Can Interim Assessments be used for Instructional Change?

Margaret E. Goertz, Leslie Nabors Oláh, Matthew Riggan

The past ten years have witnessed an explosion in the use of interim assessments by school districts across the country. A primary reason for this rapid growth is the assumption that interim assessments can inform and improve instructional practice and thereby contribute to increased student achievement. Testing companies, states, and districts have become invested in selling or creating interim assessments and data management systems designed to help teachers, principals, and district leaders make sense of student data, identify areas of strengths and weaknesses, identify instructional strategies for targeted students, and much more. Districts are keeping their interim tests even under pressure to cut budgets (Sawchuk, 2009). The U.S. Department of Education is using its Race to the Top program to encourage school districts to develop formative or interim assessments as part of comprehensive state assessment systems.

Much of the rhetoric around interim assessments paints a rosy picture. Supporters argue that these tests will provide data on student understanding; teachers' analysis of this data will in turn lead to greater differentiation of instruction and better teaching of content, leading to improved student learning. Much of the belief in the potential of interim assessments comes from the body of research on formative assessment, particularly those studies showing that "short-cycle" formative assessments—largely those that are based on information collected by teachers in their classrooms and that provide feedback to teachers within a single class period—are a powerful means to improve the quality of teaching and raise student performance (Black & Wiliam, 1998; Crooks, 1988; Hattie & Timperley, 2007; Natriello, 1987).

However, this research has not focused on interim assessments, but rather on practices that are embedded within classroom instruction.

Very little research exists on how interim assessments are actually used, by individual teachers in classrooms, by principals, and by districts. Some recent studies surveyed teachers about their use of test data in instruction. Many of these teachers reported that interim test results helped them monitor student progress and identify skill gaps for their students, and led them to modify curriculum and instruction (c.f., Christman, et al., 2009; Clune & White, 2008; Stecher, et al., 2008) These studies, however, did not examine how individual teachers actually analyzed and used these data to inform their classroom practice, the policy conditions that supported teachers' ability to use interim assessment data to improve instruction, or the interaction of interim assessments with other classroom assessment practices. Our study begins to fill that vacuum.

The purpose of this exploratory study was to examine the use of interim assessments and the policy supports that promote their use to change instruction, focusing on elementary school mathematics. We use the term "interim assessments" to refer to assessments that: a) evaluate student knowledge and skills, typically within a limited time frame; and b) the results of which can be easily aggregated and analyzed across classrooms, schools, or even districts (Perie, Marion, & Gong, 2009). Our study looked at how 45 elementary school teachers in a purposive sample of 9 schools in 2 districts used interim assessments in mathematics in 2006-07. The study focused on teachers' use of data in a cycle of instruction-

al improvement; that is, how teachers gather or access evidence about student learning; analyze and interpret that evidence; use evidence to plan instruction; and carry out improved instruction. It also considered the many factors that influence how teachers access, manage, interpret, and act on data. These include district and school policies and practices, and organizational norms and routines, as well as educator capacity.

The two study districts—Philadelphia and Cumberland, Pennsylvania[1]—share a common accountability context (i.e., the same state standards and state test), use the same elementary mathematics program, *Everyday Mathematics* (*EM*), and had adopted interim assessments in elementary mathematics. By selecting one urban and one suburban district, we sought to learn how policy supports for assessment and instructional improvement function in these different environments. Within each district, we chose schools that had made Adequate Yearly Progress but reflected a range of student performance as well as the ethnic and socio-economic diversity of the district. Seven of the nine study schools were Title I schools. In each site, we conducted interviews with district administrators, principals and instructional support staff. In each school, we observed and interviewed all 3rd- and 5th-grade teachers, our focal grades. We collected information on how the teachers analyzed and acted on their interim assessment data and how they would respond to student errors on assessment items. We also conducted a short survey designed to measure participating teachers' mathematical knowledge for teaching (Hill, Shilling, & Ball, 2004).

This policy brief reviews the study's key findings regarding the policy supports that existed to support data use and teachers' actual use of interim (and related formative) assessment data. It also presents implications for educators, policymakers, and researchers. Our findings highlight the potential and limitations of interim assessments for the four stages of the instructional improvement cycle.

We conclude that interim assessments that are designed for instructional purposes are helpful but not sufficient to inform instructional change. When well-supported by their districts and schools, teachers used interim assessment data to decide what to re-teach to and to whom, but not necessarily to change the ways in which they taught this content. Rather, teachers' instructional and assessment practices appeared to be affected more by their capacity to understand their students' mathematical learning than by the type of assessment (interim or formative) they used.

District and School Supports for Assessment Use

Philadelphia and Cumberland administered interim assessments roughly every six 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 time period. As discussed below, teachers were given a period of time after receipt of assessment results to review and/or extend development of these topics. Both districts adopted policies and created conditions that were designed to support teacher use of the interim assessment data for instructional improvement. These included setting strong expectations for data use; generating timely and accessible analyses of student performance data that could inform instruction; dedicating time to analyze data, plan instruction and re-teach students; and providing instructional support for teachers and students.

Expectations for use. Perie, Marion, and Gong (2009) discuss three possible uses of interim assessment results: a) instructional—to help teachers adjust their instruction and curriculum to address student learning needs; b) evaluative—to help educators evaluate and improve instructional programs; and c) predictive—to determine each student's likelihood of achieving a performance level on an end-of-year test. The districts in our study established and communicated expectations for the *instructional* use of interim assessment data at all levels of the system. The districts viewed these assessments as "teaching tools" that would support and guide teachers' instruction. District staff and school leaders (principals in Philadelphia and curriculum

[1] Cumberland is a pseudonym for the suburban district in the study.

2

specialists in Cumberland) expected teachers to use assessment results to reflect on their instruction, to discuss and share common problems and instructional solutions, and to provide remediation and enrichment during a dedicated period of time following the assessments.

Both districts designed their interim assessments to be part of their overall instructional guidance systems, not as "mini state tests" that mirrored the items in the high-stakes state assessment. The districts adopted curriculum in mathematics aligned to state standards, adopted common mathematics programs across schools, developed instructional timelines linked to units in these programs, and aligned interim assessment items with content of the district curriculum and materials for each instructional period.

The districts communicated their expectations to principals and teachers through several mechanisms. First, they structured their information management systems in a way that focused the attention of teachers, instructional support staff and, in Philadelphia, their principals on tested skills and learning standards. Philadelphia shaped and reinforced their expectations for analysis of the interim assessment results by mandating the use of a data analysis protocol that asked teachers to identify the weakest skills and concepts for the instructional period and instructional strategies for re-teaching these skills and concepts. Cumberland developed interim assessments at the request of teachers, and teachers participated in their development, creating more of a "buy-in" for their use. Although test results were not made public in Cumberland, the district required all teachers to enter assessment results into an electronic spreadsheet. The format of the spreadsheet, which highlighted student performance by content sub-areas, generated expectations of when and where teachers should provide additional support to students.

Second, Philadelphia held principals accountable for ensuring that teachers accessed, interpreted, and acted on the results of the interim assessments. Philadelphia principals were required to complete and review data analysis protocols for their schools with their regional superintendents and to analyze school-level results with other principals at monthly meetings. Although the intent of

these meetings was to generate constructive dialogue about instructional improvement, some educators viewed the public sharing of data as undermining the low-stakes, instructional focus of the interim assessments. In Cumberland, expectations for the use of the interim assessments were communicated through the district's curriculum and instruction staff, keeping the stakes of the interim tests low and maintaining the emphasis on instructional use.

Finally, principals in our Philadelphia study schools reinforced the district's expectations for data use by modeling and monitoring teachers' analysis of the interim assessments. The principals conducted their own analysis of assessment results to identify struggling students; areas of weak skills within a grade level; teachers whose classes might be falling behind those of their grade-level colleagues; and/or subgroup performance. Principals reviewed teachers' data analysis protocols and they discussed the results of the interim assessments with teachers in grade-group meetings. In some of our study schools, principals looked for evidence of the reported re-teaching strategies in teachers' lesson plans. District- and school-based curriculum specialists played this role in Cumberland, identifying common problems within grades and across schools and discussing assessment results with teachers.

Technology and data. The districts developed "*user-friendly*" *instructional management systems* (IMS) that scored student responses and facilitated teachers' analysis of interim assessment data. The IMS in Philadelphia enabled a teacher to view the performance of each student and the entire class on individual test items and by content standard, all with a click of the mouse. The system also displayed each student's incorrect answer choices. The Cumberland IMS was less sophisticated. After teachers entered test results, the system automatically highlighted

students who missed more than one test item for a learning standard. While teachers could not conduct item analyses using this IMS, they could easily see how many and which students were possibly weak in a particular learning area. The Philadelphia IMS also provided links to information on how to re-teach a particular standard and practice worksheets for students. Philadelphia provided professional development on its IMS, but the training focused on how to access and use the components of the system ("point and click"), rather than on how to analyze the interim assessment data.

Both districts assumed that their interim assessments, which were aligned with instructional units, would provide teachers with valid and actionable information; that is, teachers could diagnose student error and design appropriate re-teaching based on the results of the tests, including an analysis of incorrect answers. As will be discussed below, it is doubtful that these interim assessments fulfilled this goal.

Time. Both districts created dedicated time for teachers to discuss assessment results and instructional techniques, to re-teach content and skills to students, and to participate in professional development. Philadelphia created six-week cycles of instruction and assessment: five weeks of instruction (tied to the district's pacing guide) culminating with the interim assessments, and a sixth week of review and/or extended development of topics. At the end of the sixth week, teachers moved on to the next instructional unit. The cycle was somewhat different in Cumberland, where teachers generally administered the interim assessments, or practice tests, 3 to 5 days prior to giving a curriculum-embedded end-of-unit test that was part of a student's grade. In both cases, the districts expected teachers (and where available, other support staff) to provide remediation for students in areas of weakness and enrichment in areas of strength during these re-teaching periods. While schools in both Philadelphia and Cumberland created common planning time for teachers in the same grades, and set aside dedicated time for professional development during the school day, student or school issues or district-directed professional development often limited time available for teachers to discuss interim assessment results and common instructional challenges.

Professional support. The districts identified two targets for professional support: teachers and students. Each of our Philadelphia schools had a school-based teacher leader (SBTL) whose job was to assist teachers with data analysis and instruction in mathematics. In three schools this was a part-time position, while in two the SBTL was full-time. The SBTLs helped teachers analyze data and locate additional instructional materials, but they had limited time to provide instructional support to teachers. In Philadelphia, teachers occasionally had other adults in their classrooms (generally volunteers or student teachers) to instruct struggling students or work with groups of students, thereby enabling teachers to implement small-group instruction. Many of our Philadelphia teachers made time during lunch hours and before or after school to give students additional support. In contrast, Cumberland had an extensive system of instructional support for teachers and students through a district mathematics coach and full-time school-based elementary curriculum specialists (ECS) who served students as well as faculty. Cumberland schools also had mathematics aides who worked directly with teachers. This staffing facilitated the use of group instruction in the classroom and provided more intensive remediation to students identified as needing additional help.

As we see below, the type and level of district and school supports were one, but not the only, factor affecting how teachers used interim assessments. Others included teacher analysis and interpretation of interim assessment data as well as teacher capacity to assess and learn from students' mathematical understanding.

School-level Implementation of Interim Assessments

When well supported by districts, interim assessments were implemented in schools in a manner consistent with district expectations; that is, we found that teachers were attempting to use interim assessment results for instructional improvement. In Philadelphia, the majority of the teachers in our study had accessed their classroom-level interim assessment results using the IMS, and they had all visited the "item-level analysis" view that had

been featured in district and school professional development sessions. Most teachers were comfortable accessing those results directly using the IMS, although a few teachers relied on others in their building to print data out for them. In Cumberland, all teachers were inputting their classroom interim assessment scores into that district's IMS. Among our sample, interim assessments had become part of teachers' regular work.

We found, not surprisingly, that teachers used interim assessment results in the context of information from their own measures of student learning. Teachers rarely reported being surprised at the results of the interim assessments, and they mentioned that the interim assessments largely confirmed what they already knew about student learning in mathematics. These assessments did, however, provide uniform data that could serve as the basis for cross-classroom conversation. In addition, receiving the interim assessment results led some teachers to follow up by collecting more assessment information directly from students. For example, several of our teachers questioned students on their responses to individual interim assessment items and then used this information to plan further instruction.

Teacher Analysis and Interpretation of Interim Assessment Data

In both Philadelphia and Cumberland, teachers in our sample used the data from interim assessments to identify weak content areas or struggling students within their classes. We found that teachers in Philadelphia often set thresholds for student performance (a score below which instructional response would be warranted), and that these thresholds varied from school to school and often from teacher to teacher. Teachers in Cumberland were much less likely to employ specific thresholds of performance in their review of interim assessment data. Rather, they were more likely to speak of these results in the context of other information from their own formative assessment practice.

Teachers' processes for interpreting interim assessment data with an eye toward instructional planning were influenced by a variety of factors, including their knowledge about specific students' backgrounds or past

performance, student performance in relation to their peers, district factors such as the scheduling of interim assessments relative to the pacing guide, or teacher perceptions about which mathematical content was especially challenging for students. Furthermore, the IMS in each district influenced the steps teachers took in analyzing interim assessment results. Although the design of the IMS highlighted areas of weakness in student performance in both districts, the system in Philadelphia enabled teachers to easily link assessment items, curricular content, and standards. In contrast, results were presented in Cumberland's IMS by content area, with references to specific item numbers. This required teachers to take an extra step in relating results back to specific items, a step that fewer teachers took. Because much of the interest in interim assessments stems from the assumption that they can be used to improve teaching by providing data on students' understanding, a central focus of the study was the degree to which such assessments provided teachers with information about students' mathematical thinking. We explored this issue in three ways. First, we analyzed a sample of the assessments themselves, focusing on the extent to which the items provided teachers with actionable information about students' mathematical understanding. Among the Philadelphia assessments, only 6 of the 20 items in the 3rd-grade assessment and 5 of the 20 items in the 5th-grade assessment contained a set of distractors reflecting multiple errors that typical students might have. Further, distractors on only 2 to 3 items contained information on mathematical misunderstandings, as opposed to other sources of error (e.g., reading the problem correctly). The majority of the 3rd-grade Cumberland assessment items contained unclear expectations (e.g., directions requiring students to "explain," without specifying what about the answer they are to explain), answer choices that failed to reveal typical student errors, or content that was simply "not very mathematical". The majority of the Cumberland 5th-grade assessment items had either unclear instructions or a design that did not support diagnosis because the problem types were too mixed to allow for detection of misunderstanding across item types.

Second, we analyzed the ways in which teachers interpreted student error in a variety of contexts. We hypothesized that even if interim assessment items were of high quality, use of these data for instructional change or improvement required teachers to have the capacity to infer students' thinking or understanding from their responses to particular items. To explore the extent to which this occurred, we identified specific items from the interim assessments where there were plausible mathematical explanations for most or all of the incorrect answers. We then presented teachers with incorrect responses on these items, and prompted them to tell us what the student might have been thinking. In both districts, a wide range of responses were observed. In general, these responses fell along a continuum ranging from *procedural* explanations of student error to *conceptual* [2] diagnoses. For example, presented with a problem in which a student incorrectly added fractions with different denominators, some teachers would describe the student's failure to find a common denominator (procedural), while others were more likely to note that the student did not seem to understand that a denominator represents the number of parts into which the whole is divided (conceptual). Overall, Philadelphia teachers were more likely to engage in procedural diagnoses, while Cumberland teachers were more evenly distributed between procedural and conceptual explanations of student error. Therefore, student performance even on potentially high-quality assessment items led to a wide range of inferences made by teachers in our study.

Third, we explored teachers' use of interim assessments within the wider context of their formative assessment practice.

[2] In the mathematics education literature, procedural knowledge has been largely defined as "how to" and conceptual knowledge has been defined as "why" (Hiebert & Lefevre, 1986). Without joining current debates on the relationship between the two knowledge types (Baroody, Feil, & Johnson, 2007; Rittle-Johnson & Sigler, 1998), we adopt Hiebert and Lefevre's straightforward distinction as our definition for the purposes of this study. Of course, we acknowledge that both knowledge types are necessary for the development of mathematical competence.

This included "short-cycle" practices such as question routines and use of individual whiteboards or response cards, along with teacher-developed assessments such as homework tasks or selecting appropriate activities from the instructional materials. Using a combination of interview and observation data, we constructed profiles of teachers focused on how they collected, interpreted, and acted on interim, short-cycle, and teacher-developed assessment information. Overall, we found that teachers who focused on students' conceptual understanding using one type of assessment were more likely to do so for all types of assessment, including interim assessments. This suggests that analytic or diagnostic capacity underlies effective formative assessment, regardless of whether those assessments are embedded within instruction, developed by teachers, or externally designed.

In addition to diagnosing student error, many teachers in both districts attributed low performance on interim assessments to other learning challenges. These included a list of possible causes, including, but not limited to, weak reading ability, difficulty maintaining attention, and low levels of English language proficiency. Teachers in both districts also offered contextual or external explanations for students' struggles, such as lack of background knowledge, as contributing to difficulties in comprehending word problems. These factors were generally perceived to be outside of the teacher's or school's realm of influence.

Instructional Response to Interim Assessment Data

While teachers accessed and analyzed interim assessment data, we found that this information did not substantially change their instructional and assessment practice. Teachers used interim assessment results largely to decide what content to re-teach and to whom, but not to make fundamental changes in the way that this content or these students were taught. Teachers' use of classroom-based formative assessment did not necessarily lead to changes in instructional strategies either.

The interim assessment results informed decisions about *whom* to teach, and these decisions differed by district. In planning for

re-teaching time, the most common response among Philadelphia teachers was to revisit content using a combination of whole-group and small-group instruction with a smaller number of teachers adding peer-tutoring to this mix. In Cumberland, the results from the interim assessments were primarily used to inform ongoing flexible grouping based on students' various (mis-) understandings. These different ways of organizing re-teaching may have reflected, to some extent, the types and levels of support available to teachers. Individual remediation during class time was rare among the Philadelphia teachers in our study, due in part to lack of classroom support for practices like conferencing. In Cumberland, the lowest performing students could be referred to the curriculum specialist for learning support.

The interim assessment data influenced *what* teachers chose to re-teach, but not necessarily *how* to teach it. While we saw variation among teachers' responses within each district, teachers in Philadelphia largely emphasized re-teaching procedural steps in their instructional planning, perhaps reflecting their focus on students' procedural errors. In classrooms, we observed that this approach often took the form of teachers reworking examples (often problems from the interim assessments) with either the whole class or with a small group. While teachers in Philadelphia used the IMS to help with data interpretation, use of the IMS for more complex tasks, such as generating supplemental assessments or identifying curriculum, was far less common. In Cumberland, we also observed and heard about procedurally oriented approaches, but many teachers mentioned that part of re-teaching involves re-teaching a mathematical concept as well. Some Cumberland teachers made a distinction between re-teaching and "completely re-teaching," where the former might include follow-up lessons or student-worked examples on the board and the latter referred to direct instruction on a concept or skill that was already taught. This distinction is an important one, as re-teaching that emphasizes "ritualized skills and applications" is unlikely to lead to increased student learning (McMillan, 2010, p. 45).

We also examined the degree to which instructional strategies adopted by teachers during re-teaching differed from those used to teach the same content originally. To analyze change in instructional strategy, we explored the different ways in which teachers responded instructionally not only to interim assessment data, but to all classroom-based formative assessment information (e.g., teacher-developed quizzes, questioning routines, etc.). Nearly all of the teachers acted on formative assessment information with *organizational strategies*, with about half using it primarily or only in this way. For these teachers, formative assessment information was used to determine: what content to re-teach, which students need additional support, whether and how students should be grouped during re-teaching, and when to move on to the next concept or topic. Fewer than half of the teachers in our sample employed *instructional change strategies*—modifications in how they intended to re-teach specific content or students—in response to formative assessment information. Many of these teachers simply opted for teaching content "a different way," or made greater use of manipulatives in the hope that an alternate presentation might help students to grasp material with which they had struggled.

Teachers who assessed for conceptual understanding were far more likely to employ instructional change strategies than those who did not. Examples of these strategies included use of additional representations or models of mathematical concepts (e.g., the introduction of arrays for multiplication or set models for fractions) and connecting students' prior knowledge to current learning goals (e.g., relating algorithms for double-digit subtraction to triple-digit subtraction). We can hypothesize that assessing students' mathematical understanding (and not simply success or failure with procedures) affords teachers better opportunities to assess students' learning needs. Many times, these needs are inconsistent with past instructional approaches.

It is also possible that teachers who are able to assess for conceptual understanding are also more likely to have the capacity to respond with a varied instructional repertoire. When we conducted an in-depth examination of the instructional practices of a small subset of our teachers, we found that the instruction and formative assessment practice of teachers with high levels of mathematical knowledge for teaching (MKT) was generally centered around student understanding: the teachers

actively sought to learn how students think and they responded to student understandings. The level of mathematics in these rooms was relatively high, and student engagement was maintained. In contrast, the lack of small-group and individual student work in the rooms of teachers with lower levels of MKT, coupled with mathematically superficial questioning routines, was an obstacle to learning about student understanding. While these findings are based on data from only a small number of teachers, they raise questions about the possible relationships between teachers' knowledge of how to teach mathematics and their instructional practices.

Implications for Policy and Research

Research literature about the impact of interim assessments on student learning is at best inconclusive. Optimism about its potential largely derives from research on short-cycle formative assessment, which has been shown to improve both instruction and student learning. The critical question for policymakers, then, is whether interim assessments can be used formatively. Put another way, can teachers use interim assessment data to make instructional changes that are likely to improve student achievement?

Our study showed that interim assessments are useful but not sufficient to inform instructional improvement. When linked directly to a district's curriculum, interim assessments helped teachers make decisions about what content to re-teach and to whom by identifying areas in which specific students or the class as a whole were performing poorly. Where resources were available, interim assessments also allowed teachers to help students in need of additional, individualized supports.

Use of interim assessments for these purposes was facilitated by several district and school factors, including alignment of interim assessment content with standards and curriculum; expectations that interim assessment results would be used to inform instruction; a quality and accessible IMS that focused teachers' attention on content as well as on items; time to re-teach content and skills to students; and instructional supports for struggling students and professional supports for

teachers in data analysis and instruction. School leadership and a culture of data use were also critical factors in supporting teachers' use of data.

We found little evidence, however, that the interim assessments we studied helped teachers develop a deeper understanding of students' mathematical learning—a precursor to instructional improvement. Most items in the assessments did not provide actionable information on students' misunderstandings. In addition, teachers' capacity to interpret assessment data played a major role in how they used the results of interim, and even formative, assessment. Many teachers focused on procedural rather than conceptual sources of student errors on test items, diagnoses that appeared to inform their instructional planning during re-teaching. Teachers who assessed for conceptual understanding were more likely to use instructional change strategies that those who did not. Teachers' mathematical knowledge for teaching also appeared to contribute to teachers' instructional and assessment practices.

The findings from our study, along with those from related research on formative assessment and data-driven decision making, lead us to make the following recommendations about the design of assessment systems, supporting the use of interim assessments, and future research.

Focus, align, and inform. The design of interim assessments must reflect their intended use. While this study focused on the ways in which teachers used interim assessments formatively (i.e., to change instruction), interim assessments can also have predictive and/or evaluative purposes. Assessments should be chosen to serve a single purpose. If interim assessments are to be used formatively, they must be designed for instructional purposes. This may mean using other tests to meet predictive or evaluative goals. Assessments designed for instructional purposes must be closely aligned with district curriculum as well as district and state standards. This principle applies not only to the constructs that are assessed and the formats of the test, but to any supplemental components of the assessment. For example, recommended instructional strategies need to align to the instructional approach of the curriculum. Similarly, districts need to verify claims that

multiple-choice item distractors carry instructionally useful information. Mathematics items should be written so that distractors represent common errors in both procedure and conceptual understanding.

Support teachers and students. Even if interim assessments are focused, aligned with curriculum, and of high quality, their impact on teaching and learning depends on how their adoption and use is supported at the district and school level. District and school leaders need to communicate consistent and clear messages about the purpose and use of interim assessment. School leaders should model effective data use for teachers and other support staff and should allocate school-level resources to support interim assessment use for instructional purposes.

District IMS must return interim assessment data to teachers in a manner that is both timely and accessible; teachers must in turn be trained to use the IMS to its full capabilities. The goal should be to have teachers invest their time in interpreting results and planning instruction rather than navigating the IMS or entering data. Another critical factor is time. Whether highly structured or flexible, pacing schedules must allow time for re-teaching to occur. Additionally, teachers should have regular time in their schedules to analyze interim assessment results and discuss potential instructional responses.

While a major goal of interim assessment is to improve classroom instruction, our findings also suggest that a secondary use of such assessments may be to identify students in need of additional support, such as added instructional time or tutoring. Schools that already have these resources in place should consider using interim assessments (together with teacher input) to identify students in need of support. Where such supports are limited, schools should consider how to best respond to individual students who continue to struggle. This is an urgent issue given the multiple demands placed on teachers during regularly scheduled instructional time.

Build instructional capacity. Building a high-quality assessment system that is supported at the district and school levels is necessary for teachers to access, analyze, and discuss data. How well teachers use such data in the classroom, however, reflects their capacity to assess and teach for mathematical understanding. Teachers who assess for conceptual understanding do so across multiple test formats, and appear to be more apt to enact instructional change strategies than those who pay attention to students' procedural skills alone. Likewise, formative assessment, as a process, is heavily dependent on teacher capacity.

When looking to increase teacher capacity to use data for instructional improvement, districts and schools should consider that teachers need more professional development and support on interpreting data (e.g., diagnosing student error) and on connecting this evidence to specific instructional approaches and strategies. Therefore, professional development for interim assessment use should go beyond using "point and click" to locate and organize data and should emphasize analysis of student results in the context of standards and curriculum. Analysis should incorporate information from other types of assessment (e.g., in-class student work, teacher observation, etc.). In addition, part of using evidence of student learning to improve instruction is knowing how mathematical understanding develops and how to support students' progress toward a learning goal. Thus, in mathematics, professional development for teachers should focus as well on teacher content knowledge, developing teachers' instructional repertoires, and capacity to assess for students' mathematical learning.

The curriculum must be designed to allow for integration of assessment information from multiple sources and provide guidance for instructional response. In some cases, the potential for this opportunity lies within the current curriculum; for example, the program used by our study districts offers multiple types of assessment embedded within the curriculum as well as instructional suggestions for remediation and enrichment built into every lesson. In other cases, more appropriate programs or supplemental materials may need to be adopted. In addition, tools are being developed to enable the connection between interpretation and action; for example, newer technology platforms aim to link information gleaned from assessments with potential instructional responses.

However, adopting the right curriculum and tools are not, in themselves, sufficient to enable teachers to adjust instruction in response to assessment results. An extended research base supports the value of regular, facilitated teachers' analysis of student work to inform instructional decision-making. One such model features groups of teachers examining student work in collaboration with a content area expert (e.g., mathematics coach or curriculum specialist) on a regular basis throughout the school year. Teachers return to their classrooms with a list of possible instructional strategies developed by the group. The next meeting begins with teachers' reporting on the success and challenges of implementing instructional change. This information, along with new student work, forms the basis for the next discussion. It is this kind of ongoing, supported capacity-building that gives teachers the best chance at turning assessment results into increased student learning.

Research implications. This was an exploratory study focused on how teachers actually interpret and act on data from interim assessments. Below we make suggestions for further developing the field of research on interim assessments.

First, we see a need to develop a more comprehensive body of research that focuses on actual assessment use. We believe that the most potential lies in examining assessment use within particular content areas (e.g., reading, writing, mathematics, science, etc.). In this way, we can identify trends and relationships that exist within content areas as well as others that may apply more generally (e.g., the importance of timeliness of assessment results). Likewise, the role of teacher capacity for teaching and assessing within particular content areas is an important variable to consider when researching teacher assessment use.

Second, there needs to be research on the quality of data generated by interim assessments. This is a severely neglected area of research, yet poor data precludes effective data use. Claims about the validity of interim assessment results for instructional use need to be investigated as a matter of course.

Finally, research on assessment should examine interim assessment use in the context of the broader system of assessment. Current research tends to focus on individual assessments and not on the relationship among assessments. There is a need to examine the degree to which assessments of different types inform each other. For example, do teachers scaffold the information received from different assessments? To what degree do the characteristics of these assessments influence teacher use? Answering these questions necessitates observing the instruction that is part of the assessment cycle.

References

Baroody, A.J., Feil, Y., & Johnson, A.R. (2007). An alternative reconceptualization of conceptual and procedural knowledge. *Journal for Research in Mathematics Education*, *38*, 115-131.

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

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, WI: University of Wisconsin-Madison, Wisconsin Center for Education Research.

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

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81-112.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, NJ: Erlbaum (pp. 1-27).

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing measures of teachers' mathe-

matics knowledge for teaching. *The Elementary School Journal*, *105*, 11- 30.

McMillan, J.H. (2010). The practical implications of educational aims and contexts for formative assessment. In H.L. Andrade & G.J. Cizek (Eds.), *Handbook of formative assessment*. New York: Routledge (pp. 41-58).

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

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.

Rittle-Johnson, B., & Siegler, R.S. (1998). The relation between conceptual and procedural knowledge in learning mathematics: A review. In C. Donlan (Ed.), *The development of mathematical skill*. Hove, UK: Psychology Press (pp. 75-110).

Sawchuk, S. (May 13, 2009). Testing faces ups and downs amid recession. *Education Week*, *28*.

Stecher, B., Epstein, S., Hamilton, L. S., Marsh, J. A., Robyn, A., McCombs, J. S., Russell, J., & Naftel, S. (2008). *Pain and gain: Implementing No Child Left Behind in three states, 2004-2006*. Santa Monica, CA: RAND.

Suggested Readings

Bulkley, K.E., Oláh, L.N. & Blanc, S. (Eds.) (Forthcoming). Benchmarks for success? Interim assessments as a strategy for educational improvement. Special issue of the *Peabody Journal of Education*.

Corcoran, T., Mosher, F., & Rogat, A. (2009). *Learning progressions in science: An evidence-based approach to reform*. New York: Consortium for Policy Research in Education. (Free download at www.ccii-cpre.org)

Weinbaum, E.H. (2009). *Learning about assessment: An evaluation of a ten-state effort to build assessment capacity in high schools*. Philadelphia: CPRE (Free download at www.cpre.org)

About the Authors

Margaret E. Goertz is Co-Director of the Consortium for Policy Research in Education (CPRE) and a professor of education policy in the Graduate School of Education at the University of Pennsylvania (Penn GSE). She specializes in the study of state and federal education finance and governance policy. She has conducted extensive research on state education reform policies, state teacher policies, and state and federal programs for special-needs students. Her current research activities look at the impact of standards-based reform in elementary schools and high schools, the implementation of the No Child Left Behind Act of 2001, and state and local assessment and accountability policies. She also studies how school districts and schools allocate resources in support of standards-based reform.

Leslie Nabors Oláh is a Senior Researcher at CPRE and Research Assistant Professor at Penn GSE. She specializes in the relationship between cognition and instruction, including the role that teacher assessment practice plays in this relationship. Her current research activities center around formative assessment in elementary mathematics. She recently served as Co-PI on an NSF-funded study of teachers' use of mathematics formative assessments in Philadelphia. She was trained as a developmental psychologist at the Harvard University Graduate School of Education, where she also received an Ed.M. in Methodology in Developmental Research. She has published and presented work on infant vocabulary development and on K-1 mathematical development, and is a co-developer of the Penn Research-Informed Mathematics Education (PRIME) initiative, a professional development program designed to support urban teachers' instruction and assessment in mathematics.

Matthew Riggan is a researcher at CPRE and a Lecturer at Penn GSE. He specializes in school leadership, organizational learning and change, and qualitative research methods. His current research activities include evaluations of two initiatives: 1) the School District of Philadelphia's Twenty First Century Skills Project, an initiative to improve postsecondary transitions for high school students; and 2) an evaluation of the Benwood Initiative, an effort to scale-up promising reforms in low-performing schools in Hamilton County, TN. He is also involved in two studies focused on how teachers generate, interpret and act on information about student understanding of mathematical concepts.

CPRE Policy Briefs

CPRE

About CPRE

Established in 1985, the Consortium for Policy Research in Education (CPRE) unites researchers from seven of the nation's leading research institutions—the University of Pennsylvania, Teachers College Columbia University, Northwestern University, Harvard University, Stanford University, and the Universities of Michigan and Wisconsin-Madison—in a joint venture to improve, through research, the quality of teaching and learning in public schools in the United States and internationally.

For extensive information about all CPRE projects, please visit our Center websites:

CPRE main website: **www.cpre.org**

Center on Continuous Instructional Improvement (CCII): **www.ccii-cpre.org**

Strategic Management of Human Capital in Public Education (SMHC):
www.smhc-cpre.org

Study of Instructional Improvement (SII): **www.sii.soe.umich.edu**

**Sign up for CPRE's electronic newsletter, *In-Sites* (email: **insites@gse.upenn.edu), for updates on new publications, project research findings, and much more.

Nondiscrimination Statement

The University of Pennsylvania values diversity and seeks talented students, faculty and staff from diverse backgrounds. The University of Pennsylvania does not discriminate on the basis of race, color, sex, sexual orientation, gender identity, religion, creed, national or ethnic origin, citizenship status, age, disability, veteran status or any other legally protected class status in the administration of its admissions, financial aid, educational or athletic programs, or other University-administered programs or in its employment practices. Questions or complaints regarding this policy should be directed to the Executive Director of the Office of Affirmative Action and Equal Opportunity Programs, Sansom Place East, 3600 Chestnut Street, Suite 228, Philadelphia, PA 19104-6106; or (215) 898-6993 (Voice) or (215) 898-7803 (TDD).