Alaska. The majority of students are white (~55%), followed by Alaska Native and American Indian (~13%), Asian American and Pacific Islander (~11%), black (~8%), Multi-Ethnic (~7%), and Hispanic (~6%). More than 15% of Anchorage students are classified as having limited proficiency in English, and about 34% are classified as economically disadvantaged.

ASD began reforming its mathematics program based primarily on the recommendations of two researchers. First, ASD aligned assessments with content standards (Cohen, 1995), and provided clear descriptions of the skills and knowledge that will be assessed so that teachers can focus on specific knowledge and skills (Popham, 2003). In addition, as recommended by Popham, ASD produces standard-by-standard reports of each student's; by linking each item on an assessment to a specific grade-level expectation in mathematics, teachers can see which concepts individual students have mastered.

Anchorage decided to target its mathematics improvement efforts in the schools with less than the district average proficiency. The district identified 28 schools—22 elementary schools and 6 middle schools—that fit these criteria. In summer 2006, the district's assessment and evaluation staff analyzed the areas of weakness on the standards-based mathematics assessment for each school and provided their analyses to the district's Curriculum and Instructional Support Department. Mathematics curriculum specialists from the department, known as mathematics support teachers, spent the summer reviewing the data and planning 90-minute inservice trainings that would be provided at each school site.

Each mathematics support teacher worked with an assigned set of schools to offer training and follow-up support during the year. The assessment and curriculum staff worked together to develop growth charts for each school, showing the mean scale score comparison for each mathematics strand in the state standards by grade and comparing the 2004–2005 and 2005–2006 data. During the initial training at the beginning of the year, they gave the teachers charts and graphs with data that identified mathematics strand weaknesses at their school and grade level so the teachers could understand where they would need to focus instructional improvements.

Aligning assessments with content standards. Mathematics support teachers aligned K–6 mathematics curricula, even though students in K–2 are not tested, because they wanted to build an aligned curriculum that would support growth across strands and grades. In the initial training, the mathematics support teachers help the classroom teachers build target grade-level expectations within each strand; in follow-up meetings, they help teachers measure student progress during the course of a prescribed "instructional cycle." Mathematics support teachers train classroom teachers on the instructional cycle for measuring improvement in student performance on the focus mathematics strands. In the instructional cycle, teachers (or the mathematics support teachers and assessment and evaluation staff) identify the learning goals by noting on which mathematics strands their students performed poorly; they then plan instruction focused on the concepts in those strands. Next, teachers teach the material, assess student progress, and give feedback to students. Finally, teachers look at evidence of student learning to analyze areas students understand and areas where they still need help. This cycle continuously loops through the school year, as teachers reteach mathematics concepts, reassess student learning, and revise mathematics instruction.

Reporting by standards. As part of the training and follow-up meetings, assessment and evaluation staff show teachers how to analyze their students' and classes' performance using the district's Assessment Reporting System. The Assessment Reporting System is a comprehensive database that tracks individual students longitudinally, using a sequential growth model that matches an individual student's scores from one grade to the next. Cumulative data are available for individual student performance on district and state assessments. Teachers can access their own classroom rosters and look up the information on each of their students. If a student transfers to another teacher or school within Anchorage, that student is immediately placed electronically into the new classroom, so teachers have up-to-date access to all the student information they need.

While teachers can view their own classroom data, school administrators can view an entire school or any classroom within their assigned school. The Assessment Reporting System allows users to sort students based on proficiency by demographic information like race/ethnicity

or sex for an entire mathematics test or by strands. The four proficiency levels are color coded, giving teachers a visual snapshot of where students need the most help and allowing them to target specific students' weaknesses in each strand. The format of all data output has been customized based on teachers' requests, and the reports continue to be revised in response to teacher feedback. Since the system is homegrown and not an off-the-shelf product, Anchorage has the flexibility to continue to customize it so that is an optimum tool for teachers to use to inform their practice.

The Assessment Reporting System also features a grade-level expectation item bank. Teachers can pull items from this bank and use them to develop mini-assessments. The data from these items can then be used as part of the instructional cycle for measuring and improving student learning on different mathematics expectations.

Summary. ASD has connected its mathematics curriculum staff and its assessment and evaluation staff to collaborate on a set of professional development opportunities for connecting assessment explicitly to instruction. Anchorage has also developed a teacher-friendly database that offers teachers assessment data in new ways that allow them to measure change and growth for individual students over several academic years as well as for classrooms within an academic year. Although this is a new practice, it appears that Anchorage has built an infrastructure for connecting mathematics assessments to instruction, using key components from research that will encourage successful implementation. With the continuous training opportunities, district staff can gauge how well teachers are doing with this new practice and teachers can refine the ways they analyze and respond to the data.

Discussion

The school districts profiled in this study are spending time building and utilizing balanced assessment systems in three primary ways: (1) learning how to better use assessment *for* learning as part of their everyday practice, (2) continuing to cultivate the uses of data from various assessments to inform instruction, and (3) building in learning communities in which teachers are regularly reinforced about how to refine their practice.

Incorporating assessment *for* learning. Districts are defining assessment *for* learning and redefining what data from formative, benchmark, and interim assessments can provide for teachers to help them refine their instruction. This redefinition is based on the concept of using evidence of student learning to adjust instruction as it is taking place, so teachers can immediately address students' learning needs. Leahy, Lyon, Thompson, and Wiliam (2005) describe this work a minute-to-minute assessment that provides strategies for teachers to use as part of their instruction. In addition, districts spend time developing teachers' assessment literacy, which provides a framework for seeing assessments and instruction on the same continuum and using the variety of evidence of student learning in ways that can specifically address student learning needs.

Using assessments to inform instruction. Teachers are also learning about how to use and understand assessment data in various ways. In one district, teachers are involved in writing assessment items and tasks, and thus talking about assessments *for* and *of* learning because they are talking about formative, interim, and summative assessments. These teachers are learning how to make judgments about student learning on the basis of the type of student work or data they have available. In another district, teachers are given access to data in ways they have never been

provided before, so they can link student performance on various content standards to specific assessment items, which then informs how they may need to modify instruction. Having comprehensive databases allows teachers to look at individual students in their classes and use assessment data and items to find ways to specifically target student individual needs.

Building in professional development and learning communities. A cross-cutting theme in this research is that districts are implementing balanced assessment systems through sustained ongoing professional development for their teachers focused on using assessment to inform instruction. In districts that do a good job connecting the instructional design dots, professional development is used strategically to help teachers learn about new assessment practices as well as about research-based strategies to improve their teaching practices.

Collectively, these three methods help teachers see how they can use assessments as part of their regular practice. Teachers see that assessment of learning at the end of the school year is just another assessment built into the educational system that help inform how well students are learning. Assessments of learning are no longer considered separate entities that teachers need to prepare for separately from their regular instruction. Rather, such summative assessments are considered as they are intended: a reflection of student knowledge for a specific set of standards, and as part of the instructional continuum.

This work shows some of ways districts see the role of assessment systems and how they are using assessments to help shape the way they are improving mathematics teaching and learning. The goal of this work is three-fold: (1) to provide a forum for practitioners to learn from one another through these different approaches, (2) to help practitioners see assessments *of* and *for* learning as complements to one another and as an embedded aspect of instruction, and (3) to contribute ideas to the ways we view the role and practice of assessment in mathematics, nationally and internationally.

References

Ainsworth, L. B., & Viegut, D. J. (2006). *Common formative assessments: How to connect standards-based instruction and assessment.* Thousand Oaks, CA: Corwin Press.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice.* Berkshire, England: Open University Press.

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

Brookhart, S. M. (2005). *Research on formative classroom assessment*. Montreal, Canada: Paper Presented in the Symposium, Formative Classroom Assessment: Research, Theory, and Practice, at the Annual Meeting of the American Educational Research Association Montreal, Canada.

Cobb, P., McClain, K., Lamber, T. d. S., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher, 32*(6), 13–24.

Cohen, D. K. (1995). What is the system in systemic reform? *Educational Researcher*, *24*(9), 11-17, 31.

Cohen, D. K., & Hill, H. C. (2000). Instructional policy and classroom performance: Mathematics reform in California. *Teachers College Record, 102*(2), 294–343.

DuFour, R., DuFour, R., Eaker, R., & Many, T. (2006). *Learning by doing: A handbook for professional learning communities at work.* Bloomington, IN: Solution Tree.

Fullan, M., Hill, P., & Crévola, C. (2006). *Breakthrough*. Thousand Oaks, CA: Corwin Press.

Garet, M. S., Birman, B. F., Porter, A. C., Desimone, L., & Herman, R. (1999). *Designing effective professional development: Lessons from the Eisenhower Program.* Washington, DC: U.S. Department of Education.

Ingvarson, L., Meiers, M., & Beavis, A. (2005). Factors affecting the impact of professional development programs on teachers' knowledge, practice, student outcomes, and efficacy. *Education Policy Analysis Archives, 13*(10), 2–26.

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

Meisels, S. J., Atkins-Burnett, S., Xue, Y., Nicholson, J., Bickel, D. D., & Son, S-H. (2003, February 28). Creating a system of accountability: The impact of instructional assessment on elementary children's achievement test scores. *Education Policy Analysis Archives, 11*(9). Retrieved December 13, 2007, from http://epaa.asu.edu/epaa/v11n9.

National Education Association. (2003). *Balanced assessment: The key to accountability and improved student learning.* Washington, DC: Author.

Popham, W. J. (2003). Trouble with testing: Why standards-based reform doesn't measure up. *American School Board Journal*, *190*(2), 14-17.

Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education, 17*, 1–24.

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

Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states*. Research Monograph No. 18. Washington, DC: National Institute for Science Education.

Webb, N. L. (1997). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief 1*(2). Retrieved December 13, 2007, from http://www.wcer.wisc.edu/archive/nise/Publications/Briefs/Vol_1_No_2.