Instructional logs are a valid, reliable, and relatively cost-effective way to collect detailed information about classroom practice.

1

Using instructional logs to identify quality in educational settings

Brian Rowan, Robin Jacob, Richard Correnti

IN THE EFFORT TO identify educational settings that are most effective in improving student achievement, classroom process (that is, the way in which a teacher interacts with his or her students) is a key feature of interest. Over the past several decades, researchers have studied numerous aspects of the school environment in an attempt to identify what factors lead to improved achievement for students. Some scholars have attempted to determine whether inputs to schooling, such as teacher characteristics, teacher certification, or resources, are related to student achievement.[1] Others have explored more macrostructural aspects of schooling (for example, governance structures or school choice) in an effort to identify what makes certain schools effective.[2] Still others have explored whether particular curricular interventions or policy initiatives like comprehensive school reform can improve or explain the quality of an educational setting.[3]

Yet efforts to identify the key aspects of the educational environment that lead to improvements in student achievement have had limited success. Observable teacher characteristics such as

education and certification have been found to explain little of the variation in teachers' ability to improve academic outcomes for their students.[4] It has been difficult to establish a clear link between school spending and school quality, and only a few curricular interventions or policy initiatives have rigorously demonstrated that they can make substantial improvements to student achievement.[5] For example, the What Works Clearinghouse, which was established by the U.S. Department of Education to provide guidance to educators about what programs are most effective for improving student achievement, identifies only two programs in beginning reading and one program in elementary school math that they believe provide strong evidence of effectiveness. And yet the recent value-added literature demonstrates substantial variation across and within schools in teaching effectiveness or productivity.[6]

The lack of connection between many of the inputs to schooling and student achievement is due in part to the loosely coupled nature of schooling. Sociologists have long argued that formal authority exerts limited control over what actually takes place in the classroom.[7] While states, districts, or school personnel may assert a particular policy or attempt to implement a particular reform agenda, teachers maintain a considerable degree of autonomy in their classrooms, and there is no guarantee that initiatives will be implemented as intended—or that they will be implemented at all. Therefore, it is essential that we do a better job of understanding what occurs in the classroom between a teacher and students. By providing tools to measure such interactions, researchers can help school personnel to identify teachers' strengths and weaknesses, appropriately target professional development, plan and evaluate the effectiveness of particular educational interventions, and establish more effective policy initiatives. Unfortunately, high-quality assessment of the student-teacher interaction occurs all too infrequently, despite the critical role that understanding and measuring such processes can play in school improvement.

The lack of high-quality measures of classroom practices and processes is in part due to the fact that it is quite difficult and often costly to measure the interaction between teachers and students. Teaching is an extremely complex task. Over the course of a nine-month academic year, the typical elementary school teacher conducts 140 or more lessons in a given academic subject for the twenty to thirty students in her or his classroom, sometimes differentiating instructional activities by student or subgroups of students. During any given lesson, a teacher's instruction typically unfolds along many dimensions. A teacher normally covers several content objectives at different levels of cognitive demand during a single lesson, works in several different behavior settings, and uses a variety of subject-specific instructional techniques. Although some features of classroom instruction are implemented repeatedly across the school year, many others are not, making instructional practice not only multi-dimensional but also highly variable across days of the school year.[8]

In this article, we identify the classroom processes and practices that we believe are important to measure, discuss the strengths and weaknesses of two common approaches to studying these processes—direct classroom observation and annual surveys of teachers—and then describe the ways in which instructional logs can be used to overcome some of the limitations of these two approaches when gathering data on curriculum content and coverage.

Classroom observations are expensive, require extensive training of raters to ensure consistency in the observations, and because of their expense generally cannot be conducted frequently enough to enable researchers to generalize observational findings to the entire school year or to illuminate the patterns of instructional change that occur across the school year. Annual surveys are less expensive but often suffer from self-report bias and the bias that occurs when teachers are asked to retrospectively report on their activities over the course of a single year. Instructional logs offer a valid, reliable, and relatively cost-effective alternative for collecting detailed information about classroom practice and can overcome some of the limitations of both observations and annual surveys.

Identifying important aspects of classroom processes

We have identified four primary aspects of the student-teacher interaction that can play a role in the overall quality of the educational setting: (1) the social relationships that occur between students and teachers within the classroom, (2) the curricular content that is covered and its level of coherence and cognitive demand, (3) the specific pedagogical approaches the teacher uses (for example, the grouping techniques used, the types of questions asked, and the explicitness of the instruction in a particular content area), and (4) the level of order and organization in the classroom.

Social relationships include such things as the warmth and enthusiasm a teacher displays toward students, the way in which she deals with disciplinary issues that arise, and the way in which she mediates social interactions among students within the classroom. Researchers in preschool and early elementary school have been particularly interested in such aspects of classroom practice and have shown many of these characteristics, including teacher-child closeness and teacher warmth, to be correlated with academic achievement.[9]

Another area of interest to many researchers is the actual curricular content that is covered (for example, the amount of time devoted to reading instruction or to instruction in particular mathematical topics) and the level of coherence and cognitive demand embedded in that content. It is not surprising that the amount of exposure that students have to particular content has been linked to student achievement in a variety of contexts and subject areas.[10] In addition, instruction that is coherent rather than fragmented and content that is taught with a higher degree of cognitive demand or complexity have also been correlated with higher levels of student achievement.[11]

In the areas of reading and mathematics, there has been a great deal of interest in the pedagogical techniques teachers use. The National Reading Panel, for example, highlights the specific practices (grouping, explicitness of instruction, types of practice activities for students) that have been shown to be most effective in

improving students' reading achievement in different aspects of reading instruction (decoding, comprehension, and vocabulary, for example).[12] In mathematics, researchers are interested in aspects of pedagogy such as the richness of representation and explanation and the connectedness of classroom tasks to mathematical principles.[13]

A teacher's ability to use appropriate strategies for maintaining behavioral and emotional order in the classroom is also seen by many as a key component of classroom practice since it is hypothesized that it is more difficult for students to learn in an environment that is not orderly and calm.[14]

Some researchers are interested in studying a narrow aspect of one of these four dimensions, and others are interested in measuring a number of different aspects to provide a more global assessment of the quality of classroom practices. The Trends in International Mathematics and Science Study (TIMSS) video study instrument combines measures of pedagogical practice (for example, a global measure of "making connections") with measures of content coverage and student engagement (for example, time on task).[15] The Classroom Assessment Scoring System (CLASS) instrument asks observers to code both social relationships (for example, levels of emotional support in the classroom) and pedagogical practices (for example, the extent to which teachers pose why or how problems; engage students in prediction or experimentation; ask students to brainstorm, plan, and produce; and connect concepts and real-world applications).[16]

The most appropriate approach for measuring classroom practice depends largely on which factors or combination of factors the researcher is most interested in. A variety of approaches can be taken, including the collection of artifacts such as lesson plans and assignments, curricular materials, exams or teacher portfolios, shadowing students, or analyzing student discourse. However, the two most frequently used approaches to measuring classroom practice are the use of direct classroom observation and annual teacher surveys. Next we outline the strengths and weaknesses of these two approaches and then discuss an alternative to them, the use of

instructional logs, that can help overcome some of the limitations of these approaches, particularly when attempting to measure curricular content and pedagogy.

Measuring instructional practice using direct classroom observations

The most common approach to measuring all four aspects of classroom practice has been to send trained observers into schools to collect structured observational data (or, more recently, to make video recordings of selected samples of instruction for later coding by experts). Direct observation of classroom practice is often seen as the gold standard for data collection in research on teaching, was used extensively in process-product research, and is increasingly used in studies of instructional improvement interventions in order to assess fidelity of implementation and the mediating effects of instruction on intervention outcomes. Classroom observations can and have been used effectively to measure all four aspects of classroom practice.

Despite their flexibility and face validity, in-person classroom observations (and video recording) have limitations. First, they are expensive. In-person observations often require highly skilled observers or extensive training of observation staff; having observers travel to various study sites can be costly, as can video recording personnel and equipment. As a result, their use in large-sample research has been limited.

Observations, like most other methods of data collection, including surveys and logs, are also somewhat limited in the types of interactions they can capture. The behaviors being measured need to be well defined and relatively low inference, so that observers can be trained to a high degree of reliability. For example, it is fairly straightforward to train observers to accurately record whether reading instruction occurred in the classroom, but it is more difficult to train observers to reliably record whether reading instruction occurred in an explicit manner. This is especially true when

trying to use classroom observations on a large scale. Although it may be possible to identify and train a small group of highly skilled observers to reliably recognize explicit instruction, it may be difficult to find enough observers with the appropriate background knowledge and skills to reliably conduct higher-inference observations in large numbers of classrooms.

The large degree of variability in teaching that occurs across a nine-month academic period also creates a problem of generalizability when using classroom observations. To the extent that teaching varies systematically across days of the school year, and especially if it involves many rare events, attempts to sample teaching activities adequately and reliably discriminate among different teachers' yearly patterns of instruction will require more in-class observations than all but the most well-funded studies can afford to collect. Similarly, it is difficult to illuminate the daily and yearly rhythms of instruction in schools or identify the presence or absence of important activities that occur relatively infrequently using observational data. Researchers who conduct in-person observations rarely (if ever) observe on days when no instruction occurs, even though data from the Study of Instructional Improvement (SII) (described in more detail below) show that 15 percent of all days include no reading and language arts instruction. Data from SII also show that instruction starts slowly, reaches a peak from November through April, and then drops off in the spring and that reading and language arts lessons are less likely to occur on Fridays and on days just before and after holidays. Such information is missed in most observational studies.

Measuring instructional practice using annual teacher surveys

In addition to conducting classroom observations, researchers often take a second approach to gathering data about classroom practices: the use of annual surveys. Many national and international education studies, for example, include teacher surveys with a small

number of items intended to measure constructs like curriculum coverage or content-specific teaching (examples are the Early Childhood Longitudinal Study and the National Assessment of Educational Progress). Data from one-time surveys are less expensive to collect than observation or videotaped data, and thus surveys are well suited to large-sample research.

Although surveys are less costly, they are limited in the types of classroom practices they can measure. While in-person observations can capture the nature of the social relationship between teacher and student and the level of organization and order in the classroom, teacher surveys are effective primarily in trying to measure content coverage and pedagogy. Even in this arena, surveys have been shown to be limited in their ability to describe the quality of pedagogical practices and in the complexity of instructional practice they can capture, although they are generally reliable and valid for broad measures of instruction.[17]

Classroom observations can introduce error into the measurement due to inconsistency among the raters. Annual surveys require teachers to answer questions about their own behavior and as a result can also introduce measurement error. Teachers may provide socially desirable responses, what teachers believe and report that they are doing in their classroom may be different from what they are actually doing, or error can be introduced if the terminology used in a survey is interpreted differently by different teachers.[18]

Classroom observations also suffer in their ability to generalize and capture the daily rhythms of instructional practice. Single-administration questionnaires often try to overcome this problem by asking respondents to make retrospective self-reports of activities across the school year. However, questionnaires that ask respondents to report on activities that transpired over relatively long periods of time suffer from two other problems that can result in measurement error. First, when the period over which reporting occurs is long, respondents can easily forget the behaviors in which they engaged. Second, when memory is fuzzy, respondents resort to estimating the frequencies of their behav-

iors. However, respondents use different estimation strategies, and as a result, two respondents with the same pattern of behavior often make very different retrospective reports. In fact, estimation has been found to be especially inaccurate in retrospective self-reports of two kinds of behaviors: those that rarely occur and those that occur frequently.[19]

Measuring instructional practice using instructional logs

One way to overcome some of the limitations of both classroom observations and annual teacher surveys is the use of instructional logs (or time diaries). Instructional logs are less expensive than observations, provide a solution to the problem of generalizability, and overcome some (although not all) of the issues of measurement error inherent in one-time surveys. Interest in the use of instructional logs to gather data on classroom instruction began in the 1980s. Instructional logs were first used in such studies as the Beginning Teacher Evaluation Study and in research conducted at the Institute for Research on Teaching at Michigan State University.[20] Like annual surveys, logs ask teachers to self-report on their own instructional practices, using a paper-and-pencil or Web-based format, but they differ from annual surveys in two distinct ways. First, they ask only about very specific aspects of teacher practice, while teacher surveys often ask a broader range of questions about such topics as the school environment, job satisfaction, or sense of self-efficacy. Second, they are administered more frequently than annual surveys. Logs are significantly less costly than classroom observations to administer, and as we discuss in more detail below, because they are administered more frequently than either annual surveys or observation of classroom practice, they can overcome many of the technical problems associated with these two more widely used measurement techniques.

Instructional logs are similar to teacher surveys in many respects, and instructional logs also suffer from many of the same limitations. Like annual surveys, logs are limited in the types of information they

can gather about classroom processes. They are not an effective tool for measuring social interactions and are limited in their ability to measure aspects of classroom organization and order. They also suffer from some of the same self-report biases as surveys. Teachers may provide socially desirable responses on their logs, and therefore logs should be used with particular caution in a high-stakes environment. In addition, they are effective in measuring instructional practice only if they are actually completed by most of the teachers in the sample; most of the time, therefore, as with annual surveys, researchers must take care to ensure high responses rates. Finally, while frequent administration of logs can decrease coverage error and increase generalizability, these benefits come at the cost of increased respondent burden, which can increase nonresponse or lead to response bias if respondents develop time-saving (but inaccurate) patterns of filling out log surveys.

However, research conducted as a part of the SII has shown that when instructional logs are implemented properly, they can be an extremely powerful tool for measuring certain aspects of classroom practice in a valid, reliable, and cost-effective manner. In the next sections, we describe how researchers structured instructional logs in SII to measure one particular aspect of instructional practice: the enacted curriculum. We then discuss the procedures researchers in SII used to overcome some of the self-report biases and response rate issues that are inherent in this type of data collection (these same techniques could also help to improve the quality of data obtained from annual surveys). Finally, we compare some of the technical properties of logs relative to annual surveys and classroom observations and draw some conclusions about the implications of the findings from SII for study design.

Structuring logs to measure the enacted curriculum

SII was a large quasi-experiment that investigated the extent to which three widely disseminated comprehensive school reform programs (the Accelerated Schools Project, America's Choice, and Success for All) changed reading instruction and improved students' reading achievement in high-poverty elementary schools. One major chal-

lenge for SII researchers was to develop measures of instructional practice since, in the logic model guiding the study, instruction was seen as a key element explaining program effectiveness. The conceptual framework that the SII researchers chose to guide the measurement of instruction was based on the concept of the enacted curriculum, which refers to the amount of exposure to specific academic content that a student receives, as well as the content-specific teaching practices the teacher uses to teach that content, including, for example, the nature and cognitive demand of students' reading tasks or the explicitness of instruction in a particular content area.[21] From this perspective, instruction is conceptualized as a series of repeated (daily) exposures to instruction, and the key measurement problem is to obtain an estimate of the overall amount or rate of exposure to particular elements of instruction occurring over some fixed interval of time, such as a school year.

Since surveys that ask teachers to retrospectively report on their activities over the course of an entire school year suffer from reporting biases, and because intermittent classroom observations are limited in their ability to generalize across the entire school year, SII researchers chose to use instructional logs to measure the rate of exposure to the enacted curriculum. The log instrument that was created to measure the enacted curriculum in reading and language arts for SII was a self-administered, paper-and-pencil questionnaire containing over a hundred items, mostly in checklist format.[22] Prior research suggests that curriculum coverage differs across students in the same classroom. In particular, a few studies of early reading achievement showed that when teachers used in-class grouping arrangements to teach early reading skills, patterns of curriculum coverage varied more widely among students within than across classrooms.[23] The SII logs therefore asked teachers to report on the reading and language arts instruction provided to a single student on a single day. To ensure an accurate record of teachers' overall patterns of teaching across a given year, teachers rotated log reports across a representative sample of eight students in their classes during three extended logging periods spaced evenly across the academic year.

The SII log gathered data on several dimensions of instruction. The opening section (gateway items) asked teachers to report on the amount of time spent by a focal student on reading and language arts instruction on the reporting day, as well as the amount of emphasis given to each of the following reading and language arts topics: word analysis, concepts of print, oral or reading comprehension, vocabulary, writing, grammar, spelling, and research strategies. If teachers indicated that word analysis, comprehension, or writing was taught to the focal student, they completed additional items (in the "back end" of the log) about the specific content objectives that were taught to the student in that domain, the methods used to teach that content, and the tasks and materials the focal student engaged with that day.

Implementing logs in ways that improve accuracy and response rates

Instructional logs can suffer from many of the same limitations as annual teacher questionnaires with regard to accuracy and nonresponse. In developing the reading and language arts logs for SII, therefore, researchers convened an expert panel of reading researchers and solicited their opinions in order to construct items that were free of response bias. In addition, field staff from the University of Michigan's Survey Research Center carefully designed field procedures to improve the accuracy of teacher responses to log items and increase teacher response rates in SII. To improve reporting accuracy, field staff conducted a forty-five-minute training session for teachers before their first logging period. This session introduced teachers to the definition of terms used in the log and showed them how to complete the log questionnaire. That was followed by a suggested two-hour home study period, during which teachers were asked to study a glossary defining and illustrating the terms used in the log, and then by a one-hour in-school follow-up training session prior to the first logging period. Once logging began, teachers could call a toll-free phone number or ask local field staff to address any difficulties they were having with logging. The accuracy of responses obtained using these procedures appears to be quite good.

Also, an incentives plan was developed to increase teacher response rates. Using data on teachers' salaries nationwide, researchers calculated the average daily wage of teachers and then offered payments to teachers based on the expectation that a single log would take about five minutes to complete. Rather than use piece-rate incentives, teachers were paid at the end of each logging period in which they logged. The incentive was $150 per six-week logging period if a teacher was logging for the full eight students called for by the original design. In addition, field staff provided logging teachers with small gifts (coffee mugs, paper weights, pencils or pens) on a variable interval reinforcement schedule to motivate log completion further.

Overall, response rates for the reading and language arts log were quite high. About 90 percent of the teachers asked to log did so, and they completed 90 percent of the logs they were administered. Although teachers often struggled when they began logging, after about a week, they typically completed the reading and language arts log in about five minutes time. Moreover, the response rates demonstrate (and teacher comments suggest) that logging was not perceived as overly burdensome.

Data based on initial budgets and log administration data suggest a research cost of about $27.50 per log administered, which encompasses the cost of developing, printing, and distributing the logs; training the teachers; monitoring completion rates; and providing incentives. Although this is more expensive than administering a one-time annual survey to teachers, primarily due to the cost of the incentives, it is far less than the cost of conducting a single classroom observation, which requires hiring, training, and sending an observer to a classroom, as well as the costs associated with developing, printing, and distributing the observation protocol itself. Incentives are often provided to teachers who participate in classroom observations as well, though these are generally less than the incentives provided to teachers in SII. It is also less expensive than a videotaping session, which requires renting or purchasing equipment and sending trained video personnel into the classroom. The largest costs associated with the logs are the incentives provided to

teachers for completing the logs. Thus, costs increase as the number of teachers increases.

The validity of log data

The results demonstrate that it is possible to administer instructional logs to large samples of teachers over a prolonged period of time, at a reasonable cost, and to achieve high response rates in doing so. However, an important question is whether the data gathered from logs accurately capture the aspects of teachers' instructional practice that are of interest. Descriptive information about instructional practice obtained from the SII logs provides some evidence of their validity. For example, activities such as word analysis were more likely to be reported by teachers in the lower grades, while teachers in the upper grades reported more use of chapter books. Log data were also used successfully to discriminate among the three comprehensive school reform programs that were part of the study.

Another way to assess the validity of instructional logs is to compare findings from logs to direct observation of classroom practice. Prior research has suggested acceptable convergence between observers' and teachers' reports of curriculum coverage for a given lesson.[24] Consistent with these findings, SII data indicate that log data are only slightly less accurate than comparable data collected by trained observers. Camburn and Barnes conducted a study comparing data from teacher and observer log reports for the same lesson from thirty-one teachers working at various grades who completed logs over a three-month interval as part of SII. On one of these logging days, two trained observers from SII entered the classroom and completed separate logs to measure the same instructional events being logged by a teacher. Camburn and Barnes reported that the rates at which teachers and observers agreed on their log responses for a given day were only slightly lower than the rates at which the two observers agreed on their logs for that same day. Indeed, for both gateway and back-end items on the log, teacher-observer match rates were quite acceptable, ranging from 81 to 90 percent agreement on gateway items and averaging 73 percent across all back-end items.[25]

Interrater agreement rates between teachers and observers vary across items in the log instrument. In particular, Camburn and Barnes reported that interrater agreement was greater when the focus of measurement was on major curricular topics such as reading comprehension versus more discrete topics in a curricular area, such as particular reading comprehension strategies. It was also greater when measurement focused on frequently occurring topics or practices as opposed to topics or practices that were rare.[26] All of this suggests that logs are a valid and cost-effective way to collect data on curriculum coverage and pedagogy in the classroom. More research needs to be done to determine whether log data provide valid data on other aspects of classroom practice.

Camburn and Han also compared teachers' log responses for a given year to the same teachers' reports of instruction on the end-of-year teacher questionnaire. The purpose was to estimate the amount (and sources) of divergence in teachers' responses to comparable items across the log and questionnaire instruments. The study found that teachers uniformly reported higher frequencies for teaching practices on the annual questionnaires than they did on the log, with a general tendency for overreporting in the area of word analysis and for more overreporting by female teachers, African American teachers, more experienced teachers, and teachers who individualized instruction. These latter findings constitute a particular form of observer bias (bias due to the characteristics of the survey respondent).[27]

Number of logs required

Data obtained from the SII also provide valuable information about how many log data need to be collected in order to obtain reliable estimates of instructional practice. The ability to discriminate reliably among teachers' instructional practice when instructional log measures are taken repeatedly depends on the internal consistency of the instructional logs, the true variability in teachers' instructional practice, and the number of occasions in which measures are taken. In general, the ability to discriminate across teachers increases as the number of occasions of measurement increases (in this case, the

number of logs). However, if the degree of variability among teachers in their instructional practice is large but there is little variability across measurement occasions, then relatively few logs are needed to discriminate reliably among teachers. But if differences among teachers are small or there is a great deal of variability across measurement occasions, relatively more logs are needed.

Data from SII indicate that roughly 72 percent of the variance in instructional time spent on reading and language arts activities lies among measurement occasions, about 23 percent lies among teachers within schools and about 5 percent among schools. This implies that a fairly large number of logs are needed to reliably discriminate among teachers in content coverage and pedagogy. Specifically, SII data indicated that about twenty logs per year from teachers are necessary if the measurement goal is to reliably discriminate among teachers with respect to the extent to which they use teaching practices measured by the SII log. This underscores the advantage of collecting log data versus data from in-person observations since it would be exceedingly expensive to collect twenty in-person observations per teacher in large-sample studies. Note that different numbers of logs may be required depending on the outcome of interest. Similarly, if the goal is to reliably distinguish among schools rather than among teachers, reliability will be influenced by both the number of logs per teacher and the number of teachers sampled from each school.

Conclusion

As we continue to strive toward identifying quality educational settings for children, it is clear that we need better, more cost-effective tools for assessing the quality of teacher-student interactions. Different tools will be appropriate for measuring different aspects of classroom practices and processes. In the past, researchers have relied on in-person classroom observations and annual teacher surveys to obtain information about content coverage and instructional practice. However, both tools suffer from

some significant limitations. Classroom observations are expensive, require extensive training of raters to ensure consistency in the observations, and are difficult to use to generalize across an entire school year and because of their expense. Annual surveys are less expensive but often suffer from self-report bias and the bias that results when teachers are asked to retrospectively report on their activities over the course of a single year. Instructional logs offer a valid, reliable, and relatively cost-effective alternative for collecting detailed information about classroom practice and can overcome some of the limitations of both observations and annual surveys. Analyses from the SII indicate that when teachers are provided with effective training and appropriate incentives, logs can provide information on content coverage and pedagogy that is more accurate than one-time teacher questionnaires and only slightly less accurate than information from direct classroom observation. For most study purposes, administering around twenty logs (evenly spaced over the academic year) should allow researchers to reliably discriminate instructional practices in the area of reading and language arts across teachers and schools. To date, instructional logs have received limited attention as a tool for collecting data on classroom practices, and we encourage researchers interested in measuring instructional content and pedagogy to pay greater attention to the benefits of these instruments for collecting useful information about what takes place in the classroom.

Notes

1. Hanushek, E. A. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, *19*(2), 141–164.

2. Hoxby, C. M. (2003). *The economics of school choice*. Chicago: University of Chicago Press.

3. Cook, T. D., Hunt, H. D., & Murphy, R. F. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*(2), 535–597.

4. Hanushek, E., Kain, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). The market for teachers. NBER working paper 11154. Cambridge, MA: National Bureau of Economic Research.

5. Hanushek. (1997).

6. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools and academic achievement. *Econometrica*, *79*, 418–458.

7. Meyer, J. W., & Rowan, B. (1978). The structure of educational organizations. In M. W. Meyer & Associates, *Organizations and environments.* San Francisco: Jossey-Bass.

8. Rogosa D., Floden R., & Willet, J. B. (1984). Assessing the stability of teacher behavior. *Journal of Educational Psychology,* *76*(6), 1000–1027.

9. Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal,* *45,* 365–397.

10. Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. L., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms.* Madison: University of Wisconsin, Wisconsin Center for Educational Research.

11. Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading high school mathematics instruction: Improving learning opportunities for low-achieving, low-income youth. *Educational Evaluation and Policy Analysis,* *19*(4), 325–338.

12. National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: U.S. Government Printing Office.

13. Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction,* *26*(4), 430–511.

14. Raver, C., Garner, P., & Smith-Donald, R. (2007). The roles of emotion regulation and emotion knowledge for children's academic readiness: Are the links causal? In B. Pianta, K. Snow, & M. Cox (Eds.), *Kindergarten transition and early school success.* Baltimore, MD: Brookes Publishing.

15. National Center for Education Statistics. (2004). *Highlights from the trends in international math and science study: TIMSS 2003.* Washington, DC: Author.

16. Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system, manual, K–3.* Baltimore, MD: Brookes Publishing.

17. Burstein, L., McDonnell, L., Van Winkle, J., Ormseth, T., Mirocha, J., & Guiton, G. (1995). *Validating national curriculum indicators.* Santa Monica, CA: RAND.

18. Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis,* *12,* 311–329.

19. Sudman, S., & Bradburn, N. M. (1982). *Asking questions: A practical guide to questionnaire design.* San Francisco: Jossey-Bass.

20. Fisher, C., Filby, N., Marliave, R., Cahen, L., Dishaw, M., Moore, J., et al. (1978). *Teaching behaviors, academic learning time, and student achievement (phase III-B, final report): Beginning teacher evaluation study.* San Francisco: Far West Laboratory for Educational Research and Development; Smithson, J. L., & Porter, A. C. (1994). *Measuring classroom practice: Lessons learned from efforts to describe the enacted curriculum—The Reform Up Close Study.* New Brunswick, NJ: Consortium for Policy Research in Education.

21. Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum in large-scale surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal*, *105*, 75–102.

22. Throughout the article, we focus solely on the reading/language arts log developed as part of the SII. Readers interested in the mathematics log developed for this study can consult Rowan, Harrison, & Hayes (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal*, *105*, 103–127.

23. Barr, R., & Dreeben, R. (1983). *How schools work.* Chicago: University of Chicago Press.

24. Burstein et al. (1995).

25. Camburn, E., & Barnes, C. A. (2004). Assessing the validity of a language arts instruction log through triangulation. *Elementary School Journal*, *105*(1), 49–74.

26. Camburn & Barnes. (2004).

27. Camburn, E., & Han, S. W. (2006). *Factors affecting the validity of teachers' reports of instructional practice on annual surveys.* Madison: Wisconsin Center for Education Research.

BRIAN ROWAN *is the Burke A. Hinsdale Collegiate Professor in Education and a research professor in the Institute for Social Research at the University of Michigan.*

ROBIN JACOB *is an assistant research scientist in the School of Education and the Institute for Social Research at the University of Michigan.*

RICHARD CORRENTI *is an assistant professor at the University of Pittsburgh and a research scientist at the Learning Research and Development Center.*