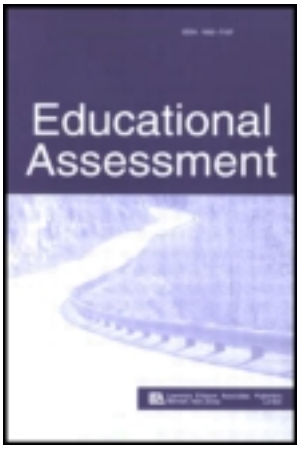

Educational Assessment
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/heda20

Locating Interim Assessments Within Teachers' Assessment Practice

Matthew Riggan [a] & Leslie Nabors Oláh [a]
[a] University of Pennsylvania
Published online: 07 Mar 2011.

To cite this article: Matthew Riggan & Leslie Nabors Olh (2011) Locating Interim Assessments Within Teachers' Assessment Practice, Educational Assessment, 16:1, 1-14, DOI: 10.1080/10627197.2011.551085

To link to this article: http://dx.doi.org/10.1080/10627197.2011.551085

Educational Assessment, 16:1–14, 2011

ISSN: 1062-7197 print/1532-6977 online
DOI: 10.1080/10627197.2011.551085

Routledge
Taylor & Francis Group

Locating Interim Assessments Within Teachers' Assessment Practice

Matthew Riggan and Leslie Nabors Oláh
University of Pennsylvania

Promising research on the teaching and learning impact of classroom-embedded formative assessment has spawned interest in a broader array of assessment tools and practices, including interim assessment. Although researchers have begun to explore the impact of interim assessments in the classroom, like other assessment tools and practices, they have been studied largely in isolation. Drawing on interview and classroom observation data for 32 teachers in two school districts, this article examines how teachers collect, interpret, and act on three types of assessment information: "short cycle" (Wiliam & Leahy, 2006) practices embedded within the flow of instruction, teacher-developed tools such as quizzes and homework assignments, and interim assessments. The article then describes the ways in which these types of assessment intersect within classroom practice and the degree to which teacher use of assessment tools or practices varies by type.

In the search for instructional practices that have a direct impact on student learning, few would seem to hold as much promise as formative assessment. "Short-cycle" formative assessment practices (Wiliam & Leahy, 2006)—largely those that are based on information collected by teachers within their classrooms—are potentially one of the most powerful means to improve the quality of teaching and raise student performance (Black & Wiliam, 1998; Crooks, 1988; Natriello, 1987).

The research base on formative assessment includes a wide range of instructional tools and practices. These include instructionally embedded practices such as choice of task, quality of discourse, questioning strategies, and quality of feedback (Black & Wiliam, 1998), teacher-constructed performance assessments (Shepard et al., 1996), student work (Young, 2006), externally designed curriculum-embedded formative assessments (Shavelson et al., 2008), and externally designed interim assessments (Christman et al., 2009; Henderson, Petrosino, Guckenburg, & Hamilton, 2008; Quint, Sepanik, & Smith, 2008).

Due in part to the breadth and variety of tools and practices to which the "formative assessment" label was applied, researchers and policymakers have more recently sought to

Correspondence should be sent to Matthew Riggan, Consortium for Policy Research in Education, University of Pennsylvania, 3440 Market Street, Philadelphia, PA 19104. E-mail: riggan@gse.upenn.edu

refine and specify the term's meaning. In 2006, a draft definition of formative assessment from the Council of Chief State School Officers (CCSSO) read, "An assessment is formative to the extent that information from the assessment is used, during the instructional segment in which the assessment occurred, to adjust instruction with the intent of better meeting the needs of the students assessed" (Popham, 2006, pp. 3–4). The final version of the definition went further in specifying the duration and use of formative assessment while explicitly defining it as a process. "Formative assessment is a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes" (CCSSO, 2008, p. 3). Although this definition is more clear and more closely tied to research showing instructional impact, it also creates a category of assessments that are intended to inform instruction yet occur outside of real-time teaching. These include assessment tools (such as homework assignments or class projects) created by teachers as well as externally developed interim assessments.

THE FORM AND USE OF INTERIM ASSESSMENTS

There is considerable skepticism about whether interim assessments, particularly those that are not embedded within the curriculum and those that are in all multiple-choice formats, can be considered formative. Shepard (2010) argued that using interim assessments formatively (i.e., translating interim assessment results into instructional modification) relies on "an idealized data-based, decision-making theory," the apparent successes of which are more often than not confounded with other supports for instruction, such as professional collaboration (p. 246). Black and Wiliam (2009) noted that although teachers may use interim assessment results to "make decisions about their next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence" (p. 29), such long-cycle assessments are actually the least likely to result in meaningful changes in instruction and deeper student understanding of content. Perie, Marion, and Gong (2009) noted that interim assessments share certain similar characteristics with formative assessments, namely, that these types of assessment are conducted more frequently and have a narrower content focus than do summative assessments. But they differentiate between interim and formative assessment due to differences in how the two types of assessment are used and the degree to which the information generated by them allows for comparisons across classrooms or schools.

The National Research Council report, *Knowing What Students Know: The Science and Design of Educational Assessment*, advised that as new forms of assessment are made available to teachers and schools, more research will be needed on the use of such assessments in the classroom. (Pellegrino, Chudowsky, & Glaser, 2001). Although we acknowledge that teachers may use information generated from interim assessments to alter instruction, this article follows Perie and colleagues in drawing a distinction between formative and interim assessment.

Whether or not they are considered formative, it is clear that different types of assessment are: (a) intended to inform instruction in some way; and (b) cohabitate within teachers' practice. Instruction itself is a multidimensional, complex practice. Attempts to measure instructional quality suggest that good teaching requires teachers to clearly communicate their expectations to students, design rich and challenging assignments, and encourage focused and productive dialogue in the classroom (Matsumura et al., 2006) as well as create classroom social environ-

ments that promote healthy social and emotional development (Pianta & Hamre, 2009). In the largest study ever conducted of instructional practice, researchers are using six different rubrics to measure instructional quality. Some focus on general practice, others specific content, and others classroom environment. Whatever its findings, the very design of this study highlights the multidimensionality of teaching practice (Educational Testing Service, RAND Corporation, & Institute for Social Research at the University of Michigan, 2010).

Like instruction, assessment practice is a mosaic of tools, routines, and practices. At the classroom level, teachers are asked to draw from a multitude of data sources to inform instruction. In addition to those just mentioned, teachers also regularly make use of state-test and English-language proficiency scores, attendance and medical records, diagnostic assessments and screening instruments, and Individualized Education Plans. Understanding the whole requires close examination of the ways in which the parts of assessment practice fit together. To date, little research has considered the ways in which different types of assessment interact within the context of individual teacher practice. In an action-research project, Torrance and Pryor (2001) found that they had to modify their initial conceptual framework to include a more holistic view of formative assessment, as the participating teacher-researchers "began to point out that, in practical situations, these processes were often embedded within another, or occurred in linked sequences or progressions" (p. 622). In their study of a small sample of middle-grades teachers implementing a specific Foundational Approaches to Science Teaching unit, Herman, Osmundson, Ayala, Schneider, and Timms (2006) examined the ways in which teachers employed both question routines and embedded written assessments to elicit students' conceptual understanding of science content, provide appropriate feedback, and respond instructionally. Overall, they found that even among a sample of highly engaged and experienced teachers with strong content knowledge, formative assessment practice across types or activities remained "basic." Teachers infrequently assessed for conceptual understanding, provided only cursory feedback to students, and did not consistently respond instructionally to feedback received. Although this study applied a holistic notion of formative assessment quality to multiple tools and practices, it did not explore the relation between teachers' use of question routines and written assessments. Indeed, Herman et al. noted that few studies have examined the ways in which teachers "orchestrate" the range of assessment tools and practices available to them, that is, the way they link, integrate, or sequence them within their instruction and planning. As increasing numbers and types of assessments are introduced into schools and classrooms, the knowledge about teachers' integration of assessments will become all the more valuable to researchers and practitioners alike.

This article addresses this gap in the research by describing the relationship between teachers' use of interim assessments and other aspects of their assessment practice in mathematics. In a previous analysis of teachers' use of interim assessments in elementary mathematics (Oláh, Lawrence, & Riggan, 2010), we found that teachers analyzed data in two ways. First, they used it to *locate* errors, focusing on whether students got items correct. Second, in response to researcher prompts, they used it to *diagnose* those errors, focusing on why students might have gotten certain items wrong. In doing so, some teachers looked at the steps or procedures used by students in solving problems, whereas others focused on underlying mathematical thinking or misconceptions. We also found that teachers' interpretation of interim assessment data was enhanced and complemented by information from other sources, some of which was obtained through other assessment practices. For example, some teachers reported asking students to

explain responses to particular interim assessment items, or encouraging students to show their work alongside their multiple-choice responses in their test booklets. Finally, we found that how teachers analyzed data, and what resources were available to them, led to different types of instructional planning. If assessment, formative or otherwise, is to ultimately improve instruction, the connection between analysis and instruction is crucial.

Taking a broader look at teachers' assessment practice, this analysis builds upon our previous findings. Our purpose is to further explore elementary school teachers' "orchestration" of assessment in mathematics—how they connect and sequence assessment tools and practices. Specifically, this analysis focuses on two sets of questions:

1. Do different types of formative assessment serve different purposes within teachers' instructional practice?
2. How do teachers connect different types of formative assessment? To what degree do different formative assessment types inform one another?

In this article, we first describe how teachers used different types of assessment: short-cycle, teacher developed, and interim. We then explore the ways in which use of different types of assessment intersected within teachers' practice.

ANALYZING ASSESSMENT

The conceptual framework from which this analysis derives locates assessment within a cycle of instructional improvement. Cycles of improvement have been applied to multiple aspects of district, school, and classroom practice. At the district level, standardized test score data have been used to make decisions about school and district performance goals, identify supports or sanctions for low-performing schools, evaluate school performance (Supovitz, 2006), and support efforts to align instruction with state standards (Hamilton et al., 2009). At the school level, state assessment data have been used to determine professional development needs, evaluate teachers, and identify students in need of intensive support (Kerr, Marsh, Ikemoto, Darilek, & Barney, 2006). Interim assessment data have been used for many of the same purposes, as well as to promote differentiated instruction at the classroom level and monitor teacher and school progress toward performance goals (Bulkley, Christman, Goertz, & Lawrence, 2008; Clune & White, 2008). To facilitate this kind of use, interim assessment results in mathematics are often reported by domain or subcontent area (e.g., operations, measurement, etc.).

Improvement cycles at the classroom level tend to utilize different types of evidence than district- or school-level cycles and be directed toward different ends (Coburn & Talbert, 2006). With respect to mathematics instruction, they tend to focus on the ways in which teachers surface student understandings and respond to errors or misconceptions in real time. This process has been referred to as "eliciting, interpreting, and acting" (Bell & Cowie, 2001) or as "eliciting, recognizing, and using information" (Ruiz-Primo & Furtak, 2004). Black and Wiliam (2006, p. 88) further specified this cycle as having six sequential stages:

1. A design, or intent, with formative opportunities built in.
2. Evoking of student responses.

3. Reception and interpretation of student responses.
4. Action based on interpretation of student responses.
5. Reception and interpretation of this action by the student.
6. Transition to the next part of the design.

Although models of the cycle of instructional improvement vary in terms of the system level at which they occur and the number of steps and processes included, they all contain three fundamental elements: the deliberate collection of information, interpretation of the information collected, and action based upon that interpretation. The framework for the study described here therefore focuses on these three, overarching steps.

METHOD

Data presented in this article were collected as part of a larger study of interim assessment use in nine schools in two school districts: Philadelphia, Pennsylvania and Cumberland, Pennsylvania.[1] Using data from interviews with teachers, principals, and district staff along with classroom observations and document analysis, this larger study considered not only how teachers used interim assessments but the overall policy supports that facilitated or impeded that use.

Study Sites

Philadelphia is among the largest school districts in the United States. It educates approximately 163,000 students in 284 schools, charter schools excluded. The student population is predominately African American (61.2%), with nearly equal proportions of Caucasian (13.3%) and Hispanic (17.6%) students. Cumberland is an economically diverse suburban district that enrolls approximately 7,400 students in seven elementary schools, one junior, and one senior high school. Most Cumberland students are Caucasian (70.7%), and 21% are African American.

These districts were selected based on a number of factors. First, both districts used the same mathematics program, *Everyday Mathematics* (*EM*), facilitating comparisons across districts. Second, both districts were in the same state and thus functioned within similar accountability contexts. Third, both districts had already adopted interim assessment systems in elementary mathematics prior to our study.

Within districts, schools were selected according to three criteria. First, all schools had made adequate yearly progress in school year 2004–05. Second, among schools that met this minimum level of achievement, we chose schools to reflect a range of mathematics performance around the district average. The average mathematics proficiency level in the participating Philadelphia schools ranged from 41% to 62% (compared to the Philadelphia 2004–05 elementary school average of 49%), whereas the average proficiency level in the participating Cumberland schools ranged from 80% to 93% (compared to the Cumberland

[1]Cumberland is a pseudonym for the suburban district in our study. Although we had permission from the Philadelphia to use the name of this district, the small number of schools in the suburban district made it impossible to use their name while maintaining the confidentiality of the schools, administrators, and teachers.

2004–05 elementary school average of 89%). Third, schools were chosen to reflect the ethnic and socioeconomic diversity within each district.

Philadelphia and Cumberland administered interim assessments roughly every 6 weeks in the elementary grades. These assessments were aligned to district curriculum and were designed to test only those concepts and objectives taught during that period. Teachers were given a period after receipt of assessment results to review and/or extend development of these topics. In Philadelphia this took the form of a scheduled "re-teaching week," whereas teachers in Cumberland had more discretion about how much instructional time should follow the administration of interim assessments.

The larger study focused on Grades 3 and 5, as these were the only elementary grades tested by the state at the start of our study. These are also focal grades for elementary mathematics instruction in that it is at these levels that the mathematical performance landmarks in computation are critical for students' academic progress. Third grade typically marks the level at which students are expected to show mastery of core addition and subtraction concepts and procedures with whole numbers and of fundamental knowledge of place value in different contexts (National Council of Teachers of Mathematics, 2006). Fifth grade is the point in the curriculum when students are expected to have mastered generalizable procedures for multiplication and division and to have developed fraction concepts and skills, including adding and subtracting fractions (National Council of Teachers of Mathematics, 2006). Fractions are crucial as foundations for continued work with rational numbers as well as algebra. With a few exceptions related to teacher attendance or schedule conflicts, we interviewed and observed all third- and fifth-grade teachers in our study schools. A total of 46 teacher interviews and observations were conducted in fall 2006, 39 in winter 2007, and 38 in spring 2007.

Data Collection

The analysis presented in this article is based on interview and classroom observation data collected in both districts throughout the 2006–07 school year. Three observations were conducted during the 2006–07 school year (fall, winter, and spring). We visited each third- and fifth-grade teacher's classroom for one mathematics period during each observation. Each observation focused on different aspects of teachers' assessment practice. The first examined teachers' short-cycle formative assessment practices, in particular how teachers addressed student misconceptions. The second and third observations occurred during the "instructional window" between the reporting/scoring of the interim assessment results and the end of that assessment period. This allowed us to make comparisons across classrooms, because all teachers within each district held essentially the same broad instructional goal during our visits (e.g., revisit content from January and February, revisit content from *EM* Unit 9, etc.). During these visits, we focused on instructional and formative assessment practices that teachers used during the reteaching period. Because we could not directly observe whether these instances of practice were linked to the information gained from the interim assessments, we asked teachers about these particular practices in the teacher interviews, which immediately followed each classroom observation.

Individual, hour-long interviews with teachers were conducted immediately following each classroom observation whenever possible. The goal of the interviews was twofold. First, we sought to learn more about what teachers were thinking during the observed lessons.

Assessment practice entails thinking and doing, often in rapid succession, so pairing interviews with observation provided the best means to capture both elements. Second, we wanted to explore the ways in which teachers analyzed, interpreted, and acted on assessment information. The interviews contained three components: background and context questions, data analysis activities, and misconception scenarios. Background and context questions (fall, winter, spring) focused on teachers' professional background, general assessment practices, and planning for reteaching.

Data analysis activities (fall, spring) allowed us to observe the ways in which teachers analyzed and interpreted interim assessment information. The first activity (fall) consisted of a hypothetical mock-up of student results based on each district's (and each grade's) interim assessment, presented in a format identical to that which they would receive for their own students. The use of hypothetical results allowed us to standardize the "results" across grades and districts to see what variation in teacher analysis or interpretation would occur in response to an identical set of results. We presented each teacher with a one-page printout of hypothetical interim assessment results, asking the teacher to imagine that this was her class and to "think aloud" for us about what she saw in the results. For the second data analysis activity (spring), we asked teachers to bring copies of their most recent interim assessment results with them. We asked both about classwide patterns of performance as well as about mathematical concepts that seemed to present difficulty for students. These questions were designed to closely mirror the first scenario. In this way, we hoped to get a more complete picture of teachers' individual processes of making sense of interim assessment results.

Misconception scenarios (winter, spring) examined how teachers responded to typical student misconceptions. Specifically, we wanted to learn: (a) whether teachers could identify student errors in mathematics and what these errors told them about students' thinking; (b) what questions they would ask students to learn the extent to which their own interpretations were correct; and (c) what instructional steps they would take to address particular misconceptions. These aims map onto the interpretation and planning steps of the instructional improvement cycle previously described.

Data Analysis

Analyzing and triangulating data across multiple sources followed a two-step process. First, interview data were coded using a structure that was aligned with the study's conceptual framework. All interviews were professionally transcribed and analyzed using Atlas.Ti qualitative data analysis software. The study's conceptual framework was used to develop an extensive set of codes, which were used to sort interview data into descriptive categories. This code set had five primary domains, three of which are pertinent to the present analysis: data collection, data analysis, and action. Additional codes were added to mark specific segments of interviews, such as data or misconception scenarios.

Second, using a combination of coded interview data and observation write-ups, profiles of assessment practice were constructed for each teacher in the study. Teacher profiles were constructed for all teachers for which multiple interviews and observations were available ($n = 39$). The profiles consolidated data from interviews and classroom observations, focusing on how individual teachers collected, interpreted, and acted on assessment information in three domains:

1. Interim assessments: Interim scored assessments designed to measure the progress of an entire class over an extended period.
2. Short-cycle assessments: *Practices* employed by teachers within a single class period to determine the extent to which students have acquired a specific concept or skill.
3. Teacher-developed assessments: *Tools* developed or adapted by teachers to gauge student understanding. Although some teacher-developed assessments may also be short-cycle assessments, others may extend across multiple class periods.

Teacher profiles were constructed using a two-step process. First, the database was filtered by teacher and data were retrieved for codes related to all aspects of interim assessment use, short-cycle assessment, teacher-administered tests and quizzes, interpretive or diagnostic processes, action based on analysis of assessment data (all types), and response to student misconceptions. Second, classroom observation notes were reviewed to identify tendencies, routines, or practices that were consistent across multiple observations, and to cross-reference those with the categorized and reduced interview data. The analysis presented in this article is built upon interview and observation data captured in profiles of 32 teachers (14 from Cumberland and 18 from Philadelphia).[2]

The analysis of teacher profiles focused on the specific ways in which teachers orchestrated assessment activities across types. For example, if a teacher asked specific questions of students based on his analysis of their performance on an interim assessment, the profile summary would note that he connected interim assessment *interpretation* with short-cycle *collection*. Similarly, if a teacher designed a performance task or test based on her interpretation of student work completed in groups the previous day, the profile summary would note that she had linked short-cycle *interpretation* with teacher-developed *collection*. In addition to noting all observed sequences for each teacher, the profile summaries contained notes on how such sequences were observed.

Limitations of this Analysis

Data collected for this study focused primarily on teachers' use of *interim* assessments. Although teacher interviews and observations yielded a considerable amount of data about short-cycle and teacher-developed assessments, data collection protocols focused less on these practices. In addition, although steps were taken to standardize both observation protocols and teacher profile formats, it must be acknowledged that variation always exists across both classes and researchers. Invariably, this results in some degree of bias at each level of analysis. Although the frequency of observations noted in this analysis is intended to serve as an indicator of the prevalence of certain relationships or practices, the reader is cautioned not to interpret these frequencies too rigidly but to focus instead on the broader relationships those frequencies suggest.

FINDINGS

Across the sample, there was widespread evidence of teachers completing the instructional improvement cycle—collecting information, interpreting it, and acting on it instructionally—

[2]The remaining seven profiles were excluded because they did not contain sufficient data to analyze across assessment types.

TABLE 1
Completion of the Formative Assessment Cycle, by Type

Formative Assessment Type	*No. of Philadelphia Teachers (out of 18)*	*No. of Cumberland Teachers (out of 14)*	*Total (out of 32)*
Interim	18	14	32
Short cycle	14	10	24
Teacher developed	13	10 (2 N/A)	23

for all three assessment types. Table 1 shows the number of teachers for whom there was evidence of completing the cycle for each assessment type.

All 32 teachers in the sample linked their analysis of interim assessment data to instruction in some way. Although it was somewhat less common for teachers to act on short-cycle information, there was evidence of such actions for three fourths of the teachers in the sample. Similarly, teacher-developed assessment feedback was linked to instruction in more than three fourths of the cases. In the remaining cases, there was generally evidence of collection and interpretation of information but no evidence of action based on that interpretation. Again, we should caution that as the focus of this study was teachers' use of interim assessment data, we may have underestimated the proportion of teachers who linked short-cycle or teacher-developed assessment to instruction.

How Teachers Employ Different Types of Assessment

This section focuses on the ways in which teachers used different types of assessment and the degree to which different teachers linked or sequenced assessment activities. The most common uses of interim assessment data by teachers in the sample were organizational: determining what to teach and to whom. In general, teachers identified areas of weakness (either content or individual students) and planned accordingly. When reviewing interim assessment results, about two thirds of teachers began by looking for weak content areas, either by looking directly at the standard to which an item corresponded or by identifying individual items that covered the same curricular content. Roughly one fifth of teachers began by looking for individual low-performing students rather than at content areas. Overall, nearly all teachers reviewed interim assessment results by both student and content area.

With regard to short-cycle practices, it should first be noted that not all questioning routines were considered to be formative assessment practice. Many such routines, like other teacher–student interactions observed, focused primarily on directing students toward a correct answer rather than eliciting information about their thinking or process. These interactions were not included in the present analysis, as their intent was not to elicit information but rather to direct student responses.

Although variation across the sample was noted, teachers used short-cycle and teacher-developed assessments in related but different ways. Short-cycle practices were most often used to get students to explain their thinking, or to allow teachers to observe their problem-solving processes. This action was most often characterized by open-ended questions, such as "How did you get that answer?" These questions were asked in response either to student vocalizations or to problems that students had solved independently. Eliciting this feedback

was at times a corrective in itself. In describing their process aloud, students would discover their own errors.

Like short-cycle practices, teacher-developed assessments were used to elicit additional information about students' problem-solving processes but were also used as postassessments to determine the degree to which students had mastered specific content. This information informed pacing decisions (e.g., whether the class could move on to a new unit) and in some cases assisted teachers in planning from one day's lesson to the next. A third-grade teacher from Philadelphia explained this function:

> Q: Thinking back to previous re-teaching, after you've spent five days presenting the material again to the children, the concepts, do you test for mastery in any way, either formally or informally?
>
> A: I would say both. Like quizzes, like maybe three problems on a particular skill. And any quizzes. Not unit tests in *Everyday Math*. No. Teacher-made tests.

In this way, teacher-developed assessments provided a valuable service to teachers as, as mentioned in our earlier description of these districts' assessment schedules, proceeding to the next instructional unit was determined by district-created pacing guides. Teacher-developed assessments gave teachers a more informed and principled way of pacing instruction.

Teacher Orchestration of Formative Assessment Types and Activities

There was evidence of orchestration of assessment activities across type for almost every teacher in the sample. Nearly all of these linkages involved the teacher moving from interpretation or action on a first type of assessment information to collection of a second type. The most commonly observed patterns showed teachers moving from interpretation of interim assessment data to collection of short-cycle information (again, this observed trend may be due to this study's focus on interim assessments). This sequence was observed for 18 of 32 teachers in the sample and served several purposes. Most often, short-cycle practices were used to elicit more information about why students answered interim assessment items in the way that they did. This pattern was observed for 12 of the 18 teachers. Two teachers specifically noted they did so to distinguish between students who genuinely did not understand the content and those who merely made "careless mistakes." Two other teachers suggested that they used interim assessment data to figure out which questions to ask students in upcoming classes. An observation of a fifth-grade Cumberland teacher illustrated this process:

> Teacher: "Yesterday, we took the practice test. After looking at the practice tests, I noticed that the things that many of you didn't get right we hadn't gone over. . . . Many of the volume questions, you didn't get right."
>
> . . .
>
> Teacher draws a 3-D rectangular figure on the overhead and asks the students how many "faces" the shape has. She calls on two different students; one says "five faces" and the second student says "six faces." Teacher then asks class, "How many say five faces?" and counts hands in the air. "How many say six faces?" and she counts hands. There are more hands for six faces than for five. "Why are there six faces?" teacher asks. "Why not five?"

Slightly less common (13 of 32 teachers) was the sequencing of interpretation of interim assessment data with collection of information from teacher-developed assessments. Ten teachers reported that, like short-cycle practices, these assessments were used to gather more information about student problem-solving processes. As noted earlier and discussed next, teacher-developed assessments were used to gauge student progress or mastery of retaught content (postassessment) or to make pacing decisions.

There was limited evidence that sequencing of assessment practices varied by district. For instance, teacher-developed assessments were employed more often in Philadelphia to assess the extent to which students had mastered content that was retaught following the administration of interim assessments. This likely resulted from the timing of the assessment cycle itself: Administration of the interim assessment was followed by a scheduled reteaching week, after which teachers were expected to move on to new content without any additional district-mandated or curriculum-embedded assessment, such as end-of-unit tests. In the absence of such assessments, teachers substituted their own to determine whether students had mastered the content they had retaught. In this way, teacher-developed assessments served a similar purpose to that of the short-cycle practices previously mentioned. In Cumberland, interim assessments were closely aligned with the timing of end-of-unit tests, which served a similar function to the use of teacher-developed assessments in Philadelphia.

There was also some evidence that Cumberland teachers used short-cycle (and in some cases teacher-developed) assessment to determine the timing of interim assessment administration. This was likely a function of the flexibility teachers had in timing these assessments and of the fact that a summative (end-of-unit) assessment normally followed closely after.

Several other linkages across formative assessment type were less pervasive but noteworthy. In six instances, teacher-developed assessments were employed to assess the impact of actions taken as a result of short-cycle assessment. This pattern was more common in Philadelphia, where it was evident in the practice of five of 18 teachers. Also in Philadelphia, three teachers reported using interim assessment findings to confirm their interpretations of feedback from short-cycle or teacher-developed assessments.

The analysis presented here suggests that, overall, teachers use different types of assessment for different (though sometimes overlapping) purposes. In some cases, they scaffold different assessment types in accordance with these purposes. The most common example was the following of interim-assessment analysis with the collection of additional short-cycle feedback focused on students' problem solving process.

DISCUSSION AND IMPLICATIONS

Taken together, these findings present a complex view of the relationship between interim, short-cycle, and teacher-developed assessments within teachers' practice. Teachers use different types of assessment for different purposes. Interim assessments are most often used to identify weak content areas or students within a class, whereas short-cycle practices are most often used to gather additional information about how students solved problems. Teacher-developed assessments played a similar role but also had a postassessment function, sometimes informing teachers' pacing decisions.

It appears that interim assessments structure and guide other types of assessment. In themselves, interim assessments appear limited in their capacity to inform teachers about students' thinking or problem solving (Goertz, Oláh, & Riggan, 2010), but among these teachers they gave direction to short-cycle assessments that may be better suited to that purpose. Other research suggests that short-cycle assessment practices may be better suited to conceptual diagnoses, as it provided teachers with more information about students' reasoning and problem-solving processes (Black & Wiliam, 1998). This suggests that although there is little evidence that directly associates interim assessments with improved student learning, such assessments may play a role within a broader *system* of assessment. Such systems are currently the focus of several development efforts (Herman et al., 2006; Shavelson et al., 2008). An important question for future research is whether the type of structure or guidance provided by interim assessments is sufficiently useful to teachers to justify the time and expense such assessments require. It is possible, after all, that teachers are perfectly capable of identifying students who are struggling or weak content areas *without* using interim assessments. This may be especially true for content areas such as early literacy, which already benefit from well-established diagnostic assessments.

This analysis suggests that future research should focus to a greater extent on how different types of assessment—both tools and processes—interact with and support one another within the context of teachers' practice. Specific attention should be given to what combinations of assessment use are most likely to help teachers to understand students' thinking and the types of professional development and support needed to help them do so.

ACKNOWLEDGMENTS

The writing of this article was supported by a grant from the William and Flora Hewlett Foundation to the Consortium for Policy Research in Education (CPRE). The research mentioned in this article was funded by the National Science Foundation (REC 0529485). Opinions expressed in this article are those of the authors and do not necessarily reflect the views of the supporting foundations, the study districts, CPRE, or its institutional members

REFERENCES

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Norwell, MA: Kluwer.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–75.

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). London, UK: Sage.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability, 21*(1), 5–31.

Bulkley, K. E., Christman, J. C., Goertz, M. E., & Lawrence, N. R. (2010). Building with benchmarks: The role of the district in Philadelphia's benchmark assessment system. *Peabody Journal of Education, 85*, 186–204.

Christman, J., Neild, R., Bulkley, K., Blanc, S., Liu, R., Mitchell, C., & Travers, E. (2009). *Making the most of interim assessment data. Lessons from Philadelphia*. Philadelphia, PA: Research for Action.

Clune, W. H., & White, P. A. (2008). *Policy effectiveness of interim assessments in Providence public schools* (WCER Working Paper No. 2008–10). Madison: University of Wisconsin–Madison, Wisconsin Center for Education Research.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of evidence-based practice in school districts: Mapping the terrain. *American Journal of Education, 112*, 469–495.

Council of Chief State School Officers. (2008). *Attributes of effective formative assessment*. Washington, DC: CCSSO FAST-SCASS.

Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438–481.

Educational Testing Service, RAND Corporation, & Institute for Social Research at the University of Michigan. (2010). *Understanding teacher quality*. Retrieved from http://www.utqstudy.org/instruments.html

Goertz, M. E., Oláh, L. N., & Riggan, M. (2010). *From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction*. Philadelphia, PA: Consortium for Policy Research in Education.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Henderson, S., Petrosino, A., Guckenburg, S., & Hamilton, S. (2008). *A second follow-up year for "Measuring how benchmark assessments affect student achievement"* (REL Technical Brief. REL 2008-No. 002). Newton, MA: Regional Educational Laboratory Northeast & Islands.

Herman, J., Osmundson, E., Ayala, C., Schneider, S., & Timms, M. (2006). *The nature and impact of teachers' formative assessment practices* (CSE Technical Rep. No. 703). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, *112*, 496–520.

Matsumura, L., Slater, S., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment* (CSE Technical Rep. No. 681). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics: A quest for coherence*. Reston, VA: Author.

Natriello, G. (1987). The impact of evaluation processes on students. *Educational Psychologist*, *22*, 155–175.

Oláh, L. N., Lawrence, N., & Riggan, M. (2010) Learning to learn from benchmark assessment data: How teachers analyze results. *Peabody Journal of Education, 85*, 226–245.

Pelligrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28*, 5–13.

Pianta, R., & Hamre, B. (2009). Classroom processes and positive youth development: Conceptualizing, measuring, and improving the capacity of interactions between teachers and students. *New Directions for Youth Development, 121*, 33–46.

Popham, W. J. (2006, June). *Defining and enhancing formative assessment*. Paper presented at the Annual Large-Scale Assessment Conference, Council of Chief State School Officers, San Francisco, CA.

Quint, J., Sepanik, S., & Smith, J. K. (2008, December). *Using student data to improve teaching and learning: Findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST–R) program in Boston elementary schools*. New York, NY: MDRC.

Ruiz-Primo, M. A., & Furtak, E. M. (2004). *Informal formative assessment of students' understanding of scientific inquiry* (CSE Rep. No. 639). Los Angeles, CA: CRESST, University of California, Los Angeles.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., … Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education*, *21*, 295–314.

Shepard, L. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education, 85*, 246–257.

Shepard, L., Flexer, R. J., Weston, T. J., Marion, S. F., Mayfield, V., & Hiebert, E. H. (1996). Effects of introducing classroom performance assessments on student learning. *Educational Measurement: Issues and Practice*, *15*(3), 7–18.

Supovitz, J. A. (2006). *The case for district-based reform: Leading, building, and sustaining school improvement*. Cambridge, MA: Harvard Education Press.

Torrance, H., & Pryor, J. (2001). Developing formative assessment in the classroom: Using action research to explore and modify theory. *British Educational Research Journal, 27*, 615–631.

Wiliam, D., & Leahy, S. (2006, April). *A theoretical foundation for formative assessment*. Paper presented at the American Educational Research Association annual meeting, San Francisco, CA.

Young, V. M. (2006). Teachers' use of data: Loose coupling, agenda setting, and team norms. *American Journal of Education, 1112*, 521–548.