CRESST REPORT 741

Margaret Heritage
Jinok Kim
Terry P. Vendlinski
Joan L. Herman

FROM EVIDENCE TO ACTION: A SEAMLESS PROCESS IN FORMATIVE ASSESSMENT?

AUGUST, 2008

National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

From Evidence to Action:
A Seamless Process in Formative Assessment?

CRESST Report 741

Margaret Heritage, Jinok Kim, Terry P. Vendlinski and Joan L. Herman
CRESST/University of California, Los Angeles

August 2008

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported in part under the National Research and Development Centers, PR/Award Number R305A050004, as administered by the U.S. Department of Education's Institute of Education Sciences (IES).

FROM EVIDENCE TO ACTION:

A SEAMLESS PROCESS IN FORMATIVE ASSESSMENT?

Margaret Heritage, Jinok Kim, Terry P. Vendlinski, & Joan L. Herman
CRESST/University of California, Los Angeles

Abstract

Based on the results of a generalizability study (G study) of measures of teacher knowledge for teaching mathematics developed at The National Center for Research, on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles, this report provides evidence that teachers are better at drawing reasonable inferences about student levels of understanding from assessment information than they are in deciding the next instructional steps. We discuss the implications of the results for effective formative assessment and end with considerations of how teachers can be supported to know what to teach next.

Assessment is essential to effective teaching and learning. Black and Wiliam (1998; 2004) stress the critical role of formative assessment, in particular. Formative assessment is a systematic process to continuously gather evidence and provide feedback about learning while instruction is underway. The feedback identifies the gap between a student's current level of learning and a desired learning goal (Sadler, 1989).

In the process of formative assessment teachers elicit evidence about student learning using a variety of methods and strategies, for example, observation, questioning, dialogue, demonstration, and written response. Teachers must examine the evidence from the perspective of what it shows about student conceptions, misconceptions, skills and knowledge. Sometimes teachers examine evidence on a moment-by-moment basis during the course of a lesson. Other times they review evidence after a lesson or series of lessons. In all instances, they need to infer the gap between the students' current learning and desired instructional goals, identifying students' emerging understanding or skills so that they can build on these by modifying instruction to facilitate growth. The analysis and interpretation of evidence is pivotal for the effectiveness of formative assessment. Inaccurate analyses or inappropriate inference about students' learning status can lead to errors in what the next instructional steps will be, with the result that the teacher, and the learner fail to close the gap.

For assessment to be formative, action must be taken based on the evidence elicited to close the gap (Black, Harrison, Lee, Marshall & Wiliam, 2003; Sadler, 1989; Wiliam &

Thompson, 2007). For teachers this means knowing what action to take based on the evidence that they have obtained so that they "adapt the teaching work to meet the learning needs" (Black et al., 2003, p. 2). However, is adapting instruction to meet learning needs always within the competence of teachers? In this report, we illustrate through the results of a generalizability study (G study) of measures of teacher knowledge for teaching mathematics, that moving from evidence to action may not always be the seamless process that formative assessment demands. While initially aimed at the properties of teacher knowledge measures, the G study results provide interesting data showing that teachers do better at drawing reasonable inferences of student levels of understanding from assessment evidence, while having difficulties in deciding the next instructional steps.

First, we will describe the measures of teacher knowledge that were used in the study, then present a description of the G study and results, and finally we will discuss some possible ways in which teachers can be supported to use evidence more effectively to inform action.

Teacher Knowledge Measures

The purpose of the teacher knowledge measures used in the G study is to gauge the effects of POWERSOURCE©, a formative assessment strategy being developed at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles. POWERSOURCE© is expected, through professional development and job aids, to influence teachers' domain knowledge and pedagogical content knowledge and assessment practices in key principles underlying mastery of algebra I—specifically, the distributive property, solving equations and rational number equivalence. The measures conceptualize teacher knowledge as central to, and embedded in, the everyday practice of teaching, irrespective of teachers' specific curriculum or approach to teaching. This includes the knowledge that teachers draw on: (a) to interpret students' understanding of mathematical ideas and plan instruction (Ball, Lubienski & Mewborn, 2001; Shulman, 1986; Wilson, Shulman, & Rickert, 1987); (b) to give students feedback (NRC, 2000; 2001a); and (c) to explain, justify and model mathematical ideas to students (NRC, 2001b).

The measure is a series of performance tasks in which teachers are asked to review student responses to assessments and answer a series of questions. Shown in Figure 1 is an example of one student's response to an assessment of understanding of the distributive property.

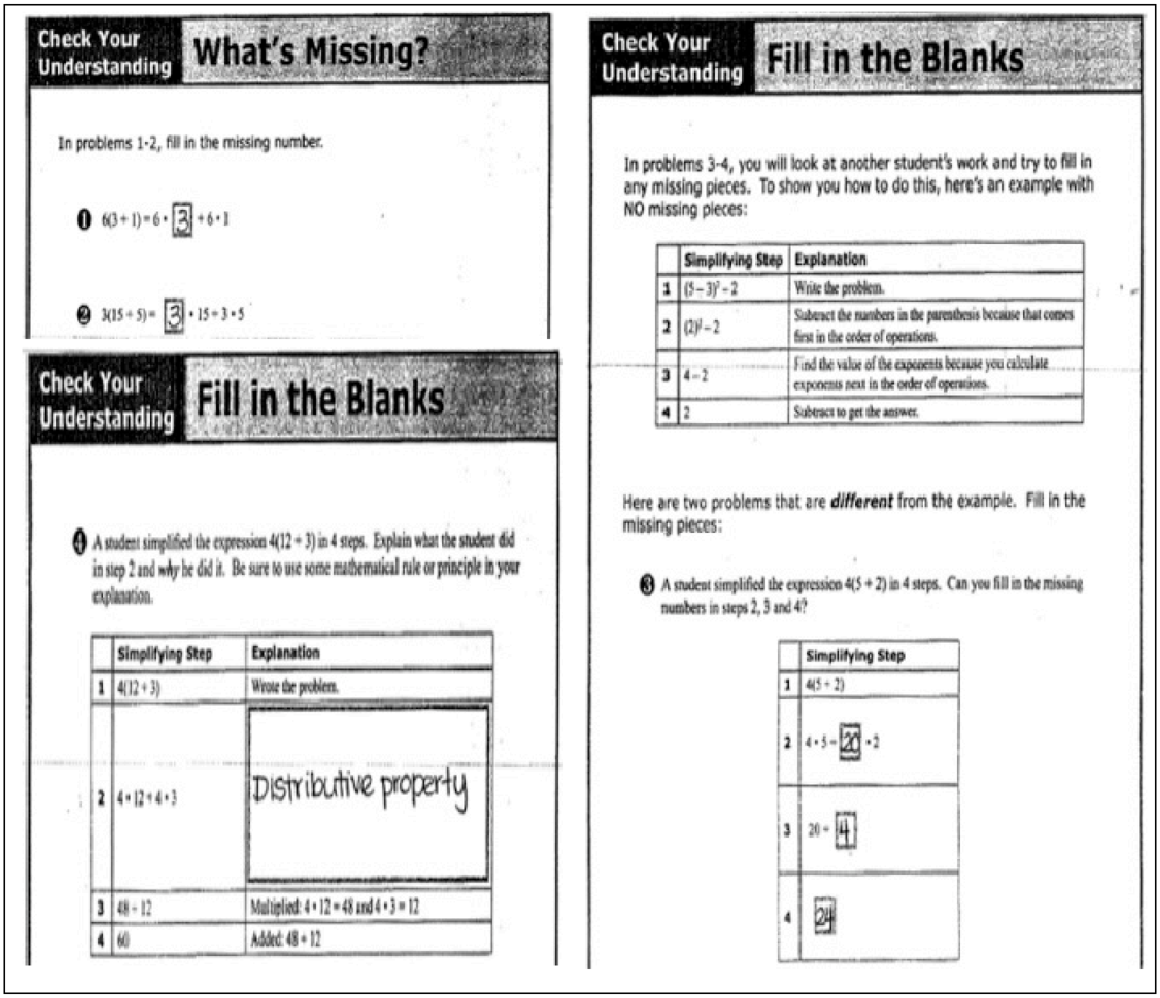

Figure 1. Student response to an assessment checking understanding of the distributive property

After reviewing the student response, teachers are asked to answer the following questions:

1. What is the key principle that these assessments address? Why do students need to understand this principle for algebra I?
2. What inferences would you draw from this student's responses? What does this student know? What does this student not know?
3. If you were this child's teacher what written feedback would you give to this student?
4. If this student were in your class, based on your responses to Questions 2 and 3, what would you do next in your instruction?

Developing the Teacher Knowledge Measures and Scoring Rubric

The student responses included in the teacher knowledge measures were drawn from a pilot study of the POWERSOURCE© assessments focusing on the distributive property, rational number equivalence and solving equations. Students in the Los Angeles Unified School District (LAUSD) completed the assessments. A group of mathematicians and expert mathematics teachers analyzed 40 LAUSD student responses to each of the POWERSOURCE© assessments and selected responses that were representative of the gaps in knowledge and misconceptions demonstrated by the students, as well as those that showed an understanding of the key principles. These student responses and the accompanying questions described above formed the teacher knowledge measures, which were designed to be scored by raters using a scoring rubric.

The development of the rubric was undertaken in several stages. First, eight mathematics teachers of varying levels of experience and expertise completed the teacher knowledge measures. Their responses represented a considerable range of knowledge. For example, in response to the question about the principle that the assessment shown in Figure 1 addresses, and why it is essential for algebra, responses ranged from "order of operations—essential for solving problems" to "distributive property, which is needed in algebra I" to "solve algebraic equations and multiply mononomials and polynomials." There were also many different ways that teachers expressed what they would teach next based on the student responses, including "repeated addition of the same thing," to "explain that the multiplier hooks up with one addend and must hook up with the other so it doesn't feel left out," to "factoring in reverse," as well as answers that bore no relationship to the distributive property, whatsoever.

Second, a group of university mathematics experts and expert teachers reviewed the teacher responses to each question and put them into four categories. The categories reflected the most rudimentary response (i.e. no mention of the distributive property) to those that the group viewed as the most sophisticated (i.e. distribution as repeated addition). They developed summaries for each category, which became the first draft of the 4-point score rubrics

Third, the draft rubrics were reviewed by another group of seven expert mathematics teachers, all involved in professional development, and with over 200 years of experience among them. This group used the draft rubrics to examine over 100 teacher responses. The content of each score point was discussed and consensus reached on revisions that needed to be made. Figure 2, for example, represents the consensus scoring rubric for the question "What would your next instructional steps be?" in relation to the student's response represented in Figure 1.

4	• Explain the distributive property as repeated addition • Explain factoring as distribution in reverse • Model the use of the distributive property with whole numbers • Model generalizing to other numbers and variables
3	*Either* • Explain the distributive property as repeated addition *Or* • Explain factoring as distribution in reverse • Model the use of the distributive property with at least whole numbers
2	• Explain procedures for how to use the distributive property, equating procedures with the order of operations
1	• No explanation of the distributive property

Figure 2. Scoring rubric for teachers' explanations of next instructional steps in teaching the distributive property.

A score of 4 shows that the teacher understands the distributive property as repeated addition, factoring as distribution in reverse, using the distributive property with whole numbers and generalizing the distributive property to other numbers and variables.

A score of 3 shows that the teacher has an understanding of either the distributive property as repeated addition or of factoring as distribution in reverse and of using the distributive property with at least whole numbers.

A score of 2 shows the teacher has a rudimentary, procedural, rather than principle-based understanding of the distributive property.

A score of 1 indicates that the teacher response contains no explanation of the distributive property.

The G Study

In our G study, the object of measurement was teachers' pedagogical knowledge in mathematics. We aimed to determine which components of variability in teachers' knowledge were likely to be responsible for overall scores on the teacher knowledge measures, and to determine how applicable our conclusions from our sample of teachers in the study were to teachers in general.

Sample of participating teachers. One hundred and eighteen sixth grade teachers from across Los Angeles County participated in the study. The teachers completed the measures online within their own time frame and at their own pace. In addition to completing the measures, teachers responded to questions about their professional background. At the time they completed the performance tasks, 97% of the teachers mostly taught Grade 6. About half of the teachers taught in a self-contained classroom, with responsibility for teaching all subjects, while the other half taught in a mathematics department, with responsibility for teaching mathematics only. Fifty percent of the teachers who taught in mathematics departments, not in self-contained classrooms, taught general mathematics and the rest were distributed across pre-algebra, remedial mathematics, special education mathematics geometry (3%), integrated mathematics program, and algebra. About half of the teachers had taught math for 5 or less years, ranging from 0–38 years.

About half the sample had taken three or more undergraduate or graduate level classes in mathematics. Approximately 25% had taken two classes, while the remainder had taken one class or none. When asked about the total time they spent on in-service education in mathematics during the past year, 24% of the teachers reported none, 23% reported less than 6 hours, 17% reported 6–15 hours, 17% reported 16–35 hours, and 20% reported more than 35 hours. In other subjects, 15% of the teachers reported none, 30% reported less than 6 hours, 22% reported 6–15 hours, 18% reported 16–35 hours, and 15% reported more than 35 hours. In response to question about the number of undergraduate or graduate level classes they took in mathematics, about 9% reported none, 17% reported one class, 23% reported two classes, and 53% reported three and more classes. In methods of mathematics, 24% reported none, 40% reported one, 17% reported two, and 18% reported three or more classes.

A number of questions were asked about teaching credentials. In terms of grade levels, more than 90% of the teachers had credentials for kindergarten to Grade 6, about 60% for Grades 7–8, and a lesser percentage for higher grades (34% for Grade 9, about 20% for Grades 10–12). In terms of subject matter 90% of the teachers had general credentials (i.e., all subjects), 14% for mathematics and 3% for science. In terms of credential status 72% had completed credentials, 20% had preliminary, and 6% had intern status.

Performance Tasks

As already described, the teacher knowledge measures used in our study were a series of performance tasks designed to be scored by raters using a 4-point scoring rubric. As in any other measurement instrument, scores based on performance tasks may be subject to various sources of error. In this study, we assumed three potential sources of error in assessing teachers' demonstrated mathematical knowledge for teaching. First, scoring of the same teacher's response may vary depending on the rater. For example, some raters may be more stringent while others more lenient, even when provided with an identical, detailed rubric. Second, the different principles may be a source of error. Among the three key mathematical principles that we focused on (i.e., the distributive property, solving equations, and rational number equivalence), the same teacher may have more knowledge about one principle than the others, which may lead to variability in scores for that teacher. Third, scores may vary depending on the assigned task. For example, variation in the same teacher's scores could be because the teacher may be better at identifying the principle and evaluating students' understanding, whereas the teacher may struggle with planning the next instructional steps. These different conditions that may give rise to the variability of a teacher's score are considered sources of measurement error, also known as *facets*.

Method and Results

We performed an ($o \times r \times p \times t$) G study to examine the magnitude of the score variation due to the main and interactions effect of: teacher—the object of measurement (o), rater (r), principle (p), and type of task (t). From this design the total variation of the observed score was partitioned into 15 terms, shown in Table 1. In addition to the partitioned terms, Table 1 presents the results of the G study analysis, showing the estimated variance components and percentages of the magnitudes of estimated variance components as compared to the total score variability.

Table 1

Estimated variance components and percentage of score variation for scores in teacher performance tasks

Source of Variability	*n*	EVC	%
Teacher (o)		0.0594	5.7
Principle (p)	3	0.0172	1.7
Type of task (t)	3	0.2591	24.9
Raters (r)	6	0.0024	0.2
op	3	0.0149	1.4
ot	3	0.0049	0.5
or	6	0	0.0
pt	9	0.0848	8.2
pr	18	0.0026	0.3
tr	18	0.0007	0.1
opt	9	0.4098	39.4
opr	18	0.0016	0.2
oqr	18	0	0.0
ptr	54	0.0086	0.8
opqr,e	54	0.1736	16.7

Note. Two small negative variance components were considered to be negligible and were subsequently set to zero.

The results show that the true-score variance (Estimated Variance Component [EVC] of teacher main effect) is only about 6% of the variability of observed scores, which indicates that a large extent of the variability is associated with other facets contributing to measurement error. Among the main effects, both the main effect of principle and rater are negligible. The main effect of the seven raters in the study is extremely small, contributing to only 0.2% of the total variability. However, one main effect large enough to be problematic (25%) is type of task. This suggests that teachers' scores may not be generalizable across the different tasks: (a) identifying key principles; (b) evaluating student understanding; and (c) planning the next instructional step based on the evaluation of student understanding. In other words, some of the tasks are more difficult, in general, for all teachers than the other tasks.

Among the interaction effects, the results show that any interaction terms that involve rater facet contribute to only a negligible percentage of the total variability. This shows that raters are very consistent; therefore, we determined that using samples of a small number of raters should be enough to obtain dependable scores. However, two interaction terms represent a noticeably large percentage of the total variability. One is a two-way interaction between principle and type of task (8%), while the other is a three-way interaction among teachers, principle, and type of task (39%). Two issues are worth noting from these interaction terms. First, the EVC of the three-way interaction is the largest magnitude (39%) among all 15 terms, which is problematic. Second, both interaction terms involve two facets: principle and type of task.

The two-way interaction term, (principle and type of task) that is of a relatively moderate magnitude, indicates that some combinations of principle and type of task are more difficult for teachers than others. The three-way interaction term (teacher, principle and type of task) that resulted in the largest magnitude implies that teachers' average scores are fairly inconsistent from one combination of principle and type of task to another. To elaborate, the two-way interaction term suggests some combinations of principle and type of task tend to be more difficult (no matter who the teacher is), whereas the three-way interaction term further suggests that some combinations are more difficult for some teachers but not for others. Thus, the large magnitude of the three-way interaction implies that the relative standings of teachers will change depending on the combination of principle and type of task.

Based on the results of the G theory study[1] we examined scores in different types of tasks and principles in order to generate hypotheses on underlying sources of such inconsistencies (i.e., inconsistencies across tasks, inconsistencies across task and principle combinations, and inconsistencies of teacher rankings across task and principle combinations). We hypothesized that teachers would have greater difficulty determining the next instructional steps from evidence than identifying the key principle addressed by the assessment and drawing inferences about student understanding.

As seen above, all of the terms that include raters—both main and interaction effects—have negligible magnitudes, which implied that we could select scores from a small number of raters and have them represent scores that teachers would have received from the population of raters. Based on these results, we averaged scores over raters. Table 2 shows the descriptive statistics of scores averaged over raters, by mathematics principle and by type of task.

[1] For readers who may be interested, G theory provides a generalizability coefficient (a G coefficient) that is analogous to the reliability coefficient, as well as partitioning and estimating variance components from various sources. G coefficients can be defined and calculated in two ways: one for the relative decision and the other for the absolute decision. The above analysis presented in Table 1 yields G coefficients of 0.52 for the relative decision, and of 0.27 for the absolute decision.

Table 2

Scores measuring teachers' pedagogical knowledge averaged over raters, by principle and by type of task

Variable	*N*	Mean	Std Dev	Min	Max
Principle: distributive property					
Task: identifying key principle	114	2.07	0.63	1	3.67
Task: evaluating student understanding	113	2.14	0.94	1	4.00
Task: planning next instruction	113	1.21	0.36	1	2.00
Principle: solving equation					
Task: identifying key principle	112	1.82	0.89	1	3.83
Task: evaluating student understanding	111	2.06	0.93	1	3.83
Task: planning next instruction	112	1.21	0.39	1	2.83
Principle: rational number equivalence					
Task: identifying key principle	105	2.94	0.47	1	4.00
Task: evaluating student understanding	102	2.07	0.98	1	4.00
Task: planning next instruction	101	1.36	0.48	1	2.83

Table 2 suggests that regardless of the math principle, determining the next instructional steps based on the examination of student responses tends to be more difficult for teachers. For all three principles, the teachers, on average, scored only 1.2–1.4 on this task, although they scored on average 1.8–2.9 in other tasks. This may underlie the large variance component of the main effect of type of task. This result illustrates, as we hypothesized, that teachers tend to be better at identifying the principle and drawing inferences about students' understanding than they are at deciding the next instructional steps.

One result from the G study that is hard to examine with the descriptive results in Table 2, is the three-way interaction among teacher, principle, and task. Since the scores are averaged over principle and task, inconsistency in teachers' relative standing across principles and tasks should not appear in the average scores—higher scores for some teachers and lower scores for others will cancel each other out, resulting in average scores that do not reflect relative ranks of teachers. Since this inconsistency is the largest component that contributes to error variability, it is important to consider why this happens. For instance, one combination of task and principle may have been more difficult for some teachers but not for others, probably because teachers have different areas of expertise and or different amounts of exposure to teaching with regard to key principles. Alternatively, teachers may have different capabilities in one type of task as opposed to the others: For example, some teachers may be better at planning the next instructional step given similar levels of knowledge in identifying key principles and in evaluating student understanding.

These results raise an important question about teachers' abilities to determine what to teach next in response to assessment information. Although they are only focused on three topics in mathematics, other researchers have also found similar issues with regard to teachers' knowing what to teach next in reading and mathematics (for example, Fuchs & Fuchs, 2008). As discussed previously, the purpose of formative assessment is to adjust teaching based on evidence about learning so that students can close the gap between their current learning status and desired goals. If teachers are not clear about what the next steps to move learning forward should be, then the promise of formative assessment to improve student learning will be vitiated. We now turn to some considerations of how teachers could be supported to improve their skills in moving more effectively from evidence to action.

Improving the Translation of Evidence to Action

To know what to do next instructionally in response to formative assessment evidence, teachers need clear conceptions of how learning progresses in a domain; they need to know what the precursor skills and understandings are for a specific instructional goal; what a good performance of the desired goal looks like; and know how the skill or understanding increases in sophistication from the current level students have reached. In this regard, conceptions of how learning progresses that are represented in typical curricula and standards at this time are not helpful to teachers. For example, standards rarely present a clear conception of how learning progresses in a domain, and curricula are often organized around scope and sequence charts, usually defining discrete objectives that are not connected to each other in a larger network of organizing concepts that show a clear trajectory in learning (NRC, 2000).

Recently, considerable interest has emerged in learning progressions (Gong, 2006, 2007; Heritage, 2008; NRC, 2001a; NRC, 2005). Learning progressions describe how concepts and skills increase in sophistication in a domain from the most rudimentary to the highest level, showing the trajectory of learning along which students are expected to progress. From a learning progression, teachers can access the big picture of what students need to learn, they can grasp what the key building blocks of the domain are, while having sufficient detail for planning instruction to meet short-term goals (for a more complete description of learning progressions, see Heritage, 2008). Teachers are able to connect formative assessment opportunities to short-term goals as a means to keep track of how their students' learning is evolving to meet the goal. Sometimes this will mean that teachers have to move backwards along the progression: for example, if formative assessment evidence shows that students are missing key building blocks. Similarly, they might move learning further forward if the evidence indicates that some students are outpacing their peers. In both cases, the progression helps teachers to make an appropriate match between instruction and the learners' needs. Such descriptions of learning will go a long way to helping teachers understand how learning progresses in a domain and to make appropriate matches between the learner and what needs to be learned next.

However, learning progressions, by themselves, will not be sufficient. A deep knowledge of the domain represented in the learning progression is also needed. Specifically, for effective formative assessment, teachers need to know what good performance of specific short-term learning goal looks like. This means they must also know what good performance does not look like. For example, in mathematics, can teachers identify what forms student misunderstandings or misconceptions can typically take? If teachers are clear about these aspects, they are better placed to respond to them when they show up in formative assessment. They have a depth of knowledge about how the concepts develop that enables them to adapt instruction to meet the learners' needs. In other words, they know what to teach next.

The depth of teacher knowledge of U.S. teachers, particularly as it relates to teaching mathematics, is an issue that has attracted attention. A number of researchers have contrasted the development of U. S. teachers' knowledge in mathematics with that of teachers in high-performing countries as measured by international assessments (for example, Ma, 1999; Stigler & Hiebert, 1999). U.S. teachers, by and large, are expected to know how and what they will teach upon graduating from teacher preparation programs (Schifter, 1996). In high-performing countries, after the completion of teacher preparation programs, teachers learn *from* teaching. This is well exemplified in Japan, where through a process of lesson study, groups of teachers collaborate to develop, evaluate and revise lessons on particular problem areas. Teachers "can collect information on how students are likely to respond to challenging problems, and they can plan which responses to introduce and in which order" (Stigler & Hiebert, 1999, p. 156.) The net result of the lesson study process is deep knowledge of how students reveal specific problems and how to address them. Teachers are able to anticipate student misunderstandings and misconceptions, and know what to do about them when they arise.

Lesson study in Japan is a meticulous, ongoing process that is ingrained in the culture of teaching and designed to build deep knowledge for teaching mathematics. While a lesson study process may not be the answer to deepening U. S. teachers' knowledge, nonetheless, it stands in contrast to the professional training reported by the teachers in our sample: 47% of teachers reported less than 6 hours of in-service mathematics training during the previous year, with only 20% reporting more than 35 hours; only 18% reported three or more classes in mathematics methods during graduate or undergraduate courses, while 64% reported one or less. These reports hardly evidence experiences that lead to deep knowledge of mathematics for teaching. In addition, as U.S. teachers have many more contact hours that teachers in Organisation for Economic Co-operation and Development (OECD) countries (OECD, 2003, cited in Wiliam, 2006) it is fair to assume that the sample of teachers, even in schools and districts with structures in place for teacher collaborative learning, do not have adequate time to engage in deep, reflective and ongoing discussion with each other that could lead to the kind of in-depth mathematical knowledge of their Japanese counterparts.

Ma (1999) makes a compelling case for the relationship between students' mathematical knowledge and teachers' mathematical knowledge. We hypothesize that if the teachers in our sample clearly understood the developmental trajectory of mathematical ideas—such as the distributive property, rational number equivalence and solving equations, represented in a learning progression—and had a deep knowledge of how the elements of these ideas manifest in student learning, they would likely have performed better on determining next instructional steps. Such knowledge would remove the uncertainty of knowing what to teach next based on evidence that is accumulated during the course of instruction through formative assessment and better realize the benefit of formative assessment to learning.

Conclusion

In this report, we have presented the results of a G study of measures designed to assess teachers' mathematics knowledge for teaching (Ball & Bass, 2000; Hill, Rowan & Ball, 2005). The measures required teachers to use their mathematical content knowledge and pedagogical content knowledge in ways that mirror classroom practice. Specifically, participants had to use assessment information to infer what students did and did not know about key principles, decide what they would teach next based on their inferences, and provide feedback that would help students improve.

The interaction effects in the G study reveal considerable complexity in assessing teachers' pedagogical knowledge with performance task measures, and highlight the challenge of developing valid and reliable measures of teacher knowledge. Many questions remain from our study, including, what are the underlying characteristics of teachers that make a particular combination of task and principle more or less easy? This will require further research.

Despite the complexity demonstrated in the G study, one finding that clearly emerges is that using assessment information to plan subsequent instruction tends to be the most difficult task for teachers as compared to other tasks (for example, assessing student responses). Given the importance of adjusting instruction to formative assessment, this finding gives rise to the question: can teachers always use formative evidence effectively to "form" action, (Shepard, 2005)? Further research will reveal the degree to which this question bears on formative assessment in other aspects of mathematics and other learning domains. Given that the teachers' ability to know what to teach next and how to adapt instruction in light of evidence is critical to formative assessment, we believe that this is an area that warrants further investigation.

We have concluded that while evidence may provide the basis for action, it cannot in and of itself "form" the action. Action is dependent on teachers' knowledge of how learning develops in the domain and on their pedagogical content knowledge. We have considered some possible directions for deepening teacher knowledge in ways that could contribute to effectively "forming" action—in other words, translating evidence into the next appropriate instructional steps that will move student learning forward. It is beyond the scope of this report to discuss the means by which this kind of teacher knowledge can be more broadly acquired in the profession. However, we see teacher knowledge as critical to effective formative assessment. It is particularly significant in knowing what to do with evidence. Until teachers have better conceptions of learning to work with better conceptions of learning, and deepening their knowledge of how the elements of student learning are manifested, then the movement from evidence to action as a seamless process will remain a somewhat distant goal. This situation inevitably diminishes the potentially powerful impact of formative assessment on student learning.

References

Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.

Ball, D. L., Hill, H. C., & Bass, H. (2005, Fall). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*.

Ball, D. L., Lubienski, S., & Mewborn, D. (2001). Research on teaching mathematics: The unsolved problem of teachers' mathematical knowledge. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.). New York: Macmillan.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. New York: Open University Press

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–73.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.), *Towards coherence between classroom assessment and accountability: Part II. 103rd yearbook of the National Society for the Study of Education* (1st ed., pp. 20–50). Chicago: University of Chicago Press.

Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.

Fuchs., D. & Fuchs, L. S. (2008, Feb). *Curriculum-based measurement, teacher decision making and student learning: A program of research*. Presentation to the Community Researchers' Forum, Center for the Continuous Improvement of Instruction, Philadelphia, PA.

Gong, B. (2006, October). *Establishing learning goals for formative assessment*. Presentation at the Reidy Interactive Lecture Series, Nashua, NH.

Gong, B. (2007, June). *Learning progressions: Sources and implications for assessment*. Presentation at the CCSSO, Large-Scale Assessment Conference, Nashville, TN.

Heritage, M. (2008, February). *Learning progressions: Supporting instruction and formative assessment.* Washington DC: Chief Council of State School Officers

Hill, H. C., Rowan, B., & Ball, D. L. (2005, Summer). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42*(2), 371–406.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum Associates.

National Research Council (NRC). (2000). *How people learn: Brain, mind, experience, and school.* In J. D. Bransford, A. L. Brown, & R. R. Cocking, (Eds.), Committee on Developments in the Science of Learning. Washington, DC: National Academy Press.

National Research Council. (2001a). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. In J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), Committee on the Foundations of Assessment, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2001b). *Adding it up: Helping children learn mathematics.* In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2005). *Systems for State Science Assessment.* Committee on Test Design for K–12 Science Achievement. In M. R. Wilson and M. W. Bertenthal (Eds.), Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119–144

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage Publications.

Shephard, L. (2005, October). *Formative assessment: caveat emptor*. Paper presented at the ETS Invitational Conference, the Future of assessment: Shaping teaching and learning, New York, NY.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14

Stigler, J. W. & Hiebert, J. (1999). *The teaching gap.* New York: The Free Press.

Wiliam, D. (2006). Formative assessment: Getting the focus right. In M. Heritage & A. L. Bailey (Eds.), *Educational Assessment, 11*(3&4), 283–289.

Wiliam, D. & Thompson, M. (2007). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping, teaching, and learning.* Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, S., Shulman, L., & Rickert, A. (1987). 150 ways of knowing: Representations of knowledge in teaching. In J. Calderhead (Ed.), *Explaining teachers' thinking* (pp.14–124). London: Cassell.