



The purpose of this Learning Policy Center brief is to demonstrate the need for measuring instruction as part of the normal, on-going professional practice of schools and school districts and to provide guidelines for how instruction can be measured most productively.

KEY POINTS

Students' test scores alone do not provide information about how districts can improve teaching and learning, whereas the measurement of instruction does.

PAGE 1

Measuring instruction focuses attention on the quality of teaching, which research indicates is the most important factor within the control of schools that influences student learning.

PAGE 2

Measuring instruction can also build a professional culture around a common vision of excellent practice within a district.

PAGE 2

To avoid the potential pitfalls of measuring instruction, districts should minimize the stakes attached to outcomes of instructional measurement and integrate measures of instruction into their professional development system.

PAGE 3

Setting the Groundwork for Quality Teaching and Learning: Measuring Instruction in Schools and Districts

by Lindsay Clare Matsumura

The quality of teaching is the most important factor within the control of schools that influences student learning.¹ Despite this fact, the quality of schools generally is assessed by student outcomes only – standardized test scores, graduation rates, absenteeism, etc. Teaching quality is notably absent in the public dialogue about school quality and in school districts' own monitoring of their success.

In this brief I argue that the quality of instruction should be measured as part of the normal on-going professional practice of schools and school districts. By measuring instruction, I mean systematically collecting information about the subject-matter content that is covered in classrooms and how that content is delivered to students – the instructional core.² More specifically, I argue that measuring instruction is important for informing district decision-making, directing attention to the quality of instruction as an important end goal of reform efforts, and building a school district's professional learning community. Some potential problems that could arise as a result of measuring instruction in districts, and suggestions for ameliorating these potential problems also are discussed.

Informing District Decision-Making

School districts across the country devote considerable resources to improve the quality of instruction and student learning in the form of professional development programs, curricula, and assessment systems. The information most readily available to districts and schools regarding how well their efforts are working, however, is only student outcomes on standardized achievement tests. This is problematic because students' test scores, on their own, provide little information to districts about how programs and policies intended to improve instruction are implemented and interpreted by teachers. In the absence of measuring instruction, district leaders cannot be sure if change, or lack of change, in students' test scores is due to:

- Lack of implementation – The reform strategy is not making its way to schools and classrooms. For example, a district decides to adopt the inquiry-based FOSS science curriculum, but the materials (e.g., FOSS kits) are not distributed to schools in a timely fashion. Or teachers have access to the materials but are not comfortable using them and continue teaching science in the way they had previously.



- **Incomplete implementation** – The reform strategy is being applied in schools, but not in a way that was intended by the developers of the reform. For example, a district adopts a new mathematics curriculum, but the enactment of the curricula resembles “old” ways of teaching, or teachers do not follow the sequence of lessons in their intended order.
- **Ineffectiveness of a reform strategy** – The reform strategy chosen by the district is being implemented “as intended,” but it is not a good strategy. For example, a district adopts a new language arts curriculum that is poorly designed, or no better than the previous curriculum that was in place.

“students’ test scores, on their own, provide little information to districts about how programs and policies intended to improve instruction are implemented and interpreted by teachers”

Each of these reasons for why a reform may not be having its intended effect likely would entail a different course of action on the part of district leadership. For example, districts could choose to provide greater incentives for teachers to implement reforms when they are not doing so; higher-quality professional development could be a path of action if reform is not being implemented in its intended form; or districts could

take up a new reform if the present reform is ineffective. Measuring instruction thus is important for monitoring the progress of reforms and helping districts make more informed decisions about the “next steps” needed in their reform strategy.

Ambitious Instruction as the Goal of Reform: Directing Attention to Instructional Quality

Measuring instruction also is important for directing attention to the quality of teaching – the features of instruction that make a difference in student learning. A growing body of research indicates that students’ scores on state standardized tests are, on their own, poor indicators of whether reform goals for instructional improvement are being met.³ This is the case for two key reasons. First, standardized tests are not always aligned with the goals of ambitious educational reforms. Students’ mastery of many important academic skills (e.g., the ability to write a research report, use multiple representations in solving a mathematics problem, conduct a scientific experiment) is difficult and expensive to assess in large numbers of classrooms. A state achievement test might not adequately measure the range of skills a district wants its students to master and that are critical to students’ academic success.

Second, teachers who are under pressure to increase test scores can narrow what they teach to the content represented in a test, and shape

classroom activities and discussion to more closely mirror a test’s question formats. Recent research indicates that excessive focus on preparing students to pass standardized tests is deleterious to student learning.⁴ When instruction focuses on the content represented in achievement tests, students may no longer have an opportunity to learn the span of knowledge and skills necessary to master grade-level content. In these instances, students may be rated “proficient” on their state’s achievement test without a corresponding increase in the knowledge and skills necessary for success in the upper levels of schooling and beyond.⁵ Focusing attention on excellent instruction as an additional indicator of school quality can serve as a counter-balance to the pressure teachers feel to teach directly to the requirements of their state’s achievement test. As such, it can mitigate some of the harmful effects of high-stakes student testing on instruction.

Building a Professional Learning Community across the Levels of the School System

A great deal has been written about the isolation of teachers in classrooms and need for the creation of learning communities in schools. Teachers now are encouraged to make their “practice public” and collaborate with their colleagues to plan for, and reflect on, instruction.⁶ Principals, in turn, are encouraged to act as instructional leaders and take an active role in this process. Increasingly it is understood, however,



that the quality of teaching is influenced by systems at both the school and district-level.⁷ For substantive change to occur in classrooms, a district needs to become a learning community. Practice needs to become public *across* schools, and district administrators need to be leaders in the process of reflecting on instruction.⁸

“Focusing attention on excellent instruction as an additional indicator of school quality can serve as a counterbalance to the pressure teachers feel to teach directly to the requirements of their state’s achievement test.”

Measuring instruction can help build this professional culture within a district. Assessments of all types communicate a vision of what is important to know and be able to do. Standards for instruction adopted by most states provide little guidance for instruction, mostly because the terms used to describe reformed practice are subject to multiple interpretations. Teachers, principals, district administrators, and educational researchers can, and frequently do, disagree about what it means to “hold a discussion” or apply “critical thinking skills.” Moreover, little opportunity exists for district administrators, principals, and teachers to develop consensus about what good teaching should look like. This is the case even when a district has adopted a common curriculum.⁹

Measures of instructional quality can

help create a common vision of excellent practice. And working toward a common vision for instruction likely would further reform goals. Benchmarks tied to these assessments (e.g., models of lesson plans, assignments and the criteria used to grade students work, excellent classroom discussions) could be useful for focusing professional development efforts, encouraging reflection on instructional practice across the levels of the school system, and supporting principals and district administrators in their roles as instructional leaders.

What Districts Should Keep in Mind When Measuring Instruction

Although measuring instruction has many potential benefits, less desirable outcomes are possible as well. Some strategies for addressing and avoiding negative outcomes as a result of directly measuring instruction are considered in the following sections.

Minimize stakes attached to the outcomes. One potential negative consequence of attempting to measure instruction in districts is score inflation - the possibility that teachers will represent their instruction on the assessment in a way that achieves high scores but is not indicative of their “true” instructional practice. Under these circumstances, measures of instruction can yield higher scores over time without improvement in the quality of instruction – students’ op-

portunity to learn. Score inflation is especially likely to occur if high stakes are attached to the outcomes of an assessment, for example, if the purpose of measuring instruction in districts is to evaluate teachers for advancement. For this reason, it would be important to make sure that the stakes attached to any measure of instruction were as minimal as possible.

“Measures of instructional quality can help create a common vision of excellent practice. And working toward a common vision for instruction likely would further reform goals.”

Minimizing any stakes attached to the outcomes also would be important for reducing the possibility that measures of instruction could have a negative influence on instruction. As described earlier, studies of the influence of high-stakes testing on instruction indicate that teachers, under pressure to improve students’ scores, can narrow the focus of their instruction to the content and format of a test. Just as standardized tests of student learning cannot measure all of the content and skills students need to be academically successful, measures of instruction also cannot measure the full range of skills teachers need to teach effectively. Minimizing the stakes attached to a measure of instruction would help ensure that an overly narrow view of instruction is not being unwittingly fostered in schools.



Additionally, to obtain a reliable indicator of an individual teacher's instruction requires many more data points than a district likely could afford to collect and rate (for a practice-based assessment at least). In other words, one would need to collect multiple assignments with student work or observe multiple lessons to be confident that what has been seen or collected is characteristic of a teacher's instruction. It would not be appropriate to attach stakes for teachers to outcomes that are not reliable at the individual teacher level.

Integrate measures of instruction into the professional development system. In her seminal article on classroom assessment, Lorrie Sheppard (2000)¹⁰ noted the discontinuity between goals for instruction and the assessments used to monitor student learning. In her words:

...if instructional goals include developing students' metacognitive abilities, fostering important dispositions, and socializing students into the discourse and practices of academic disciplines, then it is essential that classroom routines and corresponding assessments reflect these goals as well. This means expanding the armamentarium for data gathering to include observations, clinical interviews, reflective journals, projects, demonstrations, collections of student work, and students' self-evaluations (p. 8).

She further noted that for assessments to support student learning, classroom practice would need to significantly

change to resemble a "learning culture." Assessing students would need to be undertaken as a joint process of investigation between students and teachers, not an evaluative exercise or something to be "gotten through" separate from the process of learning.

"For measures of instruction to serve as a tool for learning, the assessment process would need to be integrated into the professional learning system of a school and district, and teachers would need to be engaged as partners in the assessment process"

Similar arguments could be made in terms of assessing instructional practice. Measures of teaching would need to match goals for instructional improvement. For example, if the reform goal is to improve classroom discussions about mathematics, then observations of classroom discussion would be needed to measure instruction. Or if the goal is for students to engage in more intellectually challenging work generally, then samples of assignment tasks with student work and the criteria used to grade students' work likely would be the best way to investigate this dimension of teaching practice.

The context in which teachers are assessed also would need to change to resemble a "learning culture." In many districts teacher assessment is an evaluative procedure carried out by principals and often is little more than a *pro forma* exercise or an opportuni-

Summary Policy Recommendations

- Place a priority on measuring instruction to inform district decision-making and improve the quality of teaching and learning
- Develop measures of instruction that create a common vision of excellent practice, and encourage reflection about that vision within schools and districts in order to build a professional culture
- Attach minimal stakes to the outcomes of instructional measurement - stakes can lead to score inflation and negatively impact instruction.
- Integrate measures of instruction into the professional development system by aligning those measures with instructional improvement goals, and reflect with teachers about those goals and their measurement.
- To reduce costs, measure instruction annually among a random sub-sample of teachers or teachers within one grade level; integrate formation of measures into teachers' professional development; and/or involve university partners.



ty to weed out the most incompetent teachers.¹¹ For measures of instruction to serve as a tool for learning, the assessment process would need to be integrated into the professional learning system of a school and district, and teachers would need to be engaged as partners in the assessment process. Assessment would need to be an opportunity for collaborative reflection on instruction, and school and district leaders would need to be co-accountable for teachers' development – that is, for providing professional development resources targeted to teachers' needs.

Reduce costs by selectively sampling schools and teachers. It is expensive to assess instruction using performance-based measures such as observations and samples of classroom artifacts described earlier. Other options for measuring instruction exist as well (e.g., instructional logs, surveys) that are less expensive but also may provide less grist for reflection and professional learning (that is, for establishing consensus around what intellectually challenging work for a specific grade and content area should “look like.”). The feasibility of measuring instruction in school districts is thus a real concern.

One way to reduce the burden on districts would be to collect fewer data points from teachers and look at the quality of instruction at a school level or within a grade level across schools. Alternatively, if a district is very large, all teachers at all schools might participate in the assessment as part of their

professional work and reflection (e.g., complete portfolios of their practice that might include videotaping lessons, or submitting assignments with student work), but only a sub-sample of teachers would be randomly sampled each year for independent evaluation at the district level. District administrators (e.g., Directors of Curriculum), principals, instructional coaches and teachers all could participate in assessing the quality of these artifacts. This could be done as a service (i.e. raters are elected by schools and the positions are rotated every few years), or for credit toward advancement or professional development hours.

University researchers can be a great help to districts by providing access to analytic expertise. Relying solely on a university partner to measure instruction, however, would not build capacity within districts and schools to self-assess and monitor the quality of teaching. Moreover, researchers from universities might not have the same level of credibility with teachers and their representatives that other teachers have. Universities can be useful partners, but districts and schools would need to “own” the assessment process if it is to be useful for creating consensus about good instruction within the different content areas and for serving as a tool for professional learning.

Summary

In summary, measuring instruction in districts has many potential benefits, including improved decision-making and planning, increased attention on

excellent instructional practice as an important outcome of education reform (not only students' standardized test scores), and a shared understanding of what “good” instruction looks like across the levels of a school system. Risks are present as well, though the potentially negative consequences of measuring instruction could be minimized by not attaching stakes to the outcomes and by incorporating the assessment process into a district's professional learning system. Measuring instruction could be expensive. The cost of not measuring instruction, however, of not making excellent practice transparent and of continuing to make students' scores on standardized tests of achievement the principal end of reform efforts, may be too high to pay.



References:

- ¹ Rowan, B., Correnti, R., & Miller, R.J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects Study of Elementary Schools. *Teachers College Record*, 104(8), 1525–1567.
- ² Ball, D.L., & Forzani, F.M. (2007). What makes education research “educational”? *Educational Researcher*, 36(9), 529-540.
- ³ Koretz, D., & Hamilton, L. (2006). Testing for accountability in K-12. In R.L. Brennan (Ed.), *Educational Measurement (4th edition)*, (p. 531-542). Westport, CT: American Council on Education/Praeger.
- ⁴ Allensworth, E., Correa, M., & Ponisciak, S. (May, 2008). *From high school to the future: ACT preparation – too much, too late*. Consortium on Chicago School Research, University of Chicago.
- ⁵ Linn, R.L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.
- ⁶ Talbert, J., & McLaughlin, M. (2006). *Building school-based teacher learning communities: Professional strategies to improve student achievement*. New York, NY: Teachers College Press.
- ⁷ Glennan, T.K., Jr. & Resnick, L. (2004). School districts as learning organizations: A strategy for scaling education reform. In T.K. Glennan, Jr., S.J. Bodilly, J. Galegher, and K. Kerr, (Eds.) *Expanding the reach of education reforms: Collected essays by leaders in the scale-up of educational interventions*. Santa Monica, CA: RAND.
- ⁸ Thank you to Derrick Lopez for this analogy.
- ⁹ Rowan, B. & Camburn, E. & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal*, 105, 75-102.
- ¹⁰ Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14
- ¹¹ Stein, M.K., & Matsumura, L.C. (in press). Measuring instruction for teacher learning. In D. Gitomer (Ed.), *Measurement issues and the assessment of teacher quality*. Sage Publications.

The mission of the **Learning Policy Center** is to foster high quality learning environments for both students and professionals in public schools. Toward that end, we aim to infuse into policy decisions high quality, timely research on effective teaching and learning and on school, district, and policy conditions that support their improvement. **The Learning Policy Center** utilizes the rich talent pool of the University of Pittsburgh School of Education, the Learning Research and Development Center, the Institute for Learning and other regional assets to connect high quality learning research with education policy decision-makers.

Lindsay Clare Matsumura is an assistant professor in Learning Sciences and Policy at the University of Pittsburgh. You can contact her at: lclare@pitt.edu