INFORMATION CAPSULE
Research Services

Vol. 0804
October 2008

Christie Blazer, Supervisor

How Interim Assessments Affect Student Achievement

At A Glance

Teachers have always used tests to measure students' mastery of knowledge and skills. However, educators have recently begun to incorporate formative interim assessments into the classroom learning process. These assessments allow them to adapt instructional programs to better meet students' academic needs throughout the course of the school year. Studies have indicated that frequent testing, along with the provision of corrective feedback, has a positive impact on students' performance and is particularly effective for low performing students. This Information Capsule discusses some of the reasons educators should consider using interim assessments as a teaching and learning tool and summarizes the characteristics of effective interim assessment programs. Research conducted on the impact of interim assessments on student achievement, as well as their predictive validity, is also reviewed. Finally, an overview of Miami-Dade County Public Schools' interim assessment program is provided.

Teachers have always used tests to measure students' mastery of knowledge and skills. However, educators have recently begun to incorporate formative interim assessments into the classroom learning process. These assessments allow them to adapt instructional programs to better meet students' academic needs throughout the course of the school year. There are many labels for these tests, such as interim, benchmark, common, and short-cycle assessments, but they all serve three primary purposes: to provide diagnostic information that allows instructional programs to be adapted to better meet students' learning needs; to evaluate the effectiveness of various curricular and instructional practices; and to predict performance on end-of-year state tests (Stiggins & Chappuis, 2008; Brown & Coughlin, 2007; Dufour, 2007; Vendlinski et al., 2007; Herman et al., 2006; Peariso, 2006; Perie, Gong, & Marion, 2006; Perie, Marion, & Gong, 2006; Sharkey & Murnane, 2006; Olson, 2005a; Shanahan et al., 2005; Stiggins, 2005a).

Typically, interim assessments are aligned to state or district standards for academic content and administered three to five times during the school year. Most school districts assess reading, math, and science at grades three through eight, although some districts have extended the tests to cover grades two through ten. Teachers usually have immediate access to test results to help direct the instructional process. Scores are reported in relation to specific state standards or grade-level expectations in order to provide information regarding students' strengths and weaknesses. Data are provided at the student, class, school, and district levels so results can be used not only to design instructional interventions, but to inform the district's programmatic decisions (Stiggins & Chappuis, 2008; Beauchamp, 2007; Olson, 2005a; Jones et al., 2005; Business Wire, 2003).

Research Services
Office of Assessment, Research, and Data Analysis
1500 Biscayne Boulevard, Suite 225, Miami, Florida 33132
(305) 995-7503 Fax (305) 995-7521

Reasons to Use Interim Assessments

Researchers recommend that educators should consider using interim assessments as a teaching and learning tool for the following reasons:

- Results from annual state tests arrive too late to help educators revise instructional programs or address students' learning difficulties. In addition, state tests provide only basic information, such as a student's strengths at the sub-score level, and offer little useful data on how to address performance weaknesses. In contrast, interim assessments can be administered periodically throughout the school year, generating student test score data to guide instructional interventions (Stiggins & Chappuis, 2008; Heritage, 2007; Depka, 2006; Perie, Gong, & Marion, 2006; Herman & Baker, 2005; Stiggins, 2005b; Rabinowitz & Ananda, 2001).

- Many policymakers and educators believe that interim assessments represent the most effective strategy for determining whether the curriculum is being taught and mastered (Olson, 2007; Rutherford County Schools, 2007).

- Interim assessments allow teachers to monitor student progress so instructional adjustments can be made throughout the course of the school year (Brown & Coughlin, 2007; Heritage, 2007; Wireless Generation, 2007; Herman et al., 2006; Herman & Baker, 2005; Shanahan et al., 2005).

- Interim assessments can be used to predict students' future performance and help them prepare for end-of-year or state accountability tests (Peariso, 2006; Shanahan et al., 2005).

- Locally developed interim assessments ensure that students will be evaluated on the knowledge and skills valued by the district's educators and the local community (Rabinowitz & Ananda, 2001).

- District-developed interim assessments are more efficient than assessments created by individual teachers because they promote consistency in expectations and provide timely, accurate, and specific feedback to both teachers and students (Rutherford, 2007).

- Because students experiencing difficulty are all identified at the same time, educators are better able to create a systematic program of intervention (Olson, 2007; Rutherford County Schools, 2007).

- Studies indicate that frequent testing, along with the provision of corrective feedback, has a positive impact on students' performance (Shanahan et al., 2005; Black & Wiliam, 1998; Fuchs & Fuchs, 1986).

- Some studies suggest that administration of frequent formative interim assessments is particularly effective for low performing students (Black & Wiliam, 1998).

It should be noted, however, that some researchers have pointed to the drawbacks associated with interim assessments. For example, many claim there is already too much testing taking place in classrooms and that interim assessments just add to this burden. Others have stated that teachers tend to teach only the skills covered on interim assessments, leading to a progressive narrowing of the curriculum and a focus on lower-level skills. Finally, some researchers maintain that interim assessment results don't contain enough detail to allow teachers to effectively plan instructional remedies (Beauchamp, 2007; Dufour, 2007; Heritage, 2007; Peariso, 2006; Perie, Marion, & Gong, 2006; Sharkey & Murnane, 2006; Olson, 2005b).

Characteristics of Effective Interim Assessment Programs

A review of the research literature identified the following characteristics of effective interim assessment programs:

- Effective interim assessments are aligned with content standards and give students opportunities to apply their knowledge and skills in a variety of contexts and formats (Solution Tree, 2007; Vendlinski et al., 2007; Martin, 2006; Perie, Gong, & Marion, 2006; Perie, Marion, & Gong, 2006; Turner, 2003).

- Interim assessments monitor student progress so that early intervention is a routine part of the

learning process (Brown & Coughlin, 2007; Heritage, 2007; Wireless Generation, 2007; Herman et al., 2006; Herman & Baker, 2005).

- Assessment results are used to continuously improve the effectiveness of educational programs. Teachers are guided by test results when they select instructional interventions, decide which skills to reemphasize, and group students for instruction (National Literacy Trust, 2008; Stiggins & Chappuis, 2008; Delisio, 2007a; Olson, 2007; Stiggins, 2005b; Protheroe, 2001; Black & Wiliam, 1998; Consortium for School Networking, n.d.).

- High quality interim assessments are built on solid cognitive, developmental, and educational research and provide reliable and valid results. Items are free from cultural bias and provide fair opportunities for all students, including English language learners and students with disabilities (Solution Tree, 2007; Vendlinski et al., 2007; Wireless Generation, 2007; Perie, Gong, & Marion, 2006; Perie, Marion, & Gong, 2006; Herman & Baker, 2005).

- Some researchers have claimed that locally developed tests ensure the greatest match between what is valued at the district level and what is assessed. Locally developed assessment programs complement, rather than duplicate, statewide efforts and are responsive to local constituencies, including students, parents, teachers, administrators, and the community at large (Rabinowitz & Ananda, 2001).

- Researchers agree that interim assessments are only effective when student results are immediately available to educators (Delisio, 2007a; Olson, 2007; Perie, Marion, & Gong, 2006; Herman & Baker, 2005; Rabinowitz & Ananda, 2001). Shanahan and colleagues (2005) studied over 10,000 students in southern California elementary schools and found no significant gains in student achievement when benchmark exam results were delayed. Teachers were not aware of their students' progress in meeting standards so were not able to adjust their instruction accordingly.

- Effective interim assessment systems provide educators with reports that are easy to interpret and provide clear guidance on how to use the results. Score reports provide detailed diagnostic information at the student, classroom, school, and district levels to identify patterns and trends and evaluate the effectiveness of programs and policies (Olson, 2007; Depka, 2006; Perie, Gong, & Marion, 2006; Arter & Stiggins, 2005; Herman & Baker, 2005; Stiggins, 2005b; Protheroe, 2001; Rabinowitz & Ananda, 2001).

- A key component of effective interim assessment programs is sustained attention to teachers' professional development (Wiliam, 2007; Perie, Marion, & Gong, 2006). Teachers must know how to score and interpret test results and be able to translate results into instructional actions (Heritage, 2007; Herman et al., 2006; Arter & Stiggins, 2005; Herman & Baker, 2005; Rabinowitz & Ananda, 2001). In their study of 874 teachers in southern California elementary schools, Shanahan and colleagues (2005) found that schools that posted the greatest increases in student achievement were those that used professional development to analyze and discuss the results of benchmark exams.

Research on the Impact of Interim Assessments on Student Achievement

Before engaging in a discussion of the research, it is important to understand the distinction between summative and formative assessments. Summative assessments measure student achievement after learning has occurred. Formative assessments evaluate student learning during the course of instruction (Arter & Stiggins, 2005; Olson, 2005a; Stiggins, 2005b). Starkman (2006) stated that summative assessments answer the question "How did I do?" while formative assessments answer the question "How am I doing?" For assessments to serve a formative purpose, they must diagnose gaps in students' learning and provide corrective action to help students succeed. In addition, instructional adjustments must be directly tied to the results of the assessment (Perie, Gong, & Marion, 2006).

Some researchers have concluded that interim assessments have a powerful impact on student

achievement (Delisio, 2007b; Peariso, 2006; Rothman, 2006; Olson, 2005a). Others, however, have stated that there is little, if any evidence that interim assessments have a positive effect on students' academic performance (Perie, Gong, & Marion, 2006; Popham, 2006). The disagreement appears to stem from the fact that some researchers have drawn conclusions without specifying whether the assessments studied were formative or summative. This lack of specificity has occurred in part because formative assessment has not been consistently defined in the literature (Marzano, 2006). For example, Floden and Shepard (2007) noted that some interim assessments are called formative assessments, but do not provide real-time information about students' mastery of skills and knowledge. In fact, researchers have noted that many commercially developed interim assessments claim to be formative but are only mini-standardized tests intended to predict how well students will perform on end-of-year state tests (Fair Test, 2007; Floden & Shepard, 2007; Perie, Gong, & Marion, 2006; Popham, 2006).

The existing research base is limited but indicates that formative, not summative, assessments lead to higher levels of academic achievement. Studies also suggest that frequent testing, along with the provision of corrective feedback, has a positive effect on students' subsequent performance. Following is a brief summary of research conducted on the impact of formative interim assessments on student achievement.

- Black and Wiliam (1998) conducted a meta-analysis of 250 studies and found that when teachers used frequent formative assessments to adjust ongoing instruction, students demonstrated increased mastery of content and improved their performance on external achievement tests. Typical effect sizes were between 0.4 and 0.7, which the researchers claimed were larger than the effect sizes usually found for most other educational interventions. The administration of frequent formative assessments appeared to be particularly effective for low performing students.

- Fuchs and Fuchs' (1986) meta-analysis found that the use of formative assessments produced significant learning gains across all grade levels and several content areas. Teachers in these studies conducted formative assessments between two and five times a week. Effect sizes were even greater in classrooms that used formative assessment results to guide instruction, compared to control classrooms.

- Shanahan, Hyde, Mann, and Manrique (2005) conducted a study to determine if a three-pronged intervention of standards-based curriculum guides combined with quarterly benchmark assessments and teacher professional development would lead to increases in students' math achievement. Approximately 875 elementary teachers and over 10,000 students from a large urban school district in California participated in the study. The researchers found that implementation of the intervention coincided with a statistically significant 12 percent gain in student achievement. They concluded that benchmark tests provided teachers with timely indicators of students' progress so instructional adjustments could be made prior to the administration of state tests.

- Henderson, Petrosino, Guckenburg, and Hamilton (2008) studied whether the administration of quarterly benchmark exams would lead to greater gains in eighth grade students' math scores on the Massachusetts Comprehensive Assessment System (MCAS). The researchers compared 22 middle schools that received grants to implement benchmark assessments with 44 statistically matched control schools. The study found no significant difference between the MCAS scores at schools administering benchmark assessments and control schools. One suggestion made by the authors to explain their findings was that some control schools may have independently implemented their own versions of benchmark testing. This would have resulted in analyses that compared schools with different types of benchmark programs, instead of program and control schools.

- Bangert-Drowns, Kulik, and Kulik (1991) conducted a meta-analysis of 29 studies examining the relationship between frequency of assessment and student performance. They found significantly increasing effect sizes as the

number of assessments increased from zero to 10 over a 15-week period, after which effect sizes tended to level off. Administration of 10 assessments over a 15-week period resulted in a 22.5 percentile point gain in test scores. The authors suggested that the administration of more than 10 assessments over a 15-week period did not provide students with much benefit.

Fuchs and Fuch's (1986) earlier meta-analysis of 21 studies reported findings similar to those obtained by Bangert-Drowns and colleagues. Fuchs and Fuchs found that administration of two formative assessments per week resulted in a 30 percentile point gain in test scores. Educational researcher Bruce Tuckman told *Education World* (Delisio, 2007b) that testing students frequently with shorter exams is one of the best ways to help them commit information to long-term memory. Studying smaller, rather than larger, blocks of information appears to enhance students' recall of the material.

- In a meta-analytic study that examined the impact of feedback on students' test performance, Bangert-Drowns and associates (1991) concluded that when students received feedback on a classroom assessment that only told them whether their answers were correct or not, it had a negative impact on their future performance. However, test scores increased by up to 20 percentile points when students were also provided with correct answers; understood the criteria used to judge their responses; or were provided with explanations as to why their responses were correct or incorrect. Similarly, Fuchs and Fuchs' (1986) meta-analysis found that displaying assessment results graphically was associated with gains in student achievement. They suggested that graphic displays encouraged students to take control of their own learning, helped teachers more accurately judge students' levels of understanding, and provided teachers with a more precise frame of reference for making decisions about instructional interventions.

  The manner in which feedback is communicated to students also appears to have an impact on their performance. For example, in a review of studies on feedback and its effect on learning, Kluger and DeNisi (1996) found that discouraging feedback had a negative impact on student achievement. Positive learning outcomes were more likely when feedback focused on specific task features, such as how the student could improve in relation to standards, and emphasized learning goals instead of offering non-specific praise.

  There is also some evidence that students gain the most from assessments when feedback is provided without grades. When grades are assigned, they tend to be perceived as the real purpose of the assessment (National Literacy Trust, 2008; Fair Test, 2007; Crooks, 2001). The Princeton Review (n.d.) stated that good formative assessments have no negative consequences for students, teachers, or schools.

Research on the Predictive Validity of Interim Assessments

A review of the research literature found few published studies investigating the predictive validity of interim assessments. In general, results of these studies have produced mixed findings. Although some interim assessments have predicted performance in certain cases, there has been much variability in the magnitude of these relationships.

- Brown and Coughlin (2007) studied the ability of locally administered benchmark assessments to predict performance on state tests in four Mid-Atlantic states. They found that only one of four commercially developed reading and math benchmark assessment systems (CTB/McGraw-Hill's TerraNova) predicted proficiency levels on end-of-year state tests.

- Shanahan, Hyde, Mann, and Manrique (2005) studied the predictive validity of interim assessments in a large urban California school district. When the researchers compared benchmark assessment scores with scores on the California Standards Test (CST), the benchmark exams emerged as 90 percent accurate predictors of student CST performance.

- Peariso (2006) investigated the predictive validity of interim assessments on the California Standards Test (CST) at two California high schools. He found that interim assessments did not accurately predict students' success at the "proficient" and higher levels on the CST. The assessments did, however, predict performance at the lower CST proficiency levels. Performance-based interim assessments had a closer relation to CST scores than multiple-choice assessments.

- The Oregon University System, along with the Oregon State Department of Education and the Oregon Department of Community Colleges and Workforce Development (2003) studied the relationship between tenth grade students' benchmark assessment scores in reading, writing, and math and their subsequent performance during their first year of college. The researchers found that performance on benchmark assessments was closely aligned with students' freshman year college performance two years later. Students who met or exceeded the benchmark standards were more likely to earn higher grade point averages (GPAs) in related college courses. It should be noted, however, that high school GPA correlated with college GPA at a higher level than benchmark assessment scores.

On A Local Note

Miami-Dade County Public Schools' (M-DCPS) Interim Assessment (IA) program is designed to provide educators with information about students' academic achievement and instructional needs, identify learning deficiencies, and provide timely feedback to students and teachers. The goal of the district's IA program is to enhance instructional practices by using data to make curricular decisions. The program consists of three components: interim assessments, a benchmark assessment item bank, and professional development.

Interim Assessments

During the 2007-08 school year, interim assessments were administered to students in grades 3-10 in reading and math and to students in grades 4, 5, 7, 8, 10, and 11 in science. Almost 200,000 students participated in each administration of the IA. Most students were expected to participate in testing, with exemptions available only for special education students participating in the Florida Alternate Assessment (instead of the FCAT). Schools were required to administer the science test in November 2007, January 2008, and April 2008. Administration of the reading and math tests was required in the fall and winter, with only Assistance Plus Schools (those meeting federal and state criteria for low-achieving schools) required to administer the reading and math tests a third time in the spring.

Beginning in the 2008-09 school year, administration of the IA will be required only at Differentiated Accountability schools (those meeting federal and state criteria for low-achieving schools). The IA will be administered at grades 3-10 in reading and math and at grades 4, 5, 7, 8, 10, and 11 in science. Due to current budget constraints, the IA will be administered on a voluntary basis at all other schools. However, confirming the usefulness of the interim assessment process, 91 schools have used their own funds to pay for IA booklets in order to administer the test during the 2008-09 school year.

The IA was developed for M-DCPS by the Educational Testing Service. It is a formative assessment, aligned with the district's instructional pacing guides, FLDOE item specifications, and the Sunshine State Standards. IAs contain at least four items for each tested benchmark of the Sunshine State Standards. The tests underwent a rigorous review process by M-DCPS teachers and curriculum specialists to ensure that items complied with local requirements and followed the district's pacing guides. In addition, all items comply with FCAT passage and item specifications. IA tests contain multiple-choice items.

Assessments are administered by individual classroom teachers. Results are scanned into the Edusoft Management System by either the classroom teacher, a designated teacher, or test chairperson, depending on the arrangements made at each school site. After scanning, school staff are able to retrieve results immediately. Each school uses the Edusoft system to produce its own score reports. Reports are available for individual students, classrooms, schools, regional centers, the

district, and for specific subgroups, such as LEP and ESE students. Scores are provided as the percent of items answered correctly in all three subjects and as performance levels in reading and math (science performance levels will be available for the fall 2008 administration). Performance levels are designated as satisfactory progress (students who are likely to score at FCAT Levels 3 and above), limited progress (students whose performance is not sufficient to predict success), and insufficient progress (students who are likely to score below FCAT Level 3). Teachers reported using IA results to modify instruction and to place students in new or different instructional programs.

Average correlations between students' spring 2007 FCAT scores and their IA raw scores in fall 2007 and winter 2008 were moderately high (.82 and .80, respectively). Similarly, average correlations between students' predicted performance levels and subsequent FCAT scores were .86 in the fall and winter. Therefore, the IA appears to be measuring the same academic skills as the FCAT and should prove helpful in guiding instruction. A full predictive validity study is planned for later this year.

Benchmark Assessment Item Bank (BAIB)

The Benchmark Assessment Item Bank (BAIB), available through the district's Web site using ExamView test generator software, contains items developed by the Educational Testing Service. Items have been reviewed and approved by panels of M-DCPS administrators, curriculum specialists, and teachers. The items are aligned to the FCAT Test Item Specifications and measure a range of difficulty levels. The BAIB is used for remediation purposes, based on students' IA results. The bank enables teachers to select multiple-choice, gridded response, and short and extended response items to create customized "on demand" formative assessments; target improvement efforts; adjust instruction throughout the year; and develop personalized learning strategies for students struggling with a particular standard. The BAIB contains approximately 1,500 reading items across grades 3-10 and approximately 1,000 math items across grades 3-10. The science bank contains approximately 360 items across grades 4, 5, 7, 8, 10, and 11, with an additional 400 items to be added in September 2008.

Professional Development

Professional development was provided that focused on how to read, interpret, and analyze assessment results in order to target instruction. In 2006-07, school staff attended sessions on assessment literacy; interpreting results and reading reports; planning instructional interventions; targeting instruction for maximum effect; unwrapping benchmarks; and collaborative debriefing. Five staff members from each school attended two four-day professional development sessions. In 2007-08, data teams from each regional center received professional development on assessment literacy; unwrapping benchmarks; using data to guide instruction; and integration of the IA program into the curriculum.

Assessment, Research, and Data Analysis' Interim Assessment Web page (http://oada.dadeschools.net/IAP/IAP.asp) contains resources that include program materials; an overview of the item bank; instructions for retrieving answer keys in Edusoft; instructions for generating score reports; professional development resources; and support links, such as pacing guides, FCAT Test Item Specifications, language arts standards, math standards, Sunshine State Standards, and Florida standards.

For more information on the district's Interim Assessment program, contact Ms. Felicia Mallory, Executive Director, Student Assessment and Educational Testing, at (305) 995-7520.

Summary

Teachers have always used tests to measure students' mastery of knowledge and skills. However, educators have recently begun to incorporate formative interim assessments into the classroom learning process. These assessments allow them to adapt instructional programs to better meet students' academic needs throughout the course of the school year. This report reviewed reasons educators should consider using interim assessments as a teaching and learning tool and the characteristics of interim assessment programs that are essential to their effective implementation.

Research is limited but indicates that when assessments are administered periodically throughout the school year and their results are

used to guide instruction, they can have a positive impact on student achievement. In addition, studies suggest that frequent testing, combined with the provision of corrective feedback, leads to higher levels of performance. The administration of frequent formative assessments appears to be particularly effective for low performing students. Research investigating the predictive validity of interim assessments has produced mixed findings. Although some interim assessments have predicted students' subsequent performance, there has been much variability in the magnitude of these relationships.

M-DCPS' interim assessment program is designed to provide educators with information about students' academic achievement and instructional needs, identify learning deficiencies, and provide timely feedback to students and teachers. The goal of the program is to enhance the district's instructional practices by using data to make curricular decisions. The program consists of three components: interim assessments, a benchmark assessment item bank, and professional development. Preliminary indicators suggest a positive impact from the interim assessment program, with a more detailed study scheduled in the near future.

References

Arter, J., & Stiggins, R. (2005). Formative Assessment as Assessment for Learning. *National Council on Measurement Education Newsletter, 13*(3). Retrieved from http://www.ncme.org/pubs/pdf/news0905.pdf.

Bangert-Drowns, R.L., Kulik, J.A., & Kulik, C-L.C. (1991). Effects of Frequent Classroom Testing. *Journal of Educational Research, 85*(2), 89-99.

Beauchamp, A. (2007). Not All Benchmark Assessments are Created Equal. *The California Science Project Connection, 8*(2). Retrieved from http://csmp.ucop.edu/downloads/csp/newsletters/newsletter11_2007.pdf.

Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan, 80*(2). Retrieved from http://www.pdkintl.org/kappan/kbla9810.htm.

Brown, R.S., & Coughlin, E. (2007). The Predictive Validity of Selected Benchmark Assessments Used in the Mid-Atlantic Region. *Regional Educational Laboratory Issues and Answers Report No. 017.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

Business Wire. (2003). *Vantage Learning Delivers Benchmark and Ongoing Assessments for the School District of Philadelphia.* Retrieved from http://findarticles.com/p/articles/mi_m0EIN/is_2003_Dec_16/ai_111295134.

Consortium for School Networking. (n.d.). *Data-Driven Decision Making FAQ.* Retrieved from http://www.3d2know.org/FAQ.html.

Crooks, T. (2001). *The Validity of Formative Assessments.* Paper presented at the British Educational Research Association Annual Conference, University of Leeds, Leeds, England, September 2001.

Delisio, E. (2007a). Mapping Instruction with Interim Assessments. *Education World.* Retrieved from http://www.education-world.com/a_admim/admin/admin496.shtml.

Delisio, E. (2007b). More (Short) Tests Can Help Learning. *Education World.* Retrieved from http://www.educationworld.com/a_admin/admin/admin471.shtml.

Depka, E. (2006). *The Data Guidebook for Teachers and Leaders.* Thousand Oaks, CA: Corwin Press.

DuFour, R. (2007). *Common Formative Assessments.* Retrieved from http://www.allthingsplc.info/wordpress/?p=46.

FairTest. (2007). *The Value of Formative Assessment.* Retrieved from http://www.fairtest.org/value-formative-assessment-pdf.

Floden, R.E., & Shepard, L. (2007). *Formative Assessment as an Approach to Improving Education.* Retrieved from http://www.hewlett.org/NR/rdonlyres/CCC2B32A-8F4B-464C-836C-BB5E5372472A/0/FormAssessmentevaluation2007ExecSumm.htm.

Fuchs, L.S., & Fuchs, D. (1986). Effects of Systematic Formative Evaluation: A Meta-Analysis. *Exceptional Children, 53*(3), 199-208.

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2008). A Second Follow-Up Year for Measuring How Benchmark Assessments Affect Student Achievement. *Regional Educational Laboratory Technical Brief No. 002.* Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Heritage, M. (2007). Formative Assessment: What Do Teachers Need to Know and Do? *Phi Delta Kappan, 89*(2). Retrieved from http://www.pdkintl.org/kappan/k_v89/k0710her.htm.

Herman, J.L., & Baker, E.L. (2005). Making Benchmark Testing Work. *Assessment to Promote Learning, 63*(3), 48-54.

Herman, J.L., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The Nature and Impact of Teachers' Formative Assessment Practices.* National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Center for the Study of Evaluation Technical Report 703, University of California, Los Angeles, CA. Retrieved from http://www.cse.ucla.edu/products/reports/R703.pdf.

Jones, L., Gueorguieva, J., & Bishop, S. (2005). Some Perspectives on the Recent Expansion in the Use of Assessments for Formative Purposes. *National Council on Measurement Education Newsletter, 13*(3). Retrieved from http://www.ncme.org/pubs/pdf/news0905.pdf.

Kluger, A.N., & DeNisi, A. (1996). Effects of Feedback Intervention on Performance. *Pyschological Bulletin, 119*(2), 254-284.

Martin, R.A. (2006). Wake-Up Call Brings a Jolt of Alignment to the Curriculum. *National Staff Development Council, 27*(1). Retrieved from http://www.nsdc.org/members/jsd/martin271.pdf.

Marzano, R.J. (2006). *Classroom Assessment and Grading that Work.* Alexandria, VA: Association for Supervision and Curriculum Development.

National Literacy Trust. (2008). *Formative Assessment/Assessment For Learning.* Retrieved from http://www.literacytrust.org.uk/Database/assessment.html.

Olson, A. (2007). Growth Measures for Systemic Change. *School Administrator, 64*(1), 10-16.

Olson, L. (2005a). Benchmark Assessments Offer Regular Checkups on Student Achievement. *Education Week, 25*(13), 13-14.

Olson, L. (2005b). Not All Teachers Keen on Periodic Tests. *Education Week, 25*(13), 13.

Oregon University System, Oregon State Department of Education, & Oregon Department of Community Colleges and Workforce Development. (2003). *The First Year: Student Performance on 10th Grade Benchmark Standards and Subsequent Performance in the First Year of College (2001-02).* ERIC Document Reproduction Service No. ED481709.

Peariso, J. (2006). *A Better Benchmark Assessment: Multiple-Choice Versus Project-Based.* University of La Verne Masters Thesis. ERIC Document Reproduction Service No. ED499851.

Perie, M., Gong, B., & Marion, S. (2006). *A Framework for Considering Interim Assessments.* Paper developed for the Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS) Meeting, Austin, TX, October 2006.

Perie, M., Marion, S., & Gong, B. (2006). *Interim Assessments.* Presentation at the Formative Assessment for Students and Teachers (FAST) State Collaborative on Assessment and Student Standards (SCASS) Meeting, Austin, TX, October 2006.

Popham, W.J. (2006). *All About Accountability/Phony Formative Assessments: Buyer Beware!* Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved from http://www.ascd.org.

The Princeton Review. (n.d.). *Interim Assessment with Instructional Impact: How to Use the Formative, Low-Stakes Testing System to Support Teaching and Learning in New York City.* Retrieved from http://schools.nyc.gov/daa/InterimAssessments/ela-math/NYC%20Interim%20Assessment.pdf.

Protheroe, N. (2001). Improving Teaching and Learning with Data-Based Decisions: Asking the Right Questions and Acting on the Answers. *ERS Spectrum,* Summer 2001. Arlington, VA: Educational Research Service.

Rabinowitz, S., & Ananda, S. (2001). *Balancing Local Assessment with Statewide Testing: Building a Program That Meets Student Needs.* San Francisco, CA: WestEd. Retrieved from http://www.wested.org/online_pubs/kn-01-01.pdf.

Rothman, R. (2006). (In)formative Assessments: New Tests and Activities Can Help Teachers Guide Student Learning. *Harvard Education Letter,* November/December 2006. Retrieved from http://www.timeoutfromtesting.org/0519_article_harvard.php.

Rutherford County Schools. (2007). *Common Formative Assessments.* Rutherford County Schools, Murfreesboro, TN. Retrieved from http://www.rcs.k12.tn.us/rc/instruction/plc/common_formative_assessments.html.

Shanahan, T., Hyde, K., Mann, V., & Manrique, C. (2005). *Integrating Curriculum Guides, Quarterly Benchmark Assessments, and Professional Development to Improve Student Learning in Mathematics.* Paper presented at the Mathematics and Science Partnerships (MSP) Evaluation Summit: Evidence-Based Findings from the MSPs, Minneapolis, MN, September 2005.

Sharkey, N., & Murnane, R. (2006). Tough Choices in Designing a Formative Assessment System. *American Journal of Education, 112,* 572-588.

Solution Tree. (2007). *The Power of Common Assessments.* Retrieved from http://www.whiteriver.wednet.edu/SCHOOLS/PLC/powerpoint/CommonFormativeAssessment.ppt.

Starkman, N. (2006). Formative Assessment: Building A Better Student. *Technological Horizons in Education Journal, 33*(14). Retrieved from http://www.thejournal.com/articles/19174_5.

Stiggins, R. (2005a). *Assessment For Learning Defined.* Paper written for the ETS/Assessment Training Institute's International Conference: Promoting Sound Assessment in Every Classroom, Portland, OR, September, 2005. Retrieved from http://www.assessmentinst.com/AFLDefined.pdf.

Stiggins, R. (2005b). From Formative Assessment to Assessment For Learning: A Path to Success in Standards-Based Schools. *Phi Delta Kappan, 87*(4), 324-328.

Stiggins, R., & Chappuis, J. (2008). *Enhancing Student Learning.* Retrieved from http://www.districtadministration.com/viewarticlepf.aspx?articleid=1362.

Turner, J.R. (2003). Ensuring What Is Tested Is Taught: Curriculum Coherence and Alignment. *Informed Educator Series.* Arlington, VA: Educational Research Service.

Vendlinski, T.P., Nagashima, S., & Herman, J.L. (2007). *Creating Accurate Science Benchmark Assessments to Inform Instruction.* National Center for Research on Evaluation, Standards, and Student Testing (CRESST) Report 730, University of California, Los Angeles, CA. Retrieved from http://www.cse.ucla.edu/products/reports/R730.pdf.

Wiliam, D. (2007). *What Does Research Say the Benefits of Formative Assessment Are?* National Council of Teachers of Mathematics Assessment Research Brief. Retrieved from http://www.nctm.org/uploadedFiles/Research_Issues_and_News/Briefs_and_Clips/brief_form_assessment.pdf.

Wireless Generation. (2007). *The Role of Formative Assessment in Pre-K through Second Grade Classrooms.* Retrieved from http://www.wirelessgeneration.net/pdf/Formative Assessment White Paper Perk-2 8.7.07.pdf.

All reports distributed by Research Services can be accessed at http://drs.dadeschools.net under the "Current Publications" menu.