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).
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; 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).

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).

• 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).

• 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

Reasons to Use Interim Assessments

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

- 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
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 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
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
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.

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.
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


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.


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