

# Current Developments in Assessment for Learning in Universities and High Schools in Michigan: Problems and Perspectives in Mathematics and Science Education

Steven W. Ziebarth

*Department of Mathematics, Western Michigan University, United States*

Herb Fyneweaver and George Akom

*Department of Chemistry and The Mallinson Institute, Western Michigan University, United States*

Katherine E. Cummings

*College of Education, Western Michigan University, United States*

Amy Bentz, Jonathan A. Engelman, Matthew Ludwig, Lindsay A. Noakes, Emily Rusiecki

*Assessment for Learning Project Fellows; Western Michigan University, United States*

**ABSTRACT:** Assessment for Learning (AFL) or formative assessment has the potential for raising standards and student achievement. This article describes research conducted by graduate students in an NSF-funded capacity-building project with goals to increase research in the disciplines of science and mathematics education. As background, a literature review shows parallels between assessment for learning research developed over the past two decades in England with assessment reform efforts in school mathematics and science in the U.S. A series of five research projects, conducted by Western Michigan University (WMU) AFL project scholars in fall 2008, examines (1) the prominence of assessment for learning in university course syllabi at WMU and (2) at public universities across the state of Michigan, (3) its prominence in policy statements in a sample of Michigan high schools, (4) the development of benchmarking practices with preservice teachers, and (5) a comparison of existing AFL observation protocols used in classroom research. These reports represent the range of potentially important areas that may prove fruitful for moving research on assessment in mathematics and science education forward.

This material is based upon work supported by the National Science Foundation under Grant No. 0733590.

**KEYWORDS:** *assessment for learning; student assessment; formative assessment; science and mathematics education*

Since passage of the No Child Left Behind (NCLB) Act (Public Law 107-110) in 2001, all stakeholders within the public education community have been keenly aware of its demands for accountability as a key component of assessment. In fact, for many stakeholders, assessment in the form of annual measures of progress (a key component of NCLB) now dominates what happens in classrooms as teachers map teaching and textbooks to state standards, school administrators look to satisfy annual “gain” requirements and avoid state sanctions, and state educators seek to maintain federal compliance and funding for education programs. For teachers especially, it is becoming increasingly common for school-, district-, or state-level achievement measures and comparisons to be published routinely in local newspapers. For good or bad,<sup>1</sup> NCLB is seen as one of several measures<sup>2</sup> currently used to make comparative judgments regarding the “health” of education in the U.S. (Ferrini-Mundy & Floden, 2007). To many, the mediocre overall performance of U.S. students on national and international tests raises concerns about global competitiveness while linking these woes more broadly to all scientifically oriented disciplines: science, technology, engineering, and mathematics (STEM). The message that dominates many national headlines is that we (i.e., U.S. students and educators) must do better in all STEM disciplines to maintain our global “edge.”<sup>3</sup> Yet, while these high-stakes tests and

international comparisons constitute the public, “summative” face of assessment, within the past two decades, another “formative” face has emerged that is potentially more useful and productive for producing gains in student achievement (see Black, Harrison, Lee, Marshall, & Wiliam, 2003). These two aspects of assessment have emerged in the literature from usage first coined by Scriven in 1967 in terms of distinguishing different roles for evaluation (Wiliam, 2007). For both classroom teachers and educational researchers, they provide new opportunities to improve teaching and learning in K-12 schools.

This article focuses on formative assessment, or assessment for learning (AFL), in the disciplines of mathematics and science by reporting on the initial activities of an NSF-funded project at Western Michigan University (WMU). We begin with a brief, orienting section that highlights some of the important general literature associated with assessment for learning as an important facet of the education and evaluation landscape. The following two sections describe recent developments in mathematics and science education that show how assessment for learning has been an important area of concern within the standards-based reforms that emerged in those disciplines during the early 1990s. In the final major section, we give a brief general description of the WMU Assessment for Learning Project, followed by a series of reports of ongoing research by project associates and fellows. The reports cover a wide range of activities that illustrate the breadth of interesting questions that can (and perhaps need to) be investigated within AFL research. We close with some brief comments on future directions for the project.

---

<sup>1</sup> See the spring 2008 AEA *New Directions in Evaluation Journal* (117), Consequences of No Child Left Behind for Educational Evaluation for various perspectives.

<sup>2</sup> The others are the National Assessment of Educational Progress (NAEP, 2009), given periodically to a random sample of school students in grades 4, 8, and 12; and various international comparisons, such as the most recent Trends in International Mathematics and Science Study (TIMSS, 2007); and a new, but different comparison titled Program for International Study Assessment (PISA, 2007).

<sup>3</sup> For an alternative interpretation of global competitiveness arguments as measured by international comparisons, see the many *Phi Delta Kappan* articles by

---

Bracey (2000, 2002, 2006); and most recently reaction to TIMSS and PISA on *eddra listserv*, 12/09/08.

## Assessment for Learning Literature

In some ways, it is difficult to assign a precise date for the emergence of assessment for learning as a distinct feature in educational literature. Dylan Wiliam (2007), a key early proponent of its importance to classroom teaching and learning, argues that it likely emerged as an extension of the distinction between formative and summative evaluation made by Scriven in 1967 (see the following section). Yet, when one examines some of the key components and strategies associated with current interpretations of the phrase, such as asking good questions and student motivation, these ideas were deemed important in educational theories and classroom practice throughout the twentieth century. Today, a number of Web sites have assessment for learning as their focus, and a growing body of literature has emerged to support its advocacy as good practice for teachers, its potential for improving student achievement and raising standards, and as an important area of education research.

Despite differences in how authors and Web sites lay out their presentation of assessment for learning—whether through definitions, sets of principles, or vignettes and examples from classroom practice—there is general agreement on key ideas. Most begin by equating assessment for learning with formative assessment, giving educators a familiar starting point for contrasting it with the common classroom practices that encompass summative assessments familiar to teachers, such as quizzes, tests, and cumulative final exams. In one of the seminal articles on the subject, Black and Wiliam (1998b) state that “assessment becomes formative assessment when the evidence [student oral or written work in any form] is actually used to adapt the teaching to meet student needs” (p. 140). The Qualifications and Curriculum Authority (n.d.) Web site elaborates by stating that “assessment for learning involves using assessment in the classroom to raise

pupils’ achievement. It is based on the idea that pupils will improve most if they understand the aim of their learning, where they are in relation to this aim, and how they can achieve the aim (or close the gap in their knowledge” (n.p.). Most articulations then list between six and ten principles of assessment for learning that include key ideas, such as clearly expressing the goals of learning, providing means of feedback to students, improving classroom dialogue, helping students become independent learners, and ensuring that *all* students are included in the teaching and learning process. Such notions are consistent in many of the foundation documents noted elsewhere in this article (e.g., Joint Committee on Standards for Educational Evaluation [JCSEE], 2003; National Council of Teachers of Mathematics [NCTM], 1989, 2000; National Research Council, 1996). For many stakeholders, assessment for learning in the form of Black and Wiliam’s (1998a) original notion of looking inside the “black box” strongly suggests the potential for uncovering and understanding the formative classroom strategies that will have the most impact on student learning.

Perhaps most notable in these various definitions of assessment for learning is that both teachers *and* students play key roles in the process. This can be seen in the development of *The Student Evaluation Standards* (JCSEE, 2003), a companion volume to *The Program Evaluation Standards* (JCSEE, 1994) and *The Personnel Evaluation Standards* (JCSEE, 2009b) by the Joint Committee on Standards for Educational Evaluation. Several educators played major roles in developing, applying, and disseminating *The Student Evaluation Standards*. Airasian and Gullickson, formerly the committee’s chair, who collaborated on *The Personnel Evaluation Standards* work, collaborated to draft a book (*Teacher Self-Evaluation Tool Kit*, Airasian & Gullickson, 1995) that focused on reflective teacher<sup>4</sup> evaluation practices. The book drew

---

<sup>4</sup> See, for instance, the early publication, *The Reflective Practitioner*, by D.A. Schon, Basic Books, 1983.

heavily from literature in that area, focusing on teacher data-gathering activities that now are considered assessment for learning practices. The most prominent book in this genre is titled *Classroom Assessment Techniques: A Handbook for College Teachers* (Angelo & Cross, 1993). These materials all have a common theme of getting teachers to gather data that will enable them to better assist student learning in their classrooms. These techniques and issues, addressed under the umbrella for reflective teaching practices, were then brought into the preparation of *The Student Evaluation Standards* (JCSEE, 2003). Additionally, Mark Wilson, a measurement professor from Berkeley, served as the American Educational Research Association's representative to the Joint Committee and was on the National Research Council (NRC) committee to prepare the book, *Knowing What Students Know* (NRC, 2001). That book synthesizes the early work on assessment for learning. Mark Wilson and Bob Wilson from Queen's College in Canada were current in the Black (also a member of the NRC committee) and Wiliam assessment for learning literature and brought those ideas to the Standards. This group of evaluation and measurement specialists served to ensure that many of the assessment for learning ideas were included in *The Student Evaluation Standards*. Like most early ventures, there was not a common language shared across the groups (e.g., assessment for learning was not part of these beginning conversations).

The assessment for learning literature provides encouraging evidence of its importance in improving student achievement. In their review of the literature, Black and Wiliam (1998b) examined more than 250 assessment-related studies and found that of the more than 20 that focused on student learning and achievement, effect sizes of 0.5-1.0 were typical, and gains for lower achieving students were especially noteworthy. Stiggins (2002) noted that such gains would have raised the U.S. ranking on the 1999 TIMSS international comparison to among "the top five" performing countries of

the 42 that participated. The studies noted in the Black and Wiliam (1998b) review include a range of contexts, including the training of 25 Portuguese elementary grades mathematics teachers who implemented student self-assessment techniques in their classrooms that involved more than 200 students (Fontana & Fernandez, 1994; another that involved more than 7,000 students in mastery learning connected to feedback techniques (Whiting, Van Burgh, & Reder, 1995); and a third that was an experimental design conducted with 120 American college mathematics students using various mastery learning techniques as independent variable (Martinez & Martinez, 1992). It is noted that many of these studies are now more than a decade old, and the lack of newer studies—as suggested in Wiliam (2007), especially in U.S. mathematics education research (see below)—may partly reflect the recent preoccupation with high-stakes summative testing and national standards associated with passage of NCLB. Despite this speculation, there is some evidence of continued interest in the promotion of assessment for learning teacher training and research in more recent projects.

The general literature related to assessment for learning identifies a number of important teacher pedagogies that are characteristic of good classroom practice relative to ongoing assessment. Many of these have an associated body of literature (Wiliam, 2007). For example, Black and Wiliam's (1998a, 1998b) original work with English and mathematics teachers in England focused on improving questioning strategies and classroom dialogue, changing the focus of written feedback and grading practices (such as marking of student papers), and establishing norms of peer- and self-assessment among the students. Their continued work with teachers on these assessment for learning techniques were all based on a project begun in 1999 that replicated their earlier results and began to offer practical advice for teachers for how to achieve them (Black & Harrison, 2001;

Black et al., 2003, 2004). A related line of research has attended to improving classroom dialogue in language education (van Lier, 1996) and more broadly to all classrooms, sometimes providing general practical advice (see Dillon, 1994) to teachers and more targeted suggestions for specific subject matter such as mathematics (see Steinbring, Bussi, & Sierpinski, 1998; Chuska, 1995). A survey of research evidence on dialogic teaching across disciplines and countries can be found in Alexander (2008).

A number of projects have investigated other AfL practices. For example, one project focused on helping teachers establish personal learning plans with their students (Bullock & Wikeley, 1999), which addressed a key AfL component of helping students become independent learners and assessors of their own progress. Another project directed by McCallum and Clark (2001) seeks to provide inservice teachers with practical advice for how they can help students use peer- and self-assessment tasks that help them reflect on their own work. Similar efforts that seek to provide ready-made inservice materials to help teachers address aspects of assessment for learning can be seen through downloadable Web site packages, such as *Assessment is for Learning: Self-Assessment Toolkit* (Scottish Qualifications Authority, 2006). These projects represent some of the seminal activities and research, largely conducted in England and Scotland, which brought assessment for learning into the education landscape as an important area linked to potential positive impact on student achievement.

In the following two sections, we examine related efforts in the U.S. that link directly to the STEM (science, technology, engineering, and mathematics) fields of mathematics and science education. We emphasize these two areas because they represent the focus of this particular grant, but that does not imply that the fields of technology and engineering education have no interest in assessment for learning practices as key to twenty-first-century literacy. As evidence, the National Academy of

Engineering, the National Research Council, and the Committee on Assessing Technological Literacy have recently combined efforts to produce a set of assessment standards that will guide future thinking about how we define technology literacy across all levels of education and what it means to assess that literacy. The committee report, *Tech Tally: Approaches to Assessing Technological Literacy* (Gamire & Pearson, 2006) acknowledges that “assessment of technology literacy in the United States is in its infancy” (p. 7) and the report lays out twelve specific recommendations for how we can begin to address this increasingly important part of education that now is infused into nearly every school subject and all aspects of our daily lives. In particular, Recommendation 7 calls for a study of synthesis research in mathematics learning, spatial reasoning, design thinking, and problem solving that can provide guidance on pedagogy, assessment, teacher education, and curricular issues “at all levels” that can inform “assessment in the domain of technology” (p. 11). Other recommendations, especially those directed at K-12 teachers, are more summatively oriented and call for the developments of assessment items across disciplines for students (Recommendation 4), and pre- and in-service teachers (Recommendation 5). Overall, increased attention to technology and assessing technology literacy is echoed in parallel standards documents (see below), most notably the *Principles and Standards for School Mathematics* (NCTM, 2000) where it is seen as one of six key principles of school mathematics learning and teaching for the twenty-first century.

## Assessment Trends in School Mathematics

In the most recent *Handbook of Research on Mathematics Teaching and Learning*, Dylan Wiliam (2007) notes in the opening chapter of the assessment section (Chapter 23) that the previous *Handbook* (Grouws, 1992) made only minor reference to the term *formative* in

connection with assessment and then almost “exclusively in the context of program evaluation” (p. 1057). Such an observation seems an apt starting point for characterizing assessment for learning in the field of mathematics education. William is careful to point out that this does not mean that K-12 mathematics teachers have made no use of acknowledged assessment for learning practices (such as good questioning techniques), but rather that few projects before that point had made it a focal point of good pedagogy (see Lambdin, 1993). Both teaching practices and curriculum materials consistently emphasized the summative aspects, despite a growing advocacy for broadening the notion of assessment to include formative aspects by a number of prominent mathematics educators (see Glaser, 1990; Silver, 1992; Webb, 1994; Romberg, 1995).

Given that as background, perhaps two major trends best characterize changes and developments in school mathematics assessment over the last two decades: a shift to multiple forms of assessment to better understand students’ mathematical learning (of which assessment for learning is clearly included) and using assessment for a variety of accountability purposes (NCTM, 2000). As noted in our opening paragraph, assessment for accountability receives much current attention, but changes in teaching practices that include assessment for learning have at least been paid considerable “lip service” as an important form or type of assessment in mathematics education reforms that began with the *Curriculum and Evaluation Standards for School Mathematics* (NCTM, 1989). In that document, most notably Standards 2 and 3 of the evaluation section, K-12 mathematics teachers were encouraged to understand the various “purposes of assessment,” use the appropriate assessments for “instructional feedback,” and recognize that “assessment must be more than testing; it must be continuous” (p. 203). NCTM soon followed with publication of a companion document, the

*Assessment Standards for School Mathematics* (1995), that more fully articulated the breadth and importance of assessment as a critical component of effective mathematics teaching and its basis in research related to the many proposed changes in pedagogical practices (see Webb, 1994; Wilson & Kenney, 2003). Paralleling much of what is advocated in the assessment for learning literature outlined in the previous section and *The Student Evaluation Standards* (JCSEE, 2003), NCTM’s position is best summarized in the Learning Standard, where it cautions teachers that “although assessment is done for a variety of reasons, its main goal is to advance students’ learning and inform teachers as they make instructional decisions” (p. 13). Its importance in mathematics education is also noted in the NCTM devoting its *1993 Yearbook, Assessment in the Mathematics Classroom* (Webb & Coxford, 1993) to the topic. In that volume, a number of assessment for learning ideas are reported by K-12 teachers<sup>5</sup> and a key theme is that many of them have recycled in the literature “for decades” (Lambdin, 1993). These assessment publications and position statements by the largest body of mathematics teacher educators in the U.S. has continued in the most recent reworking of these documents, the *Principles and Standards for School Mathematics* (NCTM, 2000). Consistent across all NCTM documents has been the notion of helping students achieve mathematical “power” (Standard 4, p. 205, NCTM, 1989) that represents a major goal of recent efforts to improve mathematics teaching and learning. The *Standards 2000* document argues that “students learn more and learn better when they can take control of their learning by defining their goals and monitoring their progress” (NCTM, 2000, p. 21)—clearly a key component of assessment for learning goals and principles.

---

<sup>5</sup> Themes such as using interviews, learning logs, empowering student learning, and using observations and reflections are explored in separate chapters (9-13) within the yearbook.

To help K-12 mathematics teachers achieve these assessment goals, over the past two decades NCTM has been active in publishing resources that provide practical teaching tips and activities for use in their classrooms. A good illustration is the current NCTM catalogue of resources that includes thirteen documents with *assessment* explicitly referenced in the title, including a six-volume set of assessment handbooks and discussion tips and a four-volume assessment sampler with assessment items and teaching tips that span K-12 and all content strands (e.g., algebra, number, probability, statistics, etc.). Other publishers of mathematics education materials have kept pace with NCTM by producing dozens of their own helpful resources and “how-to” manuals ranging across a variety of assessment-related topics and introducing a whole new vocabulary into the field. Such titles as *Reform in School Mathematics and Authentic Assessment*, *A Collection of Assessment Tasks and Rubrics*, *Portfolio Assessment*, *Performance Assessment and Standards-Based Curricula*, *Balanced Assessment*, and *Improving Classroom Questions* represent a small sample of what is available to the teacher consumer. And, such words as *performance*, *authentic*, *alternative*, *balanced*, and *portfolio assessments* reference a much broader picture of assessing student understanding of mathematics and are generally linked to specific classroom vocabulary that involves tasks, rubrics, questioning, discourse, open-ended problems, problem solving, and self-assessment. This vocabulary is now common within the mathematics teaching community, and books such as those noted above are replete with teaching ideas that can help move teachers from simple diagnostic and summative assessments to a more robust package of assessment for learning techniques and practices. Kulm (1994) suggested early on that a reason “for changing the direction of assessment [in K-12 mathematics] has been to focus on problem solving as a key part of the mathematics curriculum. Continued efforts are aimed at moving beyond simple word problems

toward evaluating the processes that students should use in a variety of situations” (p. 25). Yet, despite both the rationale for changing classroom assessment practices in mathematics education *and* the mounting available number of resources to help teachers do so, we have little evidence of the extent of use of such practices within K-12 classrooms or how, if at all, such resources are helpful to mathematics teachers.

A second area where mathematics education has made headway in changing teachers’ assessment practices is in the development of standards-based K-12 curricula that reflect the NCTM *Standards* documents. Largely funded through NSF grants, these curricula, now widely available and used in schools, engage students in mathematics activities that are designed to foster the NCTM process standards of reasoning, problem solving, communication, making connections, and understanding different mathematical representations across content areas at all grade levels. In many of these curricula, assessment is *embedded* into the investigations that students engage in in the classroom (see Fey, Fitzgerald, Friel, Lappan, & Philips, 1996; Hirsch, Coxford, Fey, & Schoen, 1994, 1995; Schoen, Bean, & Ziebarth, 1996). Previous mathematics textbooks and curriculum materials focused almost exclusively on providing summative kinds of assessments and practice or rote kinds of activities. The embedded assessment features were designed to help teachers develop and use formative assessment practices as they use the materials in the hope that such skills would become a natural part of their teaching practices. In addition, the materials tend to emphasize the important role that students play in their own learning processes. Despite the availability of this new (many are now in second editions) avenue of assessment resources for mathematics teachers, and despite extensive research on teachers’ use of these materials (see Remillard, Herbel-Eisenmann, & Lloyd, 2009), there is virtually none that focuses on how they support teachers’ assessment for learning practices. This remains a

potentially fruitful area of future research in mathematics education.

Perhaps the struggle for making assessment for learning an integral part of K-12 mathematics teaching is best seen by what little evidence there is from classroom-based research devoted solely to that as a focus. For example, in the 1990s, even after almost a decade of curriculum and assessment reforms, it was still the case that students' "grades" in mathematics courses were determined by a very narrow range of assessment techniques, largely through quizzes and tests. The Senk, Beckmann, and Thompson (1997) study of assessment practices in five high schools in three states illustrated this problem where these two techniques alone accounted for 77 percent of students' grades, with only small contributions coming from alternative types of assessment (when they were used at all). In addition, "test items generally were low-level, were stated without reference to a realistic context, involved little reasoning, and were almost never open-ended" (p. 187). However, in about two-thirds of the classes, small shifts toward multiple and more complex forms of assessment were found, and the critical factors associated with these changes were teachers' knowledge and beliefs and the instructional materials available to the teachers (p. 210). The authors and many others (see Bright & Joyner, 1998; Bryant & Driscoll, 1998) have suggested that increased professional development focus on assessment and other forms of exposure (e.g., through curriculum) so examples of broader assessment measures and techniques could begin to have some impact on changing teachers' practices and perceptions at the classroom level. A key issue seems to be that many teachers see little distinction between the terms *assessment*, *evaluation*, and *grading* since, for most of them, nearly everything important to their students is in some way tied to grades. Senk et al. (1997) concluded with the following observation:

Teachers who are in the process of making the types of shifts in assessment practices advocated

by the profession need guidance on how to aggregate results of students' performance on these new tasks and on how to report summary comments to students, parents, employers, and colleges. Because teachers need guidance as soon as they begin to try new forms of assessment, research on issues of aggregation and reporting of high school students' performance on complex mathematical tasks is critical and of immediate need. In the meantime, grading should come out of the shadows and onto center stage of the discussion on how to improve the quality of assessment in schools. (p. 213)

Of additional importance in reporting this study by Senk et al. is its acknowledgment by Wiliam (2007) as the most recent of research efforts related to assessment practices in mathematics classrooms, further emphasizing the potential research that could be moved in this direction. Simply put, assessment for learning as an area of research in mathematics education is rife with interesting questions relating to teachers and student learning. Wiliam (2007) cites research in a number of subareas—such as questioning, feedback, active engagement, motivation, and cooperative learning involving peer and self-assessment—that are important for our understanding of the whole field of assessment for learning, but most are approaching at least a decade old. We should be encouraging new research in each of these areas while also studying the impact of curriculum materials, teacher preparation materials, and inservice aspects that can make a difference in mathematics classrooms. As we do this, it will be important to develop research tools that focus specifically on assessment for learning practices that can be used to collect data for research into what works for teachers in classrooms. In the final section of this paper, we highlight some initial efforts at research by our project fellows with interest in mathematics education.



## Assessment Trends in School Science

There are reasons to be optimistic about the adoption of assessment for learning practices in the science classroom because a significant portion of the empirical research on AfL has taken place within the context of science courses. In Black and Wiliam's (1998a) review paper, thirty articles were from journals that are exclusively devoted to research in science teaching. In the more recent Black et al. (2003) book, twelve of the thirty-six teachers involved were science teachers. Capacity-building efforts still are limited in that they have focused on AfL in K-12 classrooms with much less attention given to higher education. We believe that the injection of AfL into the science courses taken by future teachers during their tertiary training is needed to orient and sensitize them to AfL principles and practices. This addition to the existing capacity-building efforts will provide fertile soil for still more pervasive K-12 AfL practice.

Assessment for learning as a teaching strategy is quite visible nationally for primary and secondary science teachers. Support for AfL practices has been explicit in the National Science Education Standards (NSES) since 1996 (NRC, 1996). For example, Teaching Standard C includes specific AfL principles:

Teachers of science engage in ongoing assessment of their teaching and of student learning. In doing this, teachers

- Use multiple methods and systematically gather data about student understanding and ability.
- Analyze assessment data to guide teaching.
- Guide students in self-assessment. (pp. 37-38)

The science Standards elaborate on how assessments must go beyond summative: "The word 'assessment' is commonly equated with testing, grading and providing feedback to

students and parents. However, these are only some of the uses of assessment" (p. 38). The Standards value the role of assessment in an informal and formative sense during the daily interactions between students and teachers. "During the ordinary operation of a class, information about students' understanding of science is needed almost continuously" (p. 38). The link between classroom assessment and the NSES document is further explicated in *Classroom Assessment and the National Science Education Standards* (Atkin, Black, & Coffey, 2001), a joint effort of the Center for Education, the National Research Council, and a committee convened to create the publication.

Consistent with the NCTM focus on assessment described above, the National Science Teachers' Association has supported the mandate of the *National Science Education Standards* through the publication of dozens of books, chapters, and resources for K-12 teachers. These range from research-oriented pieces such as *On the Role and Impact of Formative Assessment on Science Inquiry Teaching and Learning* (Shavelson et al., 2008) to classroom-ready activities and strategies such as *Science Formative Assessments: 75 Practical Strategies for Linking Assessment, Instruction, and Learning* (Keeley, 2008). Many of these resources are available inexpensively or without cost to NSTA members. Attending an NSTA meeting makes it seem as though all K-12 teachers know about and are hungry for using AfL principles in the classroom. We know, however, that it is only a minority of science teachers nationwide who attend such meetings and perhaps use such practices. So, there remains a need to reach more teachers through their tertiary training.

Unlike the K-12 community, tertiary education often seems unaware of assessment for learning as a paradigm. Here the word *assessment* generally has two meanings. The first is summative assessments, such as quizzes and exams. The second is program assessment, such as what a department might do at the urging of the administration to improve overall programs.

At the most recent Biennial Conference on Chemical Education, of the nearly 200 total sessions, posters, and papers, approximately 75 advocated for teaching that aligns with AfL principles. Of these, few framed these methods within an overarching paradigm of formative assessment. In particular, there were two sessions with the title “Assessing to Generate Learning Centered Undergraduate Chemistry.” But examination of the presentations made at these sessions reveals that the focus of nearly every talk was either summative or program assessment.

There is reason, however, to be optimistic about the adoption of assessment for learning principles into tertiary science education in the future. Although science educators might not know the AfL label, many who conduct research within science education advocate for specific methods consistent with AfL principles. These active areas of research promote many forms of instruction that increase the amount of feedback given to students during the learning process, either inside the classroom or while students are doing work outside of class. Inside the classroom there long has been a push for the use of guided inquiry and group problem-solving activities. More recently, the use of classroom response systems—“clickers”—has been added in many universities nationwide. Research on the effectiveness of this technology is in its infancy. Outside the classroom, students increasingly are using computer-assisted learning such as online homework delivery, simulations, and tutorials. While such delivery systems are becoming more pervasive, only a few research studies have addressed their utility and efficacy (see Fynnewever, 2008). All of these techniques are consistent with AfL principles in that they increase the amount of formative feedback given to students from their teachers, peers, and/or a computer.

## The WMU Assessment for Learning Project

In September 2007, Western Michigan University was awarded an NSF DRK-12 Capacity Building Grant (DRL-0733590) titled *Assessment for Learning (AfL) Education Research Scholars: Capacity Building in Mathematics & Science Education*. As a collaborating effort of the WMU Evaluation Center, the WMU Center for the Study of Mathematics Curriculum and the graduate program in mathematics education, and the Mallinson Institute for Science Education, the project’s main goals are to (1) bring faculty from all three programs together in a learning community to foster the development of research and leadership in evaluation, specifically AfL, (2) recruit a cadre of STEM education research scholars to be trained in and conduct research on measurement and evaluation, and (3) expose preservice teachers to assessment models in their STEM coursework, providing them from the start with a natural appreciation and understanding of assessment theory and practice. The recruited graduate students are seen as playing a key role in developing these AfL models as part of their overall graduate preparation.

At present, the WMU Assessment for Learning grant has seven active fellows who began their graduate studies by spring 2009—four in science education, two in the interdisciplinary program in evaluation with concentration in mathematics education, and one in the doctoral program in mathematics education. A guiding principle of the AfL grant is to immerse these students, from the beginning, to play an active role in carrying out research in these areas as a part of their programs of study. Through the literature reviews described above, we have shown the wide range of potential areas of research that may provide avenues for small-scale studies that are within the scope of what graduate students are normally be expected to undertake within the timeframes of their doctoral programs. Below, we provide brief summaries of the current studies under way by WMU AfL fellows and collaborators.

# 1. A Study of Syllabi: Exploring Evaluation Practices via Course Syllabi (Amy Bentz, Herb Fynewever, and Matthew Ludwig)

## *Research Question and Context*

The following questions guided this study: What evaluation and/or assessment practices are currently conveyed via syllabi in foundational-level and nonmajor science courses at Western Michigan University? What overlap exists between widely accepted principles of assessment for learning and the Joint Committee's (2003) *Student Evaluation Standard*? What should be communicated regarding evaluation in a course syllabus?

According to the Qualifications and Curriculum Authority (n.d.), assessment for learning involves using formative evaluation and/or assessment practices in the classroom to raise pupils' achievement. Teachers use assessment feedback to determine the magnitude of student learning. Using this feedback, teachers can modify classroom activities to suit the needs of their students (Black et al., 2003). One of the fundamental tenets of assessment for learning is to begin with clearly defined and measurable learning objectives that need to be communicated prior to instruction. The course syllabi could be utilized more effectively to communicate course learning objectives and the related evaluation plan. Clarification of student and instructor responsibilities offers the student direction on how to achieve course objectives and provides students with the capability of assessing their own work during the course (O'Brien, Millis, & Cohen, 2008). Communicating the specific learning objectives provides students with the framework for frequent self assessment throughout the course.

## *Methods and Results*

We collected downloadable syllabi from entry-level science courses taught at WMU in the fall semester of 2008. Two disciplines, biological sciences (BIOS) and geological sciences (GEOS) were the focus of our study. Syllabi were located using the university's internal Web search and by visiting departmental and faculty home pages. We collected a total of eight syllabi, four in BIOS, three from GEOS lecture classes, and one from a required GEOS discussion section. Our analysis began by reviewing each syllabus for evidence of student evaluation with the ultimate intent to characterize the resulting data into emergent themes. Lack of data in the syllabi prevented the development of emergent themes. Our preliminary conclusion was a general lack of specifics regarding student evaluation.

In summary, we found that two evaluation items were included in all of the syllabi. The first is a quantitative grading scale to synthesize the semester evaluation results into a single grade. Even more interesting is that none of the grading scales were identical. Specifically, the dividing line between the highest measure of achievement (A) and the next highest (BA) was found to range anywhere from 89 percent to 93 percent. The second commonality is a boilerplate plagiarism/academic honesty statement provided by the University. This statement described the University policy as the final word in any cases involving plagiarism or cheating and places sole responsibility for knowing said policy and abiding by it on the student.

A wide variety of items were found in the documents we reviewed. Among these were, in order of prevalence, including noncontent criteria (e.g., attendance, late penalties, effort, staple position, presence of dog-ears), bonus points, extra credit, and dropping the lowest test. The following items were noted on only one example we inspected: course goals, learning objectives, the "Loser's Scheme," time line for

evaluation results, how to get access to course/test grades, and mention of peer assessment. The “Loser’s Scheme” described a system where all students begin the semester with an account of several million points. The methods for students to lose points are described in detail.

### Conclusions

We contend that providing a clearer focus by concentrating exclusively on areas of overlap between *The Student Evaluation Standards* and the *AJL* principles will present educators with a more efficient set of guidelines for modifying their practice to maximize student learning. When constructed properly, a syllabus can help to obtain this goal. Specifically, a syllabus can:

- Detail for the student what the course objectives are. This list should not be merely an inventory of what topics will be covered in the course. Rather, it should itemize for the student what they can expect to learn through this course and can, therefore, be revisited by the student to assess his/her progress.
- Describe for the students how the teacher will evaluate their achievement. Most students equate what teachers evaluate with what teachers value (McKeachie, 1986). Whether this belief is warranted or not, teachers need to be aware that many students take the issue of evaluation very seriously; uncertainty regarding how they will be evaluated is sure to cause anxiety.
- Clarify instructor and student responsibilities. The syllabus is often considered to be an informal written contract for the course (O’Brien et al., 2008). The student should know what to expect in the way of help and resources from the instructor. Likewise, the instructor should let the students know how to go about meeting the course

objectives. Ideally, this should include specific expectations with due dates. The expectations should be detailed enough that the student can refer back to them for the purpose of self-assessment. O’Brien et al. suggested that detailed expectations will increase the likelihood that the work will be done properly and on time.

In this exploratory study, *none* of the syllabi examined fit the guidelines noted above. We see this as an opportunity missed. The syllabus can be the opening communication from teacher to student that sets the tone for a course rich in the scaffolding of assessment for learning.

## 2. Michigan High School Assessment Policies: A Descriptive Study (Lindsay Noakes)

### *Research Question and Rationale*

As noted in our opening section, critical analyses and impact of the 2001 No Child Left Behind legislation and high-stakes testing systems dominate the literature on assessment policies. Very little literature can be found on school assessment policies outside of large-scale testing and accountability, and virtually no literature can be found linking school assessment policies to student learning. As a result, the scope and quality of school assessment policies came into question. This study attempted to determine the extent to which Michigan high school assessment policies follow *The Student Evaluation Standards* (JCSEE, 2003) and assessment for learning themes.

### *Methods and Results*

Using Standard and Poor’s (2009) *SchoolMatters* data of the top 100 performing high schools in Michigan (based on percent proficiency on the

2006 Michigan Merit Mathematics Exam), a simple random sample of twelve schools was selected for study. Demographic information and available electronic documents such as student handbooks, course syllabi, and school policies were collected from the selected schools for which electronic information was available online (ten of the twelve). Additional in-depth interviews were conducted with personnel from three of the ten schools. All data from documents and interviews were compared with *The Student Evaluation Standards* (JCSEE, 2003), and judgments were made regarding their alignment with the three assessment for learning themes of the Joint Committee Benchmarking Project (JCSEE, 2009a).

1. Evaluation and assessment should consistently be designed to improve learning.
2. Intended outcomes of learning (and of evaluation) should be clearly stated to students and shared with other stakeholders.
3. Students should be engaged in the evaluative process and to the extent possible in planning their own next steps for learning.

The standards that were aligned to each of these themes were modified slightly for the purpose of this study. A description of each standard is presented in Table 1. Each of the ten

schools was rated as having high, medium, or low evidence of established policies for each of the aligned standards. A rating of high (green) indicates that the school had clear evidence of an established policy in that area, a rating of medium (yellow) indicates that the school had some evidence of a policy, and a rating of low (red) indicates that the school had little to no evidence of a policy.

Results of this study indicate a wide range in levels of established assessment policies among the ten schools (see Table 2). It was generally easier to find evidence of explicit policies in place when policy documents, in combination with interview data, were available for examination. In addition to the alignment with *The Student Evaluation Standards* findings, data revealed that most student handbooks and school policy documents discussed student assessment only in terms of the MEAP, MME, ACT, or other standardized tests. There was typically little, if any, mention of specific classroom assessment policies; however, those that were in place tended to focus on summative rather than formative assessments. A follow-up study is planned to continue the in-depth interviews and examine a larger sample of Michigan schools to provide a more complete picture of assessment for learning policies. Additionally, fidelity of policy implementation and impact on student learning will be included in the data collection.

Table 1  
Standards Used to Evaluate School Assessment Policies (adapted from the Joint Committee Benchmarking Project, JCSEE, 2009a)

Aligned Standard	Description of Standard in Context of Assessment <i>for Learning</i>
P1: Service to Students	Assessments are designed to provide students with useful information about their learning.
P2: Appropriate Policies and Procedures	Schoolwide and/or department guidelines on assessment are in place and are directly linked to student learning.
P3: Access to Evaluative Information	Students are able to check grades, formative comments, and progress on a regular basis. Grades, comments, and feedback are confidential.

P6/U1: Constructive Orientation and Balanced Evaluation	When providing feedback on assessments, strengths are identified along with areas for improvement. Students are given the opportunity to revise and edit their work based on formative evaluations.
U3: Information Scope	Learning objectives and rubrics for evaluation are “mapped” onto each other to demonstrate their relationship. Performance indicators within the rubric include all aspects of the learning objective.
U5: Explicit Values	Student evaluations (including self-assessments) are incorporated throughout the course/unit of study and aligned with learning outcomes.
U6: Effective Reporting	Useful feedback is provided to students in a timely manner. Students can use feedback to assess their progress toward achieving learning outcomes prior to summative assessment/evaluations.
U7: Follow-Up	Teachers and students discuss evaluative information and make a plan for remedial and/or enrichment activities. Students are held accountable for the execution of their learning plans.
A2: Defined Expectations	Explicit requirements and outcomes are shared with students and made accessible to parents.
A3: Context Analysis	Students are provided with the opportunity to demonstrate learning by using a variety of assessment strategies that allow for student choice and are individualized to meet the needs of students.
F2: Political Viability	Classroom policies are provided to parents at the beginning of the school year. Written policies are available online or in the school office for public review.

Table 2  
School Ratings on Assessment Policies for Each Standard

School	A	B	C	D	E	F	G	H	I	J
P1	High	Med	Med	Low	Low	Med	Low	Low	Low	Low
P2	High	High	High	High	High	High	High	High	High	High
P3	Med	High	Med	Med	Med	Med	Med	Med	Med	Low
P6/U1	High	Med	Med	Low	Low	Low	Low	Low	Low	Low
U3	Med	Low	Low	Low	Low	Low	Low	Low	Low	Low
U5	Med	Med	Med	Low	Low	Low	Low	Low	Low	Low
U6	Med	High	Med	Med	Med	Med	Low	Med	Low	Low
U7	Med	Med	Med	Low	Low	Low	Low	Low	Med	Low
A2	High	High	Med	Low	Low	Low	Med	Low	Low	Low
A3	High	Med	Low	Low	Low	Low	Low	Low	Low	Low
F2	High	High	Med	Med	Med	Med	Med	Med	Med	Low

### 3. Student Assessment in Michigan Teacher Preparation Programs (Emily Rusiecki)

#### *Research Question and Rationale*

This study reviewed publicly available course syllabi at Michigan public universities to determine if, and to what extent, teacher preparation programs include coursework in assessing student learning. Despite the possibility of teachers spending twenty to thirty percent of their time in assessment-related activities, most teacher preparation programs across the country do not require and/or offer a course in educational measurement. Knowledge of student assessment strategies comes primarily from colleagues and teachers' own assessment experiences as students (Stiggins, 1988). Stiggins (1999) examined individual state teacher licensing and certification requirements in the United States and found that only half require some knowledge of student assessment practices (i.e., coursework or demonstrated competence). In that study, Michigan had no licensing or certification requirements focusing on student assessment and still has none to date (Michigan Department of Education, 2008). Since Michigan does not require coursework or demonstrated competency in assessment to gain licensure, this study examined the teacher preparation programs specifically for instruction in formative assessment.

#### *Methods and Results*

All fifteen public universities in Michigan were contacted via electronic mail requesting information for this study. Eighty electronic mail requests were sent (including multiple inquiries to nonrespondents). Information was obtained from thirteen of them. The

information requested for this study (course syllabi and/or description of assessment topics covered in course) was reviewed to determine if stand-alone courses in student assessment were offered. If no stand-alone course was offered, secondary mathematics methods courses were examined to determine if, and to what extent, student assessment was a component of the course.

After reviewing course syllabi and/or electronic mail course descriptions, only two of the thirteen responding schools were found to offer a stand-alone course in student assessment. These courses were offered through the college or department of education. Several methods courses incorporated student assessment topics as part of other instruction units or as part of an individual or group lesson project. Many syllabi included student assessment as one of the learning objectives or topics covered. Some syllabi included as many as thirteen learning objectives for a one-semester course, thus limiting the amount of time dedicated to covering student assessment topics. Except in one instance, most syllabi and/or electronic mail descriptions did not provide a week-by-week breakdown of instruction topics; however, for those that did provide this information, assessment topics were addressed for only one to three weeks during the semester. Notably, neither formative assessment nor AfL techniques were referenced in the majority of course syllabi and electronic mail correspondence. Of the three universities that specified instruction in formative assessment techniques, there is no mention of AfL principles, only reference to instruction in formative and summative assessment.

#### *Conclusions and Limitations*

While all responding schools offer a stand-alone course in assessment and/or include student assessment as a component of their secondary

mathematics methods course, the extent to which student assessment is covered in the *methods course* varies from school to school. It is important to note the focus of this study was on undergraduate teacher preparation programs, not graduate programs in education. Additional stand-alone courses or certification programs in assessment are offered at the graduate level at many schools. Despite current research suggesting the importance of using student assessment as a tool to increase student learning, little emphasis is placed on student assessment in teacher preparation programs in Michigan. If student assessment is not an integral part of teacher preparation programs, teachers may continue to view assessments as merely a means of assigning grades and ranking students.

#### 4. Benchmarking the Student Evaluation Standards: A Preservice Look at Improving Classroom Assessment (Katharine Cummings, Lindsay Noakes, and Emily Rusiecki)

##### *Research Question and Context*

Many K-12 educators think of benchmarking as a process of standard setting. The focus of the process is on the result—a set of benchmarks or objectives developed by the state (or other higher authority) that will be used to evaluate student learning. A very different view of benchmarking is common for those in business and industry. In this arena, benchmarking is the process by which a company improves. According to Dodd & Turner (2000), the American Productivity and Quality Center's definition of benchmarking is widely accepted: "Benchmarking is the process of improving performance by continuously identifying, understanding, and adapting outstanding

practices and processes found inside and outside the organization" (p. 36).

The benchmarking process can be adapted easily into the K-12 environment, yet the use and understanding of benchmarking in education is still fairly limited. Brennan (1995) claimed that he could find "no mention of benchmarking in the field of school or college management, although it was a phrase used widely in initial discussions and proposals for pupil testing" (p. 36). Additionally, he claims that schools that begin to use benchmarking tend to do so in a very narrow, product-specific scope—simply comparing test scores with that of neighboring districts. This limited view has prompted research on the impact of benchmarking in an education setting.

In an effort to promote and implement the process more broadly in school environments, the National Science Foundation, as part of the Discovery Research K-12 Program, has funded a National Conference on Benchmarking Student Evaluation Practices. The conference (February 2009) and follow-up activities will engage a broad array of education organizations around the issue of improving student achievement, particularly in K-12 mathematics and science, through better student evaluation practices. Conference participants will use *The Student Evaluation Standards* (JCSEE, 2003) in concert with a benchmarking process to examine and improve current student evaluation practices.

As an ancillary project, we have begun piloting the benchmarking materials with preservice teachers. The purpose of this study is to understand the benchmarking process as experienced by preservice teachers as they begin interacting with and teaching students as part of their preinternship and internship semesters. Specific focus will be given to how preservice teachers develop and evolve their processes of best-practice benchmarking with respect to student evaluation processes and practices.



### *Methods and Results*

Undergraduate students currently enrolled in a secondary mathematics teaching methods course, an education course involving classroom placement, and student teaching courses have been invited to participate in the study. Thirty-five students completed a presurvey and have begun the benchmarking process. Over the course of the 2009 spring semester, these students will be interviewed and postsurveys and reflections on the benchmarking process will be collected. Interview, survey, and reflection questions focus on the perceptions of the benchmarking process, the underlying assumptions of student assessment and evaluation, and specific student evaluation practices.

Since the data are qualitative in nature, the method of data analysis will involve four different forms of analysis and interpretation. These include categorical aggregation (where multiple instances are used to find emergent themes), direct interpretation (where a single instance is used to extract meaning), establishing patterns, and making naturalistic generalizations. Preliminary data show that preservice teachers have very little knowledge of the benchmarking process prior to participation in the study; most believed it was related to the more familiar process of standard setting rather than a tool for personal reflection and improvement. Some have indicated that the process of interviewing exemplary teachers is more difficult and time-consuming than what they initially thought. Additionally, those who used e-mail as a method of interviewing found it was difficult to communicate the meaning of the questions asked accurately.

### *Conclusions*

Benchmarking as a process for professional development may provide preservice teachers with the structure needed to elicit practical knowledge from exemplary teachers. The

candidates, however, appear to require greater experience with assessment practices, especially practices related to assessment for learning, before they can effectively conduct interviews about those practices. Moreover, preservice teachers have little experience on which to base the self-evaluation aspects of benchmarking. The pilot project suggests that both the benchmarking materials and the process for implementing benchmarking will need to be adapted to address the knowledge and experience deficits of preservice teachers.

Most importantly, the benchmarking process is a way for teachers to use assessment for learning strategies for their own improvement. It provides a systematic method for teachers to analyze their own teaching practices, learn from themselves and others, implement new practices, and reflect on their own learning. Having teachers model assessment for learning by using the benchmarking process emphasizes the importance of *A/L* in our educational system.

## 5. Student Assessment in Classroom Observation Protocols (Jonathan Engelman)

### *Research Question and Rationale*

A key to understanding assessment for learning in classrooms will be gathering data specifically focused on those *formative* pedagogical assessment practices. Classroom observations have been an important educational research tool for many decades, and a variety of protocols have been developed to study a range of classroom practices and behaviors. Motivated by the research question, “What does assessment for learning look like in the classroom?” this study researched existing classroom observation protocols to document the extent to which assessment for learning ideas were included as a part of or, more importantly, a focus of the observations. This

study is intended to be a first step in developing a valid and reliable observation protocol that can be used to collect research data specifically focused on assessment for learning classroom phenomena.

### Methods and Results

An initial literature review of current classroom observation protocols using a Google scholar search was completed using the following words and phrases: *classroom observation protocol*, *mathematics*, and *assessment*. This search yielded a handful of unique protocols that had assessment in them, such as the much-used and modified Local Systemic Change Classroom Observation Protocol developed by Horizon Research, Inc. (1999). Additionally, the search yielded a number of protocols that used observation items similar to or derivative of other protocols. These partially redundant protocols were not analyzed further.

This literature review found a lack of quality tools dedicated to observing formative assessments in the classroom. Some protocols asked the observer to *provide a rating* based on

the observation. Other protocols require the observer to *describe* the assessments found in the observation. Table 3 reports a comparison of AfL-related items to total number of items for five different protocols examined. The table shows that many items used in a protocol are not assessment for learning items. Rather, most protocols took a holistic approach and focused on entire lessons, with no specific AfL concentration. Two of the protocols (D, E) are composed of more than one-third, and one (B) has more than one-quarter of AfL observation items. However, their treatment is not comprehensive. All protocols could search for broader range and more in-depth AfL-related observations. Some of the protocols have formats that are unique, and one protocol that included few AfL items allows for a good assessment observation. In this case, a “check-off” grid (incremented into five-minute segments) is provided to record when certain types of observations, such as assessment, take place in class.

Table 3  
Observation Protocol AfL Question Comparison

Protocol	Total Items	AfL Items	Percentage of AfL Items
A	29	3	10.3%
B	30	8	26.7%
C	15	2	13.3%
D	21	9	42.9%
E	16	7	43.8%

### Conclusions and Future Research

Table 3 represents a small sample of the classroom observations investigated thus far, and the search for more classroom observation protocols is ongoing. Further investigation will include those intended for use in science classes for comparative purposes. When this phase of the literature review is complete, a new classroom observation protocol, one that

focuses exclusively on assessment for learning, will be developed and validated.

### Closing Remarks

The WMU Assessment for Learning project is in its beginning stages. The examples reported above of work accomplished by our research fellows within their first semester of graduate studies represent a range of potentially fruitful

avenues that will be useful for moving the field forward. Despite advocacy of the importance of assessment for learning from the prominent groups in the STEM fields of mathematics and science, it is not at all clear how we will mesh this meaningfully into an education environment currently dominated by high-stakes testing and results. Our initial work is already revealing on a number of fronts. Do we focus attention on preservice or inservice teachers? The two reports by Bentz, Fynewever, and Ludwig and Rusieki show the paucity of attention to assessment for learning within college course syllabi across our own campus and then again across the state, suggesting gaps at the preservice levels. The report by Noakes suggests it may not be on the radar screen within high school policy statements, either. In all three cases, the hunt was for evidence of assessment for learning embedded within existing syllabi or policy statements, yet the yield was largely about summative aspects; perhaps confirming much of the tenor of this article. One wonders where to start if such results are indicative of the broader landscape. Is K-12 school policy and awareness the place to put our efforts? The report by Cummings, Noakes, and Rusiecki shows that awareness is at least on the NSF radar screen with respect to the notion of AfL in conjunction with the benchmarking process, and research on that aspect with preservice teachers may be productive as they enter into the teaching profession. Addressing a quite different research need, the report by Engelman indicates that existing protocols for collecting assessment for learning classroom data have not had assessment as a focal point. Perhaps developing new instruments may be a place worth expending energy, since more data are fundamental to making the case for its inclusion in any teacher education program, preservice or in-service. In moving the WMU AfL project research agenda forward, we are mindful of the initial work that many in the STEM education and evaluation fields have contributed to making assessment for learning a

visible part of the education landscape. We hope that our initial and future research contributions will be useful to the continuing work of helping teachers become better formative assessors and students to become better independent learners.

## References

- Airasian, P., & Gullickson, A. R. (1995). *Teacher self-evaluation tool kit*. Kalamazoo: The Evaluation Center, Western Michigan University.
- Alexander, R.J. (2008). *Towards dialogic teaching: Rethinking classroom talk*. York, England: Dialogos.
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass.
- Atkin, J. M., Black, P., & Coffey, J. (Eds.) (2001). *Classroom assessment and the national science education standards*. Washington, DC: National Academies Press.
- Black, P., & Harrison, C. (2001). Feedback in questioning and marking: The science teachers' role in formative assessment. *School Science Review*, 82(301), 55-61.
- 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., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box. *Phi Delta Kappan*, 86(1), 8-21.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-71.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Bracey, G. E. (2000). The tenth Bracey report on the condition of public education. *Phi Delta Kappan*, 82(2), 133-144.
- Bracey, G. E. (2002). The twelfth Bracey report on the condition of public education.

- condition of public education. *Phi Delta Kappan*, 84(2), 135-150.
- Bracey, G. E. (2006). The condition of public education. *Phi Delta Kappan*, 88(2), 151-166.
- Brennan, J. (1995). Benchmarking: An alternative to OFSTED: Lessons for higher education. *Quality Assurance in Education*, 3(1), 36.
- Bright, G. W., & Joyner, J. M. (Eds.). (1998). *Classroom assessment in mathematics: Views from a national science foundation working conference*. New York: University Press of America.
- Bryant, D., & Driscoll, M. (1998). *Exploring classroom assessment in mathematics: A guide for professional development*. Reston, VA: National Council of Teachers of Mathematics.
- Bullock, K., & Wikeley, F. (1999). Improving learning in year 9: Making use of personal learning plans. *Educational Studies*, 25(1), 19-33.
- Chuska, K.R. (1995). *Improving classroom questions: A teacher's guide to increasing student motivation, participation, and higher-level thinking*. Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Dillon, J. T. (1994). *Using discussion in classrooms*. Buckingham, England: Open University Press.
- Dodd, J. L., & Turner, M. A. (2000). Is benchmarking appropriate for small businesses? *National Public Accountant*, 45(6), 36-38, 44.
- Ferrini-Mundy, J., & Floden, R. (2007). Educational policy research and mathematics education. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1247-1280). Reston, VA: National Council of Teachers of Mathematics.
- Fey, J. T., Fitzgerald, W. M., Friel, S. N., Lappan, G., & Phillips, E. D. (1996). *Getting to know CMP: An introduction to the Connected Mathematics Project*. Arlington, VA: National Science Foundation.
- Fontana, B., & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, 64.
- Fynewever, H. (2008). A comparison of the effectiveness of web-based and paper-based homework for general chemistry. *The Chemical Educator*, 13, 264-269.
- Gamire, E., & Pearson, G. (Eds.). (2006). *Tech tally: Approaches to assessing technological literacy*. Washington, DC: National Research Council and The National Academy of Engineering.
- Glaser, R. (1990). *Toward new models of assessment*. Pittsburg, PA: Learning and Research Development Center.
- Grouws, D. A. (Ed.). (1992). *Handbook of research on mathematics teaching and learning*. Reston, VA: National Council of Teachers of Mathematics.
- Hirsch, C. R., Coxford, A. F., Fey, J. T., & Schoen, H. L. (1994). *Implementing the CPMP curriculum*. Kalamazoo: Core-Plus Mathematics Project, Western Michigan University.
- Hirsch, C. R., Coxford, A. F., Fey, J. T., & Schoen, H. L. (1995, November). Teaching sensible mathematics in sense-making ways with CPMP. *The Mathematics Teacher*, 88(8), 694-700.
- Horizon Research, Inc. (1999). Local systemic change through teach enhancement classroom observation protocol. Retrieved June 28, 2009, from [www.horizon-research.com/instruments](http://www.horizon-research.com/instruments)
- Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards: How to improve evaluations of students*. Thousand Oaks, CA: Corwin Press.
- Joint Committee on Standards for Educational Evaluation. (2009a). *The personnel evaluation standards: How to assess systems for evaluating educators*. Thousand Oaks, CA: Corwin.

- Joint Committee on Standards for Educational Evaluation. (2009a). *Assessment for learning benchmarking project*. Unpublished document.
- Joint Committee on Standards for Educational Evaluation. (2009b). *The personnel evaluation standards: How to assess systems for evaluating educators*. Thousand Oaks, CA: Corwin.
- Keeley, P. (2008) *Science formative assessments: 75 practical strategies for linking assessment, instruction, and learning*. Arlington, VA: National Science Teachers Association Press.
- Kulm, G. (1994). *Mathematics assessment: What works in the classroom*. San Francisco: Jossey-Bass.
- Lambdin, D. V. (1993). The NCTM's 1989 evaluation standards: Recycled ideas whose time has come? In N. L. Webb & A. F. Coxford (Eds.), *Assessment in the mathematics classroom*, NCTM 1993 Yearbook (pp. 7-16). Reston, VA: National Council of Teachers of Mathematics.
- Martinez, J. G. R., & Martinez, N. C. (1992). Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology*, 62.
- McCallum, B., & Clarke, S. (2001). *Gillingham partnership formative assessment Project teacher inservice materials*. London: Institute of Education.
- McKeachie, W. (1986). *Teaching tips: A guidebook for the beginning college teacher*. Lexington, MA: D. C. Heath.
- Michigan Department of Education. (2008). *Frequently asked questions for MITTC*. Retrieved December 10, 2008, from [www.michigan.gov/mde/0,1607,7-140-6530\\_5683\\_5857-116214--,00.html#mttc1](http://www.michigan.gov/mde/0,1607,7-140-6530_5683_5857-116214--,00.html#mttc1)
- National Assessment of Educational Progress (NAEP)*. Retrieved June 20, 2009, from the National Center for Educational Statistics, U.S. Department of Education Institute of Education Sciences Web site: [nces.ed.gov/nationsreportcard/about/](http://nces.ed.gov/nationsreportcard/about/)
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (1995). *Assessment standards for school mathematics*. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2001). *Knowing what students know*. Washington, DC: National Academy Press.
- O'Brien, J. G., Millis, B. J., & Cohen, M. W. (2008). *The course syllabus: A learning-centered approach*. San Francisco: Jossey-Bass.
- Program for International Student Assessment (PISA)*. (2007). Retrieved June 30, 2009, from the Organisation for Economic Co-Operation and Development Web site: [www.pisa.oecd.org](http://www.pisa.oecd.org)
- Qualifications and Curriculum Authority. (.n.d.). *Assessment for learning guidance*. Retrieved June 27, 2009, from [www.qca.org.uk/qca\\_4334.aspx](http://www.qca.org.uk/qca_4334.aspx)
- Remillard, J. T., Herbal-Eisenmann, B. A., & Lloyd, G. M. (2009). *Mathematics teachers at work: Connecting curriculum materials and classroom instruction*. New York: Routledge.
- Romberg, T. A. (Ed.). (1995). *Reform in school mathematics and authentic assessment*. Albany, NY: SUNY Press.
- Schoen, H. L., Bean, D. L., & Ziebarth, S. W. (1996). Embedding communication throughout the curriculum. *Communication in mathematics: K-12 and beyond*, NCTM 1996 Yearbook (pp. 170-179). Reston, VA: National Council of Teachers of Mathematics.
- Scottish Qualifications Authority. (2006). *Assessment is for learning: Self-assessment toolkit*. Glasgow: Learning and Teaching Scotland.
- Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne, & M.

- Scriven (Eds.), *Perspectives of curriculum evaluation, 1*, pp. 39-83. Chicago: Rand McNally.
- Senk, S. L., Beckmann, C. E., & Thompson, D. R. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*, 187-215.
- Shavelson, R. J., Yin, Y., Furtak, E. M., Ruiz-Prino, M. A., Ayala, C. C., Young, D. B., Tomita, M. K., Brandon, P. R., & Pottenger, F. M. III. (2008). *On the role and impact of formative assessment on science inquiry teaching and learning in assessing science learning: Perspectives from research and practice*. Arlington, VA: National Science Teachers Association Press.
- Silver, E. A. (1992). *Assessment and mathematics education reform in the United States*. Pittsburg, PA: Learning and Research Development Center.
- Standard & Poor's. (2009). *SchoolMatters*. Retrieved June 27, 2009 from [www.schoolmatters.com](http://www.schoolmatters.com)
- Steinbring, H., Bussi, M.G., & Sierpiska, A. (1998). *Language and communication in the mathematics classroom*. Reston, VA: National Council of Teachers of Mathematics.
- Stiggins, R. J. (1988). Revitalizing classroom assessment: The highest instructional priority. *Phi Delta Kappan, 69*(5), 363-368.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.
- Stiggins, R. J. (2002). Assessment crisis: The absence OF assessment FOR learning. *Phi Delta Kappan, 83*(10), 758-765.
- Trends in International Mathematics and Science Study (TIMSS)*. (2007). Retrieved June 30, 2009 from the National Center for Educational Statistics, U.S. Department of Education Institute of Education Sciences Web site: [nces.ed.gov/timss/](http://nces.ed.gov/timss/)
- van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and authenticity*. White Plains, NY: Addison Wesley.
- Webb, N. L. (1994). Thoughts on assessment in the mathematics classroom. In N. L. Webb & A. F. Coxford (Eds.), *Classroom assessment in mathematics* (pp. 101-114). Lanham, MD: Rowman, Littlefield.
- Webb, N. L., & Coxford, A. F. (1993). *Assessment in the mathematics classroom*. NCTM 1993 Yearbook. Reston, VA: National Council of Teachers of Mathematics.
- Whiting, B., Van Burgh, J. W., & Reder, G. F. (1995). *Mastery learning in the classroom*. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction No. ED 382 688)
- William, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 1053-1098). Reston, VA: National Council of Teachers of Mathematics.
- Wilson, L. D., & Kenney, P. A. (2003). Classroom and large-scale assessment. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 53-67). Reston, VA: National Council of Teachers of Mathematics.