UCLA IMPACT: Integrating Research and Practice Using Multiple Measures

Authors: Jon Schweig, Mollie Appelgate, and Karen Hunter Quartz
Date: July 2010
Series Number: IPMS023-X045-2010

Center X is part of the Graduate School of Education and Information Studies of UCLA
1320 Moore Hall Box 951521
Los Angeles, CA 90095
www.centerx.gseis.ucla.edu

ABSTRACT

Since 1995, UCLA's Center X has run an innovative teacher education program focused on providing high quality teachers for urban schools. Last year, UCLA added another pathway to its teacher preparation options, an Urban Teacher Residency program called IMPACT: Inspiring Minds through a Professional Alliance of Community Teachers. This program is built from a strategic partnership of UCLA's Teacher Education Program (TEP), the Los Angeles Small Schools Center and Los Angeles Unified School District (LAUSD) Local District 4. By creating this new approach to teacher learning within this partnership, IMPACT is a design experiment that seeks to integrate research and practice. In this overview paper, the authors briefly describe IMPACT and the efforts of their collaborative research team to study this program using multiple measures of teaching practice and use the findings for the purposes of evaluation, research, and data-driven school improvement.

UCLA IMPACT:
Integrating Research and Practice Using Multiple Measures

Jon Schweig, Mollie Appelgate, Karen Hunter Quartz

July 2010

Since 1995, UCLA's Center X has run an innovative teacher education program focused on providing high quality teachers for urban schools. Last month, UCLA added another pathway to its teacher preparation options, an Urban Teacher Residency program called **Inspiring Minds through a Professional Alliance of Community Teachers (**IMPACT). This program is built from a strategic partnership of UCLA's Teacher Education Program (TEP), the Los Angeles Small Schools Center and Los Angeles Unified School District (LAUSD) Local District 4. By creating this new approach to teacher learning within this partnership, IMPACT is a design experiment that seeks to integrate research and practice. In this overview, we briefly describe IMPACT and the efforts to date of our collaborative research team at UCLA to study this program using multiple measures of teaching practice and use the findings for the purposes of evaluation, research, and data-driven school improvement.

What is IMPACT?

Urban Teacher Residencies such as IMPACT have recently emerged as a third pathway option for individuals entering the teaching profession. They offer a response to a long-standing debate about the merits of traditional teacher preparation programs, which are often seen as too "distant" from the communities they serve, and the merits of alternate route certification programs, which are often seen as being "trials by fire" that cause many promising teachers to leave the field before establishing a career. UCLA's Center X is supported by the US Department of Education to develop and study the residency pathway in the high-need fields of

math, science, early childhood education, and special education; if it proves successful, it will replace the university's traditional program across all content areas.

The IMPACT program has many features that make it unique. The program design places apprentices into classrooms for an entire school year. The apprentices work through this extended period of clinical practice with an experienced mentor who has been selected and trained. Once apprentices have become full time teachers of record, the IMPACT program continues to offer induction support for two full years.

In addition, IMPACT is designed to integrate education theory and classroom practice together by designing programmatic experiences that provide opportunities for collaboration and partnership between practitioners, administrators, teacher educators, and university-based education researchers. These programmatic experiences have been carefully developed to maximize opportunities to improve apprentice and mentor teacher practice, improve the IMPACT program overall, and to advance theories about teaching and learning.

As the new research priorities outlined by the Institute of Educational Sciences (IES), recommend, educational researchers should be working closely with practitioners and policy makers in order to improve student learning. "IES believes that effective education research must be guided by the voice and interests of education practitioners and decision makers" (Easton, 2010, para. 10).

Consistent with these priorities, a research group consisting of teacher educators, researchers, statisticians and program administrators has begun work on a collaborative research project that is focused around the question, "How does this new approach to teacher education lead to increased teacher quality?" The central work of that inquiry focuses on selecting a set of instruments to collect data about teaching practice that have programmatic, research and administrative value. Understanding how these tools work together systematically, when best to use them, and what kinds of data are most helpful to stakeholders, ultimately, will maximize the opportunity to advance teacher practice and to build reflective and data-driven communities of practice in schools.

Complexity of Teaching Requires Varied Measures

The IMPACT program offers a rare opportunity to think systematically about collecting data about teacher practice in ways that are meaningful to the apprentices, teacher educators, administrators, and researchers. No single measurement tool – whether it be student test scores or principal observations (perhaps the two most commonly used tools in educational settings)–is adequate to capture all of the dimensions of teaching practice. Measuring complex things requires multiple sources of information. Multiple measures provide better, more in-depth information. As John and Soto (2007) point out, "because each data source has unique strengths and limitations, we should collect data from lots of different sources" (p. 480). They continue to say that "measures based on [other kinds] of data can help evaluate and provide evidence for the validity of more easily and commonly obtained self-report measures tapping the same construct" (p. 481). Using only one method to measure teacher practice would pose a threat to the validity of the inferences that could be drawn from the research.

With this in mind, a set of tools was carefully and deliberately chosen to measure multiple aspects of teaching. These tools will be integrated into the programmatic experiences of the apprentices and mentors, and will be embedded in their everyday experiences. All of the tools have been selected to have multiple purposes. For example, logs may serve as a source of data both for research and professional growth. As the program progresses, using multiple measures also provides the opportunity for additional research questions to evolve.

Six instruments were selected to capture information about teaching practice and quality. These instruments capture information in a variety of ways. Some data are self-reported; some are generated by external observers. Some are collected over short amounts of time, others over a period of days and weeks. Some are based on classroom-based activity, and some are based on work that is done outside of the classroom. Some are based on student work, and some are based on teacher plans. Some are longitudinal – looking at changes in teacher practice over a period of years. Because they capture multiple traits using multiple methods, each instrument gives information not only about what the teacher is doing, but also about how the other tools are working. The specific instruments include surveys, logs, observations, classroom artifacts (collected in an Instructional Quality Assessment (IQA)), Performance Assessment for California Teachers (PACT), and California Standards Test (CST) scores. We will collect these data from both apprentices and mentors, according to the following data collection plan:

Study Proposal				
Program Phase	**Entry**	**Year 1**	**Year 2**	**Year 3**
Calendar Year	**2009-2010**	**2010-2011**	**2011-2012**	**2012-2013**
Program Activities	Cohort selection, program begins June 2010	Residency and coursework (summer and school year)	Coursework (summer + seminar), Masters project (completed by Dec.), full-time teach of record, induction with BTSA provider	Full-time teacher of record, continued induction and BTSA support
Indicators ***Apprentice***	Surveys (program survey and MKT)	Surveys (program survey only) PACT Observations (including classroom video)	Surveys (program survey and MKT) CST scores Observations Retention data Artifacts (IQA) Logs	Surveys (program survey only) CST scores Observations Retention data Artifacts (IQA) Logs
Mentors	Surveys	Surveys Artifacts (IQA)		

Data collection will be most intense for the first cohort of IMPACT math, science, and early childhood apprentices because we have support to follow their progress through the end of the program and their third year of teaching. Data from subsequent cohorts will also be collected, in particular the special education cohort that does not start up until the summer of 2011. Following a description of each tool, we close this overview with a brief description of how these data will be used for the purposes of evaluation, research, and school improvement.

Our Set of Tools

1. *Surveys*

Surveys are perhaps the most administratively and economically feasible way to measure teacher practice. For information on some things, such as psychological traits like systems of beliefs or

efficacy, the self-reported information collected on surveys may be the only way to gather that data. There has been a rekindled interest, beginning in the early part of the 2000's, in exploring the relationship between conceptions of efficacy and a variety of outcomes, including student performance, teacher persistence and teacher instructional behavior (Tschannen-Moran & Woolfolk, 2001). As John and Soto (2007) point out, "the logic underlying [self-report] data is that individuals are in a good position to report about their psychological processes and characteristics—unlike an outside observer, they have access to their private thoughts and experiences and they can observe themselves over time and across situations" (p. 481). This type of data collection can be reliable. Koziol and Burns (1986) found, "agreements between teacher self-reports and observer ratings tended toward the .60s and .70s. Moreover, they rose to the high .70s and .80s when teachers used the report a second time and became more familiar with what was being asked" (as cited in Kennedy, 1999, p. 353).

We are administering two types of surveys. The first is an annual program survey designed to capture key background characteristics about the apprentices and mentors in order to understand their beliefs about teaching in high-poverty communities. The second type of survey is designed to measure apprentice's pedagogical content knowledge. For our mathematics apprentices, we will be surveying their mathematical knowledge for teaching (MKT). As developers and investigators of this survey, Hill, Rowan, and Ball (2005), define this knowledge as "explaining terms and concepts to students, interpreting students statements and solutions, judging and correcting textbook treatments of particular topics, using representations accurately in the classroom and providing students with examples of mathematical concepts and proofs (p. 373). Multiple-choice measures for mathematical knowledge for teaching were found by Hill,

Schilling and Ball (2004) to, "reliably discriminate among teachers and meet basic validity requirements for measuring teachers' mathematical knowledge for teaching" (as cited in Hill, Rowan & Ball, 2005). Additionally, teachers' scores were found to predict increases in student achievement.

Similar instruments for pedagogical content knowledge in science, early childhood education, and special education are currently being investigated.

2. *Logs*

Kennedy (1999) describes the log as "a paper-and-pencil form, filled out by teachers either daily or weekly, that describes the details of their curriculum for a specified period of time in a specified list of courses" (p. 349). Rowan and Correnti (2009) explain that in order to use the logs as a tool to measure instruction, that instruction has to be conceptualized as a, "series of repeated (i.e., 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 (e.g., a school year)" (p. 120).

There are a number of advantages to using logs to capture teacher practice. They can be used to understand patterns of bias in other forms of self-report data, like end-of-year teacher surveys. Additionally, logs have been demonstrated to match well with observer ratings of classroom practice. Rowan and Correnti (2009) report match rates between 81% and 90% on the gateway items (those about frequency of coverage), and averaging 73% across the other items (those having to do with cognitive demand). Also, compared to gathering information from

observations, logs are much less expensive. Finally, logs also offer the opportunity to study changes and variation in aspects of teacher practice at various points throughout the academic year. We are currently developing logs as both a programmatic and research tool.

3. Observations and Classroom Video

Observations, as a tool for gathering data, have been found to make important contributions both to accountability-driven measures of teacher quality, but also for purposes of professional development (Pianta and Hamre, 2009).

In their own investigations, Pianta and Hamre (2009) noted that:

> There is a reasonable body of evidence (see Gordon et al., 2008; Jones, Brown, & Aber, 2008; McCaslin et al., 2006), that teachers' performance in classrooms, in terms of their actual behavioral interactions with students, can be assessed observationally using standardized protocols, analyzed systematically with regard to various sources of error and in turn shown to be valid for predicting student learning gains. (p. 115)

Rowan and Correnti (2009) go so far as to say that sending trained observers into schools is "often seen as the gold standard for data collection in research on teaching" (p. 121).

Observations differ from both surveys and logs in that observations rely upon an external reporter and therefore do not have the issues of bias related to self-reporting. Unlike logs however, observations capture only one class on one day. We are currently developing an observation rubric to systematically collect and analyze observational data from both apprentices and mentors.

Classroom videos have been used since the 1960s to support teachers in understanding and developing instructional practices. Videos continue to be used as a means to engage teachers in conversations and reflection within teacher training and professional development (Putnam & Borko, 2000). However, more recently, classroom video has been used to measure teacher effectiveness (Spencer, Park & Santagata, 2010), as well as quality of instruction in mathematics as seen in the Third International Mathematics and Science Study (TIMSS) Video Study (Hiebert et al., 2003).

Video creates learning opportunities for teachers by situating the activities in classroom contexts (Borko, 2004; Putnam & Borko, 2000). As part of teacher learning, IMPACT apprentices and mentors will analyze classroom videos of others and of themselves throughout the program. Apprentices and mentors will use video to capture routines of practice around mathematics discourse and then analyze the video's content.

4. *Classroom Artifacts/Instructional Quality Assessment (IQA)*

When gathering artifacts, researchers typically ask teachers to "collect and annotate a set of materials such as classroom exercises, homework, quizzes, projects, exams, and samples of student work" (Borko, Stecher, Alonzo, Moncure & McClam, 2005, p. 76). In many ways, artifact collection is like "core sampling." This analogy points immediately to one of the relative advantages of artifact collection. It is less time consuming and less costly than classroom observation, particularly when employed across a large number of classrooms. It gives a way to understand and explore classroom practice at a distance.

Matsumura et al. (2006b) point out that artifacts and assignments provide insight on the "opportunity students have to produce high-quality written work on their own" (p. 7). They can also provide a further check on "the rigor and implementation of lesson activities, and application of classroom assessments" (p. 7). Classroom observation, for example, can only capture the parts of instruction that take place during the classroom period. However, outside-of-class work is also an important component of the way in which students experience instruction. Artifact collection has the potential to capture these dimensions, which may elude observational study. Our collection of artifact data will be structured by the Instructional Quality Assessment (IQA). This instrument, originally developed with protocols for assessing classroom instruction and teachers' assignments or artifacts, will be used in this research solely for its classroom artifact/assignment protocols and rubrics. In evaluating classroom assignments, the IQA, assesses the "potential of the task" to engage students, the "implementation of the task", the "rigor in students' responses to the task" and the "academic rigor in teacher expectations" (Matsumura et al., 2006a, p. 52-55).

5. *Performance Assessment for California Teachers (PACT)*

The Performance Assessment for California Teachers (PACT) is a culminating assessment for pre-service teachers, a test which they must pass to earn their credential. Using both classroom video and a teaching event, it measures certification candidates' abilities in Planning, Instruction, Assessment, Reflection and Academic Language. PACT provides the opportunity to have a baseline understanding of IMPACT participants' abilities prior to employment as the full-time teacher of record. Apprentices complete the PACT as a culminating assessment for their credential.

6. *California Standards Test (CST) scores*

CSTs (California Standards Tests) are administered to all public school students in California from grades 2-11. They test a variety of subjects, including ELA, Mathematics, Writing, History, Social Science and Science (California Department of Education). Standardized test scores offer several important pieces of information. First, it is imperative to find measures of teacher practice that improve not only instruction, but also student outcomes. As Matsumura, Garnier, Pascal and Valdes (2002) state, the question is "how changes in specific aspects of classroom practice may (or may not) influence student learning" (p. 208).

Student achievement gains are an integral part of evaluating the IMPACT program. Finding achievement gains, in this case using CST scores, is a signpost of program success, and thus evaluating those gains is an important part of evaluating the program.

How Will We Use these Data?

Data will be collected and used for three purposes: evaluation, research, and school improvement. Evaluation of the program, including choice of the research tools, is organized by the National Center for Research and Evaluation Standards and Student Testing (CRESST) at UCLA. The evaluation is designed to serve both summative and formative purposes; that is, it will provide results pertinent to overall program effectiveness as well as information that the program can use on an on-going basis for program improvement and refinement. Second, educational researchers will be investigating other aspects of the program. For example, one professor is investigating how videotaping classroom teaching can be used to improve

instruction. Lastly, data gathered will be used in schools by educators—including the apprentices and mentors—to reflect on practice and inform educational decisions. We are currently designing professional development and support for schools to build their capacity to collect and use multiple measures of teaching practice to improve instruction and student learning.

References

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher,* 33(8), 3-15.

Borko, H., Stecher, B. M., Alonzo, A. C., Moncure, S. & McClam, S. (2005). Artifact packages for characterizing classroom practice: A pilot study. *Educational Assessment*, 10(2), 73 -104.

California Department of Education. *California Assessment System.* Retrieved from http://www.cde.ca.gov/ta/tg/sa/caassessment.asp

Easton, J. Q. (June 29, 2010). New Research Initiatives for IES - IES Research Conference Keynote Address. Retrieved from http://ies.ed.gov/director/speeches2010/2010_06_29.asp

Hiebert, J., Gallimore, R., Garnier, H., Givvin, K. B., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching mathematics in seven countries: Results from the TIMSS 1999 video study*, NCES (2003-013), U.S. Department of Education. Washington, DC: National Center for Education Statistics.

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

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 461-494). New York, NY: Cambridge University Press.

Kennedy, M. M. (1999). Approximations to indicators of complex student learning." *Educational Evaluation and Policy Analysis*, 21(4) 345-363.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., & Resnick, L. (2006a). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment* (CSE Technical Report 681). Los Angeles: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Matsumura, L.C., Slater, S.C., Wolf, M.K., Crosson, A., Levinson, A., Peterson, M., & Resnik, L. (2006b). Using the Instructional Quality Assessment toolkit to investigate the quality of reading comprehension assignments and student work." (CSE Technical Report #681). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Matsumura, L.C., Garnier, H., Pascal, J. & Valdes, R. (2002). Measuring instructional quality in accountability systems: Classroom assignments and student achievement. *Educational Assessment*, 8(3), 207 - 229.

Pianta, R., & Hamre, B. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119.

Putnam, R.T. & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher,* 29(1), 4-15.

Rowan, B. & Correnti, R. (2009). Studying reading instruction with teacher logs: Lessons from the study of instructional improvement." *Educational Researcher, 38*(2), 120-131.

Spencer, J., Park, J. & Santagata, R. (2010). Keeping the mathematics on the table in urban mathematics professional development: A model that integrates dispositions toward students. In M. Foote (Ed.), *Mathematics Teaching and Learning in K-12*. New York: Palgrave.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive concept. *Teaching and Teacher Education*, 17, 783-805.