

# ***BACK TO THE FUTURE***

## ***Reclaiming Effective Mathematics Assessment Strategies***

**Gerald Kulm**

*Texas A&M University*

In this article, I describe some efforts initiated about 2 decades ago to reform mathematics assessment. Many, if not most, of those ideas have gradually eroded not only from school practice but also from the knowledge base of current day thinking and research in mathematics assessment. In addition to exploring and explaining the policy, political, and economic factors that led to the demise of these ideas, I want to recall and encourage current day mathematics educators to reconsider them and explore ways that they might be adapted and implemented once again. It is necessary first to provide a brief account of these developments and the forces that I believe led to the status of mathematics assessment today.

More than 20 years ago, I edited a book titled *Assessing Higher Order Thinking in Mathematics* (Kulm, 1990) which was one of the first attempts by some leading mathematics educators and assessment experts to develop ideas for reforming the traditional paper-and-pencil, multiple-choice approaches to assess-

ing student achievement. In the book, I stated that “More comprehensive and thoughtful approaches should begin to replace the simplistic *report cards*, group comparisons, and minimal competency emphasis of earlier accountability and indicator efforts” (p. 77). The authors of chapters in the book addressed the issues of the day, including comprehensive assessment of higher order thinking (Baker), alignment of assessment and curriculum (Pandey), new assessment frameworks (Romberg, Zarinnia, & Collis), assessment of problem solving (Lester & Kroll), uses of technology in assessment (Lesh; Strong; Lipson, Faletti, & Martinez). Authors also reported research on assessing multiple dimensions of student learning (Nichols, Cobb, Yackel, Wood, & Wheatley), student knowledge schemas (Marshall), quantitative reasoning (McKnight), and structured problem solving items (Wilson).

The book was inspired in part by a pilot study by the National Assessment of Educational Progress (NAEP), which explored and

---

• **Gerald Kulm**, Professor Emeritus and Senior Professor of Mathematics Education in the Department of Teaching, Learning & Culture, Texas A&M University. E-mail: [gkulm@tamu.edu](mailto:gkulm@tamu.edu)

---

Middle Grades Research Journal, Volume 8(2), 2013, pp. 1–10  
Copyright © 2013 Information Age Publishing, Inc.

ISSN 1937-0814  
All rights of reproduction in any form reserved.

provided examples of innovative assessment items that emphasized student “hands-on” performance (NAEP, 1987). I was a member of the team of mathematics and science educators working with NAEP to develop about a dozen tasks, including group-administered open-ended problems, “station activities” that required students to use equipment, and “experiments” administered to individual students to investigate and discuss findings with the administrator. The tasks assessed skills and knowledge in classifying, observing and making inferences, formulating hypotheses, and interpreting data. NAEP field-tested the tasks with national sample of about 1,000 third, seventh, and eleventh grade students. The results indicated that conducting hands-on assessment is “feasible and extremely worthwhile” (p. 7). Participating teachers and administrators were enthusiastic about the promise of new information about student learning that might be obtainable through these new types of assessment tasks.

As the momentum grew for these new assessment ideas, teachers, schools, and some states began to pilot test them. Various terms to describe the new assessments began to appear: authentic, performance, alternative. Grants from states and the federal government became available to develop new assessments (e. g. Kulm, 1994c), train teachers in their use (e. g. Kulm, Telese, & McMullen, 1993), and do research on their effectiveness (e.g., Kulm, 1994a, 1994b). At the same time, several states experimented with, and some implemented the new assessments as part of their state mathematics achievement and accountability measures. There was a national sense, in parallel with the reforms and mathematics standards movements (National Council of Teachers of Mathematics, 1989), that these new assessment approaches could reflect and reinforce the goals that students should learn not only the skills and procedures of mathematics, but also the concepts, processes, and problem solving applications that are essential in the modern world (Mathematics and Science Education Board, 1993; Webb, 1993). The tradi-

tional, standardized, mainly multiple-choice tests were not well-suited to assessing these cognitive outcomes. The availability of new assessments in mathematics could serve teachers in their efforts to implement the mathematics standards.

### ***THE NEED FOR AND PURPOSES OF ASSESSMENT***

In the past decade or so, I have come to the conclusion that many educators, from classroom teachers, to university professors, to administrative leaders have either forgotten or chosen to narrow their ideas of the meaning and purpose of assessment. Sadly, this may include mathematics educators. The focus on tests for the sole purpose of accountability, competition for funding, or international comparisons has resulted in a distorted view of assessment. This constricted view has resulted in several disastrous consequences including constraining curriculum to the content that is easily tested, forcing teachers to “teach to the test,” spending inordinate amounts of the school year on preparing for and taking tests, and unreliable judgments of students’ abilities. Ironically, all of these deficiencies were the intended target of those, including myself, who favored new and reformed assessments. This is not to say that assessment is not necessary or important. However, assessment should not drive the curriculum, instruction, funding, and everything else in education. Is it possible to regain some of the original intents and successes of assessment? Can the earlier pitfalls and mistakes be corrected and overcome? To even begin to answer to these questions requires a review of the true purposes of assessment. We can use mathematics assessment as an example because it was been the first target content area for new innovations in assessment, and also the first content area to fall to the ills of the current state of disrepair.

What are the needs for and purposes of assessment? We can try to start with a defini-

tion. There are clearly two different views of the meaning of assessment [emphases added]:

1. Assessment is “the process of gathering evidence about a *student's knowledge* of, ability to use, and disposition toward mathematics and of making inferences from that evidence for a variety of purposes” (National Council of Teachers of Mathematics, 1989, p. 3); determining *what students know* (Webb & Briars, 1990); “finding out *what each student is able to do*, with knowledge, in context (Wiggins, 1996/1997, p. 19).
2. Assessment is the process of collecting, interpreting, and synthesizing information for *decision making* (Airasian, 1991); doing evaluation—“an *accounting* of what is” (Webb, 1992, p. 663).

The first definition addresses the primary purpose of assessment—finding out what student know, think, and are able to do. The second definition focuses on accountability and evaluation; that is, assigning grades or other values to the obtained results of the assessment process.

These two definitions also have implications for the audience of assessment. There are many potential recipients and users of assessment data. The evolution of assessment reflects a gradual but inexorable shift from the original, intended users to those who are most removed from the classroom, yet have powerful leverage and effect on the classroom where mathematics learning takes place (Kulm, 1994d). The list of those who make use of assessment results includes:

- students;
- parents or guardians;
- teachers;
- school curriculum and professional development specialists;
- school and district administrators;
- school boards;
- state-level departments, agencies, and administrators;

- state-level school boards;
- state-level legislators and governors;
- federal-level departments, agencies, and administrators; and
- U.S. President.

There are others that could be added to this list, including textbook publishers, test publishers, colleges and universities, and even real estate agents.

It is clear to see that I have ordered this list from “bottom-to-top.” In my opinion, the most important entity on the list is the student, who is at the same time the one who is the object and “producer” of assessment outcomes and the most important consumer and decision-maker in the process. Do the results of assessment impel the student to become confident about doing math, try harder, work toward a STEM career, become anxious about math, dislike the teacher, or drop out of school? Any of these reactions and many more can be the reaction of a student to assessments. A similar list of reactions and decisions could easily be generated for the next most important groups on the list: parents and teachers. As one goes down this list, further removed from the student and the classroom, the nature of the reactions to the results and types of decisions change. For example, state and federal departments and administrators may decide to begin new initiatives for curriculum or research, or to provide funding for particular programs or groups. From the individual student to the very top levels, all of these consumers, uses, and decisions made on the basis of assessment results are legitimate and important. The decisions were made even more important with the passage of the No Child Left Behind Act (2002), which resulted in high-stakes testing and assessment in general being thrust upon students, parents, and educators. Currently, most U.S. schools are using high-stakes testing to measure student performance.

According to the American Psychological Association (2001), tests should not only provide valuable feedback to teachers regarding individual student performance, curriculum,

and teaching methods, but they can also help to hold schools and school districts accountable for students' performances. On the other hand, high-stakes testing is typically a single "snapshot" of a student's academic achievement. Schools give the students one chance to successfully pass a paper-and-pencil test. In this situation, tests often "have unintended and potentially negative consequences for individual students, groups of students, or the educational system more broadly" (American Psychological Association, 2001, p. 5). The bottom line is: we have nearly lost sight of a very important fact: The type of assessment depends on the use or purpose of the assessment result. In particular, a once-every-year, high-stakes, easily-administered, paper-and-pencil mathematics test required of every student in the state does not serve the purpose of every entity on the list of consumers of assessment. This type of test, which has evolved as the de facto model of student mathematics assessment, has been the result of a shift toward the top level of the list as being the most important consumer. Simultaneously, parents and teachers have accepted, or have been coerced, into believing that the results of these tests can serve their needs and purposes as well.

It is not difficult to point out the many reasons such tests do not effectively meet the needs of teachers, parents or, least of all, students in mathematics classrooms. Necessarily, these tests "cover" only a fraction of the curriculum; they provide only surface level information about the extent or depth of students' knowledge or understanding; they do not provide early diagnostic information for planning instruction; they are susceptible to bias and underestimation of many students' multiple abilities; and worst of all, they promote the notion that learning mathematics is all about passing a test. Researchers (cf. Dorr-Bremme & Herman, 1983; Herman & Golan, 1991; Kellaghan & Madaus, 1991; Shepard, 1991; Smith & Rottenberg, 1991) have concluded "the time focused on test content has narrowed the curriculum by overemphasizing basic-skill

subjects and neglecting higher-order thinking skills" (Herman, 1992, p. 74). Although some test scores have improved through the questionable practice of "teaching to the test," the approach does not result in meaningful learning for the students (Herman, 1992).

### **RECLAIMING MATHEMATICS ASSESSMENT**

The good news is that it is possible to change the dire situation we now find in mathematics assessment. Mathematics education leaders in schools and universities can reclaim the true meaning and purpose of enlightened mathematics assessment. This movement back to best practices and research-based assessment does not require huge amounts of funding, just some effort and a little risk-taking. Some years ago, I directed an exploratory research project to help K-12 teachers in Texas learn about new strategies for mathematics assessment. The results of the project are published in the book *Mathematics Assessment: What Works in the Classroom* (Kulm, 1994a). One of the high school mathematics teachers in that project was assigned in her school to teach a class of seniors who had failed the state mathematics test twice. They had one more chance to pass it or not graduate. The school had assigned these students to a math class devoted solely to preparing for the test. I asked each of the teachers in the project to select one of their classes to try some new assessment strategies that we would develop and pilot test during the year. Along with the new assessment strategies, the project also encouraged and supported the teachers to try more student-centered instruction, including problem solving and group work. This brave teacher decided to use her math *test-prep* class as the group to try these new approaches. The usual daily routine in these kinds of classes was to review math procedures they had failed to understand for 11 years (or perhaps not tried to understand), practice multiple-choice test items, and do more review and practice on the items they missed. Instead, the

teacher selected open-ended problems to solve, working in small groups. She asked students to write journals to explain to a friend how to do math procedures, and gave innovative assignments such as pretending to be a geometric figure and describing their “personality” and jobs that the figure could perform. In short, she regained their engagement, interest, and self-confidence in doing mathematics. Her assessments of their work and her feedback to them were not multiple-choice items from previous state tests. She developed rubrics for the journals, problem solutions, and class presentations that reflected the level and depth of their knowledge and skill. She involved the students in developing these rubrics so they knew what was important to learn and how they would be evaluated. In the spring, every student in the class passed the state test! This result should not really be surprising. That year, the state mathematics supervisor wondered why seniors were not passing the test and decided to do some interviews of the students. The great majority of these seniors, across the state, replied that they probably could have passed; they just didn’t care or didn’t try very hard.

### ***THE RISE AND FALL OF NEW ASSESSMENT STRATEGIES IN MATHEMATICS***

Most educators know, even intuitively, that there are two general types of assessment: formative and summative. Formative assessments are intended to improve or modify instruction and learning. Summative assessments are tests whose purpose is to make success or failure decisions about instructional activities that have already taken place (Popham, 2005). As stated above, summative assessments have received the major share of recent attention, especially at the state and national levels. This attention has translated to the mathematics classroom level through pressures for accountability. As a consequence, few teachers have the time, training, or opportunity to use forma-

tive assessments, which are the most relevant to their work. Why has this shift in emphasis to the accountability aspect of mathematics assessment taken place?

In the early 1990s, as mathematics reform began to gain momentum, schools, districts, and states began experimenting with alternative forms of assessment, mainly because they realized that the traditional standardized tests failed to measure many of the important aspects of learning. State wide testing programs, which employed traditional standardized tests, were the target for the assessment reforms. Because these traditional tests did not support many of the teaching strategies that mathematics teachers were being encouraged to implement in their classrooms (Darling-Hammond, Ancess, & Falk, 1995), the state tests appeared to offer a strong leverage point for reforming mathematics classroom practice. I was among many mathematics educators who believed that teaching to these alternative tests would result in improved student learning and achievement (Wiggins, 1989). This belief was supported, for example, by results such as those of my exploratory study (Kulm, 1994b). Looking back at this period of excitement about alternative assessments, it is apparent why the widespread use and adoption of these approaches failed to materialize. The early state adopters (e.g., Kentucky, Maryland, Vermont) of mathematics assessment alternatives, including performance assessments such as those initiated by the NAEP (1987) pilot study, encountered several insurmountable barriers. The amount of time and effort needed for the proper development, design, and actual implementation of performance assessments was enormous for a statewide effort. Also, research findings began to indicate that these assessments were not reliable for the purpose of accountability. Assessment of student understanding usually cannot be inferred from a single response on a single task. Instead, a “variety of tasks are needed to generate a profile of behavioral evidence” (Hiebert & Carpenter, 1992, p. 89). Because performance tasks took longer to complete, only a few of

them could be included in a test, making reliability difficult to establish. The scoring of the performance assessments can be burdensome and typically adds additional threats reliability due to the subjective judgment by the raters. For these reasons, performance assessments in mathematics were gradually discarded as tools for large-scale tests and states returned to the more traditional paper-and-pencil, usually multiple-choice, items. Unfortunately, this resulted in the proverbial “throwing out the baby with the bath water.” That is, alternative assessments were largely eliminated in mathematics classrooms at the same time, since mathematics teachers were under increasing pressure to prepare for high-stakes tests, with the advent of No Child Left Behind. The teachers’ natural reaction was to use the multiple-choice formats as the primary tool not only for their students to practice but also to guide instruction.

### ***Alternative Assessment in Mathematics Classrooms***

The following summary is an attempt to recall some basic and important ideas about mathematics assessment. Partly due to the pressures at the top, and partly due to new teachers entering classrooms, it seems that much of this knowledge and practice has been forgotten or overlooked, especially in the past decade. Much of the information provided here draws from the active and productive work of mathematics educators and measurement specialists about 20 years ago. Here, I suggest three strategies for mathematics classroom assessment that are superior to the current approaches that rely mainly on practicing test items. As I stated earlier, teachers can use these strategies, along with student-centered instruction and, simultaneously, their students will do well on state mathematics tests. More importantly, these strategies are flexible for diverse students, making it possible for more students to exhibit their mathematical knowledge and understanding.

### ***Performance Assessment<sup>1</sup>***

This strategy provides students with multiple ways to demonstrate their understandings and skills by completing tasks in which they are asked to respond orally, in writing, and/or by constructing a product (Gronlund, 2003; Nitko, 2001; Popham, 2005). “Performance tasks give the teacher and students feedback as to the current level of achievement and suggest ways for teacher and students to improve future outcomes” (Huetinck & Munshin, 2004, p. 373). This type of assessment allows the mathematics teacher to observe “students’ work and their performance in ways that can evaluate the performances of students, classes, and whole schools” (Darling-Hammond et al., 1995, p. 10). Student work is evaluated according to criteria expressed in the form of a rubric, which emphasizes and communicates to the student, important components and level of achievement. Performance assessments are broader in scope than traditional pencil and paper assessments, since they require and incorporate the entire context of a task. They are more authentic and are more likely to elicit a student’s full repertoire of skills and reflect a range of goals the teacher wants the students to meet (Campbell, Melenyzer, Nettles, & Wyman, 2000).

There are various forms of performance assessment including oral presentations along with student written work, experiments, debates, teacher observation and inventories, inquiries, and portfolios (Archbald & Newman, 1988; Meisels, 1996/1997; O’Neil, 1992). These various forms of performance assessments are also valuable tools in special education where paper-and-pencil tests are not applicable or able to test important skills (Stiggins, 1987). Although there are various forms of performance assessment, all forms have the same three features in common: multiple evaluative criteria, prespecified quality standards, and judgmental appraisal (Popham, 2005). Because performance assessments are integrated within the curriculum and not separate tests, they are ideal for classroom use. Perfor-

mance assessments can become an integral part of a classroom mathematics instruction, reducing student stress and test anxiety. Instead, students are challenged to meet their own and their teacher's expectations of proficiency on tasks (Huetnick & Munshin, 2004). Finally, performance assessments can be used in group or individual settings, increasing their flexibility for use in the mathematics classroom.

Implementing performance assessments at the middle grades level in mathematics has many advantages for students, teachers, parents and administrators. Performance assessments allow students the opportunity to display more than just speed and accuracy. In addition, students have more opportunities to do their own organizing and thinking as they solve more authentic problems, allowing them to focus more on the process rather than simply memorizing rules or procedures. In performance assessment, students are involved in real tasks that can be tailored to their interests so they are engaged and motivated to do mathematics. For teachers, assessment is an integral part of instruction and student learning. By utilizing performance assessment in their mathematics classrooms, teachers can increase the quality of information they can obtain about student understanding and ability to do mathematics. They are able to gain more comprehensive information for instructional decisions about student misconceptions or errors. For administrators and parents, the use of performance assessment in classrooms provides direct evidence of students learning to think and use mathematics in new situations.

### ***Informal Classroom Assessment***

In many lessons, a teacher presents new material, or initiates a task for student exploration, then asks students to work individually or in groups to explore or practice. In order to make a formative or informal assessment of student learning as they complete steps in the exploration or practice on the skills or con-

cepts, teachers often walk around the classroom, checking work by individual students or groups, and asking questions. Formative assessment, as defined by Black and William (1998), is "all those activities undertaken by teachers, and/or by their students, which provide information used as feedback to modify the teaching and learning activities in which they are engaged" (pp. 1-2). When teachers are trained to help students with self-assessment and when students participate in self-assessment on a daily basis, research shows achievement will improve (Fernandes & Fontana, 1996).

According to Cole (1999), walking around the classroom is an immediate, efficient assessment method available to teachers, allowing them to determine student progress and remedy developing problems on the spot. Rodriguez (2004) found that teacher assessment through observations of student work was highly correlated with the teachers' use of this information, whereas there were only modest correlations between other types of assessment and teachers' use of them to diagnose or remediate learning difficulties. This result may be due to the fact that with other assessment methods such as journals, homework, or quizzes, feedback is delayed by a day or more. Cole (1999) noted that teachers usually observe each group working for about half a minute—just long enough to verify that the group is on task and is not having major difficulties with the work. This conclusion is supported by the work of Ding, Li, Piccolo and Kulm (2007) who found that many teachers simply walked by and made brief comments rather than interacting with groups. These short observations are essential, but they do not supply much information about group processes or mathematics understanding. Some teachers conduct more active interactions with students or group, lasting three to five minutes, or longer during which the teacher is able to probe for understanding or guide students to revise their thinking (Cole, 1999; Ding, et al., 2007).

### ***Classroom Discourse and Questioning***

Student-teacher discourse can have a significant impact on students' learning when the students are trying to understand a concept (Clarke, 1988). In assessment it is important that the student construct the same meaning of the task as the teacher intended (Lorsbach, Tobin, Briscoe, & LaMaster, 1992). According to Bangert-Drowns, Kulik, Kulik, and Morgan (1991), students who received feedback to correct their errors were less likely to make those errors again. Carpenter, Fennema, and Franke (1996) developed Cognitively Guided Instruction (CGI) which "provides a framework in which teachers construct and test their own models of students' thinking to guide their instructional practices" (p. 5). Teachers who understand how students think and construct mathematics ideas and concepts are more successful and effective.

An important way that teachers can assess student understanding in a classroom is by asking questions (Sahin & Kulm, 2008; Wilson & Kenney, 2003). However, it has been well-established that the majority of the questions that teachers ask require simple recall and do not evaluate higher order thinking (Stiggins, Griswold, & Wikelund, 1989). As teachers assess student understanding, Leinhardt (1983) found that novice and expert teachers differed in their explanations about student's achievement and potential for academic success. The expert teachers were more precise in analyzing their students' success and the reasons behind it. Leinhardt and Greeno (1986) have found that less knowledgeable teachers are less effective in their questioning strategies about students' performance and do not have sufficient skills to gather information. Novice teachers need training in order to learn how to better ask questions that promote student higher order thinking, "to enhance students' learning, classroom assessment must include adequate dialogue between teachers and students" (Wilson & Kenney, 2003, p.56). Mathematics achievement could be improved dramatically simply

by developing teachers' skills in asking questions.

### ***Summary***

The changing student demographics and the importance that all students learn and understand mathematics make it critical that we use the very best practices that are available (Telese & Kulm, 1995). I believe that we have allowed those who are far removed from the classroom to influence mathematics teaching and assessment in ways that are counter to research results and best practice. The pressure of accountability at the state and national level as the primary purpose of assessment has overshadowed what we know is really the main purpose: student mathematics learning.

I acknowledge that alternatives such as performance assessment are hard work and require time and effort by teachers. There are at least two ameliorating factors that can help. First, many mathematics educators have recognized that the "mile wide inch deep" mathematics curriculum must be changed to focus on fewer topics in more depth. Performance assessment is very much in line with this idea. By significantly reducing the amount of time and effort on test practice and reteaching, mathematics teachers can integrate performance assessment with authentic, longer term, problem-based tasks that promote in-depth understanding. These kinds of tasks naturally require and involve higher levels of teacher-student discourse and teacher questions that promote student understanding.

A second issue of time and effort is the increased work in scoring and recording that is required in alternative assessment strategies. The technology that has developed during the past decade can provide tools to support mathematics teachers and ease this burden in implementing alternative assessment. Hand-held smart devices can be used in real time by teachers as they observe, interact, and evaluate students' work on tasks. Web-based systems can also support the communication of outcomes among teachers, parents, and adminis-

trators. These technologies can bring to reality the visions of mathematics assessment strategies previously imagined.

## NOTE

- 1 This section on performance assessment is adapted from Mohr (2006).

## REFERENCES

- Airasian, P. W. (1991). *Classroom assessment*. New York, NY: McGraw-Hill.
- American Psychological Association. (2001, May). *Appropriate use of high-stakes testing in our nation's schools*. Retrieved from <http://www.apa.org/pubs/info/brochures/testing.aspx>
- Archbald, D. A., & Newman, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.
- Bangert-Drowns, R., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238.
- Black, P., & William, D. (1998). Assessment and classroom Learning. *Assessment in Education*, 5, 7-74.
- Campbell, D. M., Melenzyer, B. J., Nettles, D. H., & Wyman, R. M., Jr. (2000). *Portfolio and performance assessment in teacher education*. Boston, MA: Allyn and Bacon.
- Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal*, 97(1), 3-20.
- Clarke, J. (1988). Classroom dialogue and science achievement. *Research in Science Education*, 18, 83-94.
- Cole, K. A. (1999). Walking around: Getting more from informal assessment. *Mathematics Teaching in the Middle School*, 4, 224-227.
- Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work*. New York, NY: Teachers College Press.
- Ding, M., Li, X., Piccolo, D., & Kulm, G. (2007). Teacher interventions in cooperative-learning mathematics classes. *Journal of Educational Research*, 100, 162-175.
- Dorr-Bremme, D., & Herman, J. (1983). *Assessing student achievement: A profile of classroom practices*. Los Angeles, CA: UCLA, Center for the Study of Evaluation.
- Fernandes, M., & Fontana, D. (1996). Changes in control beliefs in Portuguese primary school pupils as a consequence of the employment of self-assessment strategies. *British Journal of Educational Psychology*, 66, 301-313.
- Gronlund, N. E. (2003). *Assessment of student achievement* (7th ed.). Boston, MA: Allyn & Bacon.
- Herman, J. L. (1992). What research tells us about good assessment. *Educational Leadership*, 49(8), 74-78.
- Herman, J. L., & Golan, S. (1991). *Effects of standardized testing on teachers and learning—Another look*. CSE Technical Report #334. Los Angeles, CA: Center for the Study of Evaluation.
- Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65-97). Reston, VA: National Council of Teachers of Mathematics.
- Huetinck, L., & Munshin, S. N. (2004). *Teaching mathematics for the 21st century: Methods and activities for Grades 6-12* (2nd ed., pp. 119-160). Upper Saddle, NJ: Pearson.
- Kellaghan, T., & Madaus, G. (1991). National testing: Lessons for America from Europe. *Educational Leadership*, 49(3), 87-93.
- Kulm, G. (1990). *Assessing higher order thinking in mathematics*. Washington, DC: American Association for the Advancement of Science.
- Kulm, G., Telese, J., & McMullen, B. G. (1993). *Alternative assessment in mathematics*. Texas Mathematics Staff Development Module 31. Austin, TX: Texas Education Agency.
- Kulm, G. (1994a). *Mathematics assessment: What works in the classroom*. San Francisco, CA: Jossey-Bass.
- Kulm, G. (1994b). The effects of alternative assessment on mathematics teaching. *Proceedings of the EMT Conference on 30 Years of Excellence*, University of Georgia, Athens, GA.
- Kulm, G. (1994c). Situated performance assessment. In *A database and catalog of alternative assessments, Second Edition*. Portland, OR: Northwest Regional Education Laboratory.

- Kulm, G. (1994d). The control of assessment. In G. Kulm & S. M. Malcom (Eds.), *Science assessment in the service of reform*. Hillsdale, NJ: Erlbaum. Originally published in 1991 by AAAS Press.
- Leinhardt, G. (1983, April). *Routines in expert math teachers' thoughts and actions*. Paper presented at the annual meeting of American Educational Research Association, Montreal, Canada.
- Leinhardt, G., & Greeno, J. G. (1986). The cognitive skill of teaching. *Journal of Educational Psychology*, 78(2), 75-95.
- Lorsbach, A. Q., Tobin, K., Briscoe, C., & LaMaster, S. U. (1992). An interpretation of assessment methods in middle school science. *International Journal of Science Education*, 14, 305-317.
- Mathematical Sciences Education Board. (1993). *Measuring what counts: A conceptual guide for mathematics assessment*. Washington, DC: National Academy Press.
- Meisels, S. J. (1996/1997). Using work sampling in authentic assessments. *Educational Leadership*, 54(4), 60-65.
- Mohr, M. J. (2006). *An assessment of preservice teachers' mathematics knowledge for teaching: Middle grades mathematics*. Unpublished doctoral dissertation, Texas A&M University.
- National Assessment of Educational Progress. (1987). *Learning by doing: A manual for teaching and assessing higher order thinking in science and mathematics*. Report No. 17-HOS-80. Princeton, NJ: Educational Testing Service.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluations standards for school mathematics*. Reston, VA: Author.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 1114, 115 Stat. 1473 (2002). Washington, DC: U.S. Government Printing Office.
- O'Neil, J. (1992). Putting performance assessment to the test. *Educational Leadership*, 49(8), 14-19.
- Popham, W. J. (2005). *Classroom assessment: What teachers need to know* (4th ed.). Boston, MA: Pearson.
- Rodriguez, M. C. (2004). The role of classroom assessment in student performance on TIMSS. *Applied Measurement in Education*, 17(1) 1-27.
- Sahin, A., & Kulm, G. (2008). Sixth grade mathematics teachers' intentions and use of probing, guiding, and factual questions. *Journal of Mathematics Teacher Education*, 11, 221-241.
- Shepard, L. (1991). Will national tests improve student learning? *Phi Delta Kappan*, 73, 232-238.
- Smith, M. L., & Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10(4), 7-11.
- Stiggins, R. J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and practice*, 6(3), 33-42.
- Stiggins, R. J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26, 233-246.
- Telese, J., & Kulm, G. (1995). Assessment scenarios for at-risk students in mathematics. *Urban Review*, 27, 121-140.
- Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 661-683). New York, NY: Macmillan.
- Webb, N. L. (Ed.) (1993). *Assessment in the mathematics classroom*. Reston, VA: National Council of Teachers of Mathematics.
- Webb, N. L., & D. Briars, D. (1990). Assessment in mathematics classrooms, K-8. In T. J. Cooney (Ed.), *Teaching and learning mathematics in the 1990s* (pp. 108-117). Reston, VA: National Council of Teachers of Mathematics.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 141-147.
- Wiggins, G. (1996/1997). Practicing what we preach in designing authentic assessments. *Educational Leadership*, 54(4), 18-25.
- 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.