Annenberg Institute for School Reform | *Voices in Urban Education*

VUE

Winter 2005

Evidence-Based Practice

Voices in Urban Education
Number 6, Winter 2005

Publisher
Annenberg Institute for School Reform
at Brown University
Warren Simmons, Executive Director

Editor
Robert Rothman

Managing Editor
Susan Fisher

Copy Editor
Margaret Balch-Gonzalez

Production & Distribution
Mary Arkins

Design
Gilbert Design Associates, Inc.

Illustration
Carolina Arentsen

Voices in Urban Education (ISSN 1553-54X) is published quarterly at Brown University, Providence, Rhode Island. Articles may be reproduced with appropriate credit to the Annenberg Institute. Single copies are $10 each, including postage and handling. A discount is available on bulk orders. Call 401 863-2018 for further information.

The Annenberg Institute for School Reform was established in 1993 at Brown University. Its mission is to develop, share, and act on knowledge that improves the conditions and outcomes of schooling in America, especially in urban communities and in schools serving disadvantaged children. For program information, contact:

Annenberg Institute for School Reform
Brown University, Box 1985
Providence, Rhode Island 02912
Tel: 401 863-7990
Fax: 401 863-1290
Web: *www.annenberginstitute.org*

Annenberg Institute for School Reform | ***Voices in Urban Education***

VUE

Evidence-Based Practice

VUE's Web site at <*www.annenberginstitute.org/VUE*> offers more information about VUE, including excerpts from previous issues, audio and video clips, ordering information, and an on-line message board.

A Fad or the Real Thing? Making Evidence-Based Practice Work

Robert Rothman

Among the many buzzwords swarming around education reform, "evidence-based practice" has become one of the hottest. Spurred in part by No Child Left Behind – with its more than one hundred references to research and evidence – and, in part, by efforts by the business community and others to help infuse educational decisions with data, schools and school systems are quickly lining up to demonstrate how their curricular and spending decisions reflect evidence about what works and what's needed.

To a great extent, this trend is a positive one. Educators will admit that many decisions have been based more on history (the way schools have always worked) and on politics (the wishes of a favored constituency) than on evidence. And with budgets tight, administrators are eager to show that schools are producing results.

Yet, like many concepts, "evidence-based practice" can mean many different things, and the way it is interpreted and applied can determine whether it represents a real change in the way schools operate or just another fad.

The notion of evidence-based practice implied in No Child Left Behind is a limited one. The law is generating reams of data, but nearly all of it is standardized-test data that cannot sufficiently inform decisions about programs and practices. While it is useful to know whether certain groups of students are performing less well than others in mathematics, it is also important to know what the classrooms are teaching so that schools can know what to change.

Robert Rothman is a principal associate at the Annenberg Institute for School Reform and the editor of Voices in Urban Education.

In addition, the law's emphasis on knowing "what works" is based on a limited model of research that assumes that a program that works in one school will work in any school. The type of randomized trials the law holds up as the "gold standard" for education research, similar to the kind of studies used in medical research, say little about how to implement an innovation.

This issue of *Voices in Urban Education* examines evidence-based practice and its application in urban school systems. In an introductory essay, Warren Simmons points out the limitations of the data and research methods implied by No Child Left Behind. He argues for expanding data systems to include measures of the conditions of instruction and for expanding the notion of research to include collaborations between researchers and practitioners that would take into account the context of educational innovations.

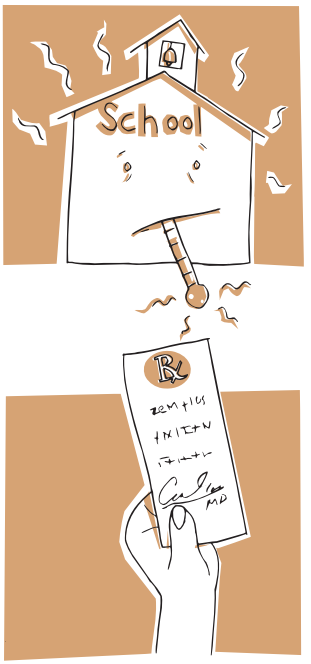
David Abbott describes an accountability system being developed by the Rhode Island Department of Education that aims to help districts provide appropriate support to a range of schools. To that end, the state collects a broad range of evidence on school and district practice.

Meredith Honig and Cynthia Coburn scan the evidence on how district administrators actually use evidence. They find that, contrary to conventional wisdom, administrators rely on evidence fairly heavily. However, they also find that the evidence does not

necessarily shine "floodlights" on practice; administrators need models for how to use evidence effectively.

Dale Mezzacappa describes how journalists use research in examining school and district practices. She provides examples of newspaper articles that combine data with in-depth stories of classrooms and families; together, these two kinds of evidence give readers a good sense of how schools work.

All of the articles make clear that evidence-based practice will only make a difference if decision-makers are skillful in analyzing data and in using it to inform decisions. Too often, school systems ignore this lesson and leap to structural solutions – for instance, building an information system without considering what the information is supposed to be used for and who is supposed to use it. Unless practitioners have the knowledge and skills to use information effectively, all the data in the world will do little good. Perhaps that's one reason Honig and Coburn found little evidence that evidence-based practice has so far improved school outcomes.



The growing interest in evidence-based practice suggests that many believe the concept has the potential to produce genuine improvements. The good news is that more and more people are unwilling to take those assurances on faith. With a broad range of data and an expanded notion of how to analyze it, we might begin to see real evidence about evidence-based practice very soon.

Evidence-Based Practice: Building Capacity for Informed Professional Judgment

Warren Simmons

The current emphasis in education reform is on scientific research that answers the question "What works?" However, this approach often fails to provide useful guidance for school practitioners. To improve student outcomes, researchers and practitioners must collaborate on research that explains why a particular practice meets students' needs and what conditions are needed to enable the practice to work.

Education reforms over the past two decades have tended to emphasize *will* rather than *skill*. The assumptions behind standards-based reform seem to be that educators know what to do, but that they lack the incentives and the flexibility to take appropriate steps to improve performance. As Richard Elmore and Robert Rothman (1999) observed, the implied theory of action underlying standards-based reform held that if schools were given more resources and flexibility in exchange for being held accountable for getting all students to meet high standards, teachers and principals would have the motivation needed to foster continuous improvement in student achievement.

While the prevalence of school-funding litigation casts doubt on whether resources are adequate for improvement, serious questions have also been raised about whether or not practitioners – particularly, those in urban settings – have the knowledge needed to improve conditions of learning and thereby raise student outcomes. Most notably, highly qualified teachers are in short supply. Dysfunctional human resource systems and chronic funding shortfalls leave many urban school districts poorly prepared to compete for skilled teachers. These same human resource systems also operate to assign novice or struggling teachers to urban schools serving students with the greatest academic and economic needs (DeStefano & Foley 2003).

Warren Simmons is executive director of the Annenberg Institute for School Reform.

Research as a Capacity-Building Tool

Standards-based reform, however, has not been entirely blind to the need for greater expertise to guide school and classroom improvement. America 2000, the education reform proposed by President George Bush in 1990, created the New American Schools (NAS) Development Corporation to sponsor a set of research-based whole-school reform designs that schools could adopt to help students reach high standards.

This approach assumed that schools lacking the internal expertise to promote student achievement could be improved significantly by acquiring a carefully articulated design that was research based. In 1998, Congress affirmed this reasoning by passing legislation authorizing $150 million to pro-

Research-based designs or programs, by themselves, do not guarantee improvement if conditions supporting their effective use are not present.

vide schools with the resources needed to implement what were then called *research-based comprehensive school reform designs*. By this time, the eleven original models created by NAS had been joined by over forty designs possessing various forms of research that demonstrated their efficacy and/or origins in a particular line of research.

The term *research-based,* as applied to designs, has two separate meanings. One has to do with the degree to which a design and its various elements are rooted in research on learning, development, and other areas related to education. The second involves the degree to which a particular design has evaluation data demonstrating its effectiveness. The models listed as being research-based vary along these two dimensions of research considerably, yet they all carry the federal government's stamp of approval.

The weight placed on research-based school designs as a solution for failing schools rests on the quality of available research, on the one hand, and the capacity of schools and districts to implement a design faithfully and effectively, on the other. In a report evaluating the progress of the eleven original NAS school designs, Thomas Glennan (1998) observed that the "success" of a design in a particular school was a function of several factors:

> The performance of a school would clearly be a function of the design, the assistance received as it was implemented, the nature of the school itself, and a variety of qualities of a district's operating environment. Attributing a school outcome to the design is clearly inappropriate. (p. 79)

In short, research-based designs or programs, by themselves, do not guarantee improvement if conditions supporting their effective use are not present. Charles Payne and Mariame Kaba (2001) made similar observations about the difficulty low-capacity schools in Chicago encountered in making effective use of research-based school reform designs that required a baseline set of conditions (e.g., collaborative culture, shared vision) that were often weak or absent in chronically failing schools. Moreover, the conditions that predict improvement include a school's ability to match its needs with the elements of a design, the ability of an external provider to offer adequate professional development and support services, and the capacity of the district governing the school to align its supports and services with those provided by the design (Hassel 2002).

This cautionary tale about the transferability of research-based designs and programs has not dampened belief in the power of research to inform practice and improve schools. If anything, No Child Left Behind (NCLB), the 2001 Reauthorization of Title I of the Elementary and Secondary Education Act, has amplified the importance of research by making the receipt of

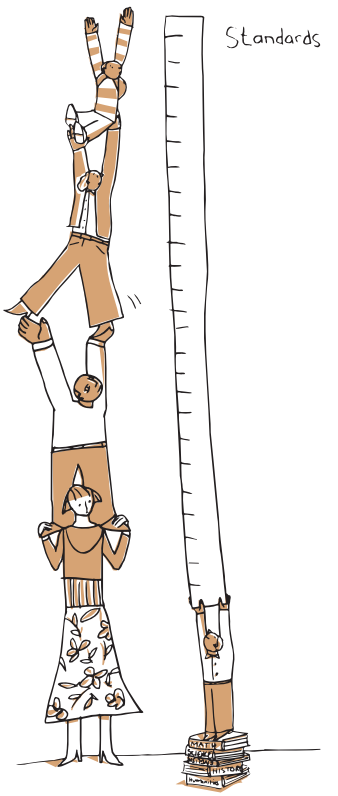
Title I dollars contingent on schools' and entire districts' adopting research-based curricular programs, along with offering incentives for the use of research-based school designs. Moreover, NCLB has expanded the importance of research, while narrowing the definition of acceptable research to studies using randomized assignment of subjects to experimental and control groups. This restriction is based on the largely unsubstantiated assumption that the weak connection between research on what we know and what we are able to achieve in practice has had more to do with the quality of existing research than with the conditions affecting its use in practice.

NCLB and the U.S. Department of Education's new Institute of Education Sciences attempt to address the "quality" problem by creating strict scientific standards for research supported by the department and by tying federal Title I dollars to the use of programs or designs documented by research that meets these standards.

Expanding Available Evidence

NCLB also bolsters research or evidence-based practice in another way. The law requires states to issue school and district reports that reveal whether schools and their major student subgroups – African Americans, Hispanics, etc. – are making adequate yearly progress toward proficiency in mathematics and reading.

This federal requirement, in effect, establishes a base of data that educators and the public can use to inform their judgments about the needs of particular schools and districts. Presumably, the widespread availability of data on the status and needs of students, schools, and districts, coupled with more robust research-based curricular programs and school designs, should

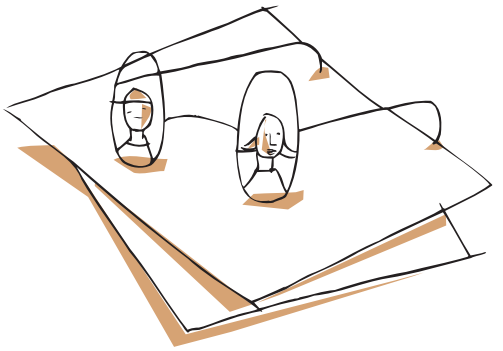


serve to accelerate the impact of research on practice and foster improvements in school and student performance.

Self-Imposed Limitations

NCLB should be applauded for stressing the importance of research and of making disaggregated data on student performance at the school and district levels widely available and publicly reported. But the potential impact of these measures is undermined by the narrow definition of research and limited scope of the data covered by the law.

NCLB's reliance on standardized tests as a primary source of data to inform public judgments about student and school performance has been criticized widely. These measures typically assess a narrow range of desired skills and knowledge and provide little information about the conditions of learning that produce performance disparities. They often cause schools to narrow the curriculum in a desperate effort to raise test performance (Herman 2003). At the same time, the law's emphasis on randomized trials provides little incentive for states and districts to invest in other types of research that might be more valuable and could inform practice in more useful ways.



Monitoring Conditions of Instruction

The shortcomings of standardized tests are well known, but alternatives that would provide broader and more useful information and lessen the possibility of perverse effects are less familiar. The National Research Council's Title I Assessment Committee argued that schools and districts should be aided in making the links between assessment, learning, and practice by developing measures that monitor the conditions of instruction. These kinds of measures would assess the quality of the *enacted curriculum* – shorthand for what Richard Elmore (2002) describes as the intellectual and material resources that influence the interactions of teachers and students with content (e.g., subject-matter knowledge combined with students' prior knowledge).

To gauge the quality of the enacted curriculum, schools, school districts, and, to some extent, states would have to invest in the systematic and ongoing use of *formative assessments*. These assessments are needed to shed light on the alignment of state and district standards with the intellectual rigor and quality of

- instructional texts and materials;
- teacher lessons and assignments;
- classroom discourse between and among students and teachers; and
- student work.

Models of such formative assessments exist and include Lesson Study,

an approach modeled after teaching practices in Japan, and the Teaching and Learning Review, a method developed by the Annenberg Institute for School Reform to assess the match between the *intended curriculum* and the enacted curriculum. To date, though, these and other models for ongoing formative assessment remain confined to networks of teachers and schools. To become tools that bridge the gap between standards and achievement, they must be elevated to become part of a system of assessments used by schools and districts to improve instruction in the manner described by Elmore.

Aligning Clinical and Research Knowledge

The development and systematic use of formative assessments – focused on conditions of instruction in schools and the relation of these conditions to district policies and supports – would provide practitioners the data they need to guide their professional judgment. Formative assessments of classroom, school, and district practice would also help practitioners weigh the appropriateness and relevance of basic and evaluation research touting specific programs and practices.

The current emphasis on scientific research that demonstrates "what works" overlooks a second critical question for the practitioner: What works, given the needs and values of my students and community and the condition and capacity of my school and district? If this second question is ignored, schools can be led to choose research-based designs and programs that don't address the needs of their learners and practitioners. For instance, many of the research-based whole-school reform designs lack supports for English-language learners and students with disabilities, yet they carry the U.S. Department of Education's seal of approval as being "proven." In this case, it is left up to the "buyer" to decide if *proven* translates into *effective* for the constellation of teachers and students in a particular school or school district.

The current emphasis on scientific research that demonstrates "what works" overlooks a second critical question for the practitioner: What works, given the needs and values of my students and community and the condition and capacity of my school and district?

Unfortunately, the context-specific nature of effective practice is overshadowed by the implication of the "what works" incantation that highlights results but obscures the conditions needed to ensure effectiveness. To advance practice and improve student outcomes, researchers and practitioners must collaborate to conduct research that addresses five basic questions:

- What are the needs of the students being served and the practitioners charged with supporting their learning?

- What are the conditions of instruction within and across challenged schools or schools needing improvement?
- How do district policies and supports work to affect these conditions for better or worse?
- Given answers to the first three questions, which research-based programs or designs are best suited to addressing the needs and conditions present in my school/district?
- What are the short-term, intermediate, and long-term resources and strategies I must consider to ensure effective implementation of a particular program or design?

These questions are rarely answered by basic or evaluation research because the design of such studies pays more attention to whether an effect occurred or not, as opposed to how the effects hold up under a range of conditions. It reminds me of my youthful rush to buy a BMW, the "ultimate driving machine." It might have been the ultimate driving machine in California, but in the hilly and icy driving conditions in Ithaca, New York, where I attended graduate school, it became a "near-death driving machine" that I left parked in the garage for much of the time.

Unfortunately, schools and districts often adopt effective programs and practices that wind up like my BMW, parked in the garage due to a poor match between the needs of the school or district and the components of a particular program or design (Corcoran, Fuhrman, and Belcher 2001).

Integrating Research and Practice

Scientific research alone will not build a stronger knowledge base about the relevance of research-based educational programs and designs to specific contexts found in urban schools. Research

addressing this issue must be sensitive to the challenges and conditions in urban schools and make this variability a focal point of study rather than a factor to be controlled. Denis Newman (2004) makes a similar point in a recent critique of the emphasis on scientific research in education:

> An instructional program cannot be "proven effective" by a single program of research. Unless decision-makers can recognize their own situation in the site where the research was conducted, they have little reason to think that the same effect will occur for them. The best evidence will always be that collected locally. (p. 9)

Rather than dismissing the importance of scientific research, Newman and others argue for the need to augment randomized, controlled experiments with what Fritz Staub (2004) calls "design experiments." According to Staub:

> In design experiments, researchers engage in close collaboration with practitioners, and they are jointly accountable for the experiments carried out. Design-teams consisting of practitioners and researchers allow detailed local knowledge from the field of practice to be included in the design process, and the close collaboration is instrumental in ensuring that interventions are implemented as planned. Design experiments are to develop theories as well as new forms of practice through repeated cycles of designing, implementation, and analyses. (p. 44)

I would only depart from Staub's position by giving priority to generating new forms of practice over developing new theories, given the moral imperative of school improvement. The kinds of research Staub describes are now being carried out by a new generation of locally focused, applied-research organizations such as Research for Action in Philadelphia, Education Matters in Boston, and the Center for Research on the Context of Teaching at Stanford University. Typically, these organizations define problems and conduct research in collaboration with teams of researchers and practitioners.

This kind of close relationship helps bridge the gap between research and practice, as well as between research and clinical knowledge, and ensures that research is attuned to local conditions. These kinds of collaborations also ensure that practitioners' concerns with *how* a program works are given equal weight with researchers' emphasis on describing *what* works. Moreover, this approach circumvents the need to figure out how to translate research into practice by involving practitioners in the research itself. As a result, practitioners serve as researchers and translators simultaneously.

These kinds of collaborations also ensure that practitioners' concerns with *how* a program works are given equal weight with researchers' emphasis on describing *what* works.

While this type of research has the potential to quicken the pace of learning, development, and improvement, the conditions and incentives needed to advance its development bear some examining. If school and central office practitioners are to be continually involved in these kinds of endeavors, district structures and policies must support collaborative inquiry and provide

the time and incentives needed to build partnerships with researchers.

Districts provide fertile ground for practitioner collaboration when they support the development of embedded professional learning communities (groups of practitioners working within and across schools, areas or regions of a district, and the district central office) with the features outlined by the Annenberg Institute (2004) and by Milbrey McLaughlin and Joan Talbert (2003). This fertile ground is further enriched by an emphasis on evidence-based practice, tools, and strategies that support the collection and sharing of data, a shared vision of instruction, strong internal accountability, time for planning and reflection, distributed leadership, and continuous cycles of inquiry.

Researcher-practitioner collaboration also requires changes in the incentives and structure of higher education. The current reward structure for research in higher education values contributions to the field over contributions to practice.

Researcher-practitioner collaboration also requires changes in the incentives and structure of higher education. The current reward structure for research in higher education values contributions to the field over contributions to practice. Junior faculty facing tenure decisions must publish in refereed journals where scholars determine the merits of the research based on the extent to which it contributes to discipline-based theory. Design experiments, in contrast, are often interdisciplinary in nature and are oriented toward practice and product rather than driven by theory. Recent efforts by some universities to create clinical professorships reflect a nascent but significant recognition by a segment of higher education of the central importance of this kind of research to the mission of the university.

As "evidence-based practice" becomes a new watchword in education reform, its potential to improve conditions and outcomes in urban schools, particularly those serving disadvantaged children and youths, depends on how we understand and use it. We must be able to grasp the "deep structure" needed to attain meaningful evidence-based practice. We must also be able to advocate for the changes required in K–12 and higher education to make this work a valued and common expectation for researchers and practitioners. If we fall short in these tasks, evidence-based practice will find itself "garaged" with other promising reforms of the past.

References

Annenberg Institute for School Reform. 2004. *Professional Learning Communities: Professional Development Strategies that Improve Instruction.* Providence, RI: Annenberg Institute for School Reform, Brown University. Available for download at <www.annenberginstitute.org/publications/plc.html>.

Corcoran, T., S. H. Fuhrman, and C. L. Belcher. 2001. "The District Role in Instructional Improvement," *Phi Delta Kappan* 83, no. 1:78–84.

DeStefano, J., and E. Foley. 2003. "The Human-Resource Factor: Getting – and Keeping – Good Teachers in Urban Districts," *Education Week* (April 16). Available online at <www.edweek.org/ew/articles/2003/04/16/31destefano.h22.html>.

Elmore, R. F. 2002. *Bridging the Gap between Standards and Achievement: The Imperative for Professional Development in Education.* Washington, DC: Albert Shanker Institute.

Elmore, R., and R. Rothman, eds. 1999. *Testing, Teaching, and Learning: A Guide for States and School Districts.* Washington, DC: National Academy Press.

Glennan, T. 1998. *New American Schools After Six Years.* Santa Monica, CA: RAND.

Hassel, B. 2002. *Making Good Choices: A Guide for Schools and Districts.* Revised edition, updated by Maria Ferguson. Naperville, IL: North Central Regional Educational Laboratory. Available online at <www.ncrel.org/csri/choices/makegood/title.htm>.

Herman, J. L. 2003. "State Test Lessons: The Effects on School Reform," *Voices in Urban Education* 1 (Spring):46–55. Providence, RI: Annenberg Institute for School Reform, Brown University.

McLaughlin, M., and J. Talbert. 2003. *Reforming Districts: How Districts Support School Reform.* Seattle: Center for the Study of Teaching and Policy, University of Washington.

Newman, D. 2004. *Generating the Scientific Evidence Schools Need so as to Leave No Child Behind.* Palo Alto, CA: Empirical Education Inc.

Payne, C., and M. Kaba. 2001. "So Much Reform, So Little Change: Building-Level Obstacles to Urban School Reform." Unpublished manuscript, Duke University.

Staub, F. C. 2004. "Transforming Educational Theory into Usable Knowledge: A Case of Co-constructing Tools for Lesson Design and Reflection." In *Quality in Practice-Oriented Research in Science Education,* edited by B. Ralle and I. Eilks, pp. 41–52. Aachen, Germany: Shaker.

Bringing Measurement to District-Based Accountability: The Challenge for State Education Departments

David V. Abbott

In response to new demands, state departments of education are changing the way they measure school and district performance. An effective approach is to look at a broad range of indicators and to provide guidance to districts so they can better support low-performing schools.

Over the past few years, federal and state policies have dramatically transformed the role of state education departments in evaluating school and district performance and program quality. The No Child Left Behind Act has placed accountability for student achievement – as measured by performance on standardized state assessments – at the forefront of state agendas. At the same time, many states, such as Rhode Island, have transformed their own accountability systems from *authority-based systems,* which issue rules based on school performance, to *inquiry-based systems,* which give districts information and tools to help the districts give needed assistance to schools.

This combination of factors makes designing state-level accountability systems a difficult task. State education agencies are not only concerned with schools' meeting federal expectations, but must also be aware of how well those schools and districts are progressing toward developing their own internal accountability systems. Effective quality assurance and program evaluation rest on an understanding of the complex forces at work in schools and school systems.

Effective accountability requires information from multiple sources to inform state departments of education as they conduct analyses and perform their new functions. Standardized-test results are not enough; states need to collect and analyze a broad range of evidence about school and district programs so they can support districts in their efforts to improve practices that support learning.

Rhode Island has adopted a system of accountability in public education that focuses on the school district's role in maintaining an internal system of continuous improvement. Schools are answerable to their districts, which in turn are responsible to, and supported by, the state education agency. Direct state intervention in individual schools is limited to schools in need of corrective action. Building a district's capacity to design, implement, and evaluate its ongoing reform efforts has become the state's highest priority.

This focus on the capacity of districts to embrace standards-based reform has greatly changed the demands upon the state education agency. The state must set clear standards and expectations for district actions across a num-

David V. Abbott is assistant commissioner of the Rhode Island Department of Elementary and Secondary Education.

ber of content areas, and those standards must be supported by an infrastructure that identifies a range of proven management tools for districts to use to meet those expectations.

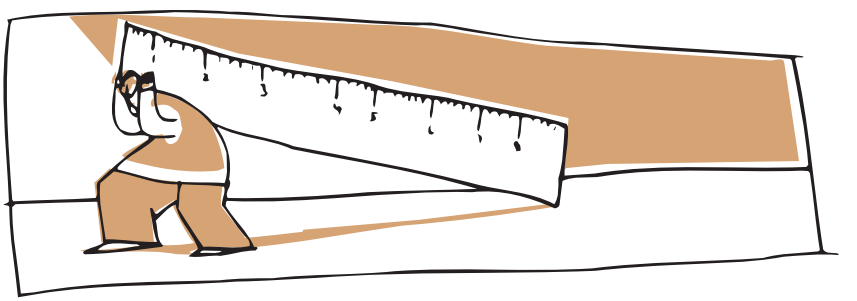
The variation of demands among urban, suburban, and rural school districts calls for a wide variety of tools. Bringing all students to high standards requires both statewide and local student assessments, curricula that are aligned with state standards, effective instructional practices, high-quality teacher preparation and training, integration of parents into the instructional process, and a coordinated accountability structure that facilitates informed decision making. Measuring the efficacy of these program and practice components is a significant challenge for both states and districts.

Rhode Island relies on two complementary accountability systems for improving student performance: School Accountability for Learning and Teaching (SALT) and Progressive Support and Intervention (PSI). Both systems rely on the use of assessment results and other evidence of student performance to inform systemic, ongoing improvement efforts. The SALT process is an annual cycle of school-based improvement activities consisting of self-study, planning, implementation of action plans, program evaluation, and public reporting. PSI is a results-driven system of accountability in which the state works with districts to address the demands of schools identified as "in need of improvement." Both SALT and PSI focus not only on ultimate student outcomes, but also on the systemic components that support effective teaching and learning practices.

Rhode Island's school- and district-based systems of accountability are inquiry based and evidence dependent. There is an expectation of sustained and continuous improvement of student achievement in all school systems, as measured against the state's grade-level expectations (grades 3–8), grade-span expectations (secondary), and the broad-based standards for student learning in the four areas of the state's Common Core of Learning: communication, problem solving, body of knowledge, and personal and social responsibility (State of RI, RI Bd. of Regents, and RIDE 2001). Measures of student proficiency in such a system cannot be limited to what can be measured by standardized assessments.

Building Districts' Capacity for Improving Student Achievement

For a state to use measurement to *support* as well as *monitor* districts' improvement, it must meet the challenge of coming up with appropriate measures of practice. The goal is to



The essence of accountability lies in knowing what to do when confronted with evidence that performance is not meeting expectations.

improve performance, not just *measure* it. The essence of accountability lies in knowing what to do when confronted with evidence that performance is not meeting expectations. Standards, plans, and measures are of little use if the actions generated by these data are poorly designed or ineffectively implemented.

When, three years ago, Rhode Island switched its focus to improving districts' capacity to support schools, there was very little research available to guide our approach. Most studies on accountability had looked at *individual* schools, but little had been done to identify what districts needed to do to become effective at managing reforms *across* schools. Our first step was the creation of clear expectations for district performance. With the support of our Regional Education Laboratory at Brown University, we convened a group of state and district educators, who created a set of outcome-based standards called District Expectations for Improving Learning and Achievement.

Many states are now developing similar sets of district-level content and performance expectations to complement existing standards for students, teachers, and schools. Rhode Island chose to focus on seven areas representing distinct subsystems, each of which is characterized by a list of measurable expectations:

- Leading the focus on learning and achievement
- Ensuring equity and adequacy of fiscal and human resources
- Guiding the selection and implementation of curriculum, instruction, and assessment
- Recruiting, supporting, and retaining highly effective staff
- Using information for planning and accountability

- Engaging families and the community
- Ensuring safe and supportive environments for students

For each of these seven areas, we have combed the knowledge base to identify effective practices, which has enabled us to select appropriate performance measures. As with standards for student achievement, district expectations and their indicators allow the state education agency to work with school and district personnel to assess current district capacities in order to better inform resource allocation, strategic planning, and differentiated instructional supports to schools.

The PSI system of accountability is designed to build district capacity to improve teaching and learning. School districts start with a self-assessment, using a tool designed to measure their capacity across the seven areas of the district expectations. The Rhode Island department of education performs a similar task, based on an in-depth review of a wide range of data from different sources about the district, which are then compiled into a comprehensive district profile. These two perspectives on current gaps in the district's capacity to implement reforms – the district's and the state's – are reviewed together at a face-to-face meeting between the state agency and the district. Strategic and resource-allocation decisions are made at this meeting, based on agreed-upon priorities and sequencing of action steps. All decisions are made within the context of bolstering the district's capacity to oversee its own school-improvement efforts.

This emphasis on internal accountability for continuous improvement requires an understanding of the complex and overlapping operations in schools and school systems. Each district central office has a primary responsibility to ensure that its schools have the capacity to lead and support their faculties in activities aimed at improving the quality and efficacy of learning and teaching. But systems capacity alone is not sufficient to improve student performance. Effective accountability initiatives must yield fundamental changes in the actions of teachers, administrators, students, and parents in ways that improve learning. People within these systems must both value those changes and believe that positive change is possible.

There is virtually no aspect of education operations that cannot benefit from established standards for content and performance.

Performance expectations for schools and districts must therefore be considered in concert with standards for educational professionals and expectations for curriculum development, instructional practices, parent engagement, school safety, and school governance. There is virtually no aspect of education operations that cannot benefit from established standards for content and performance.

Generating Multiple Sources of Meaningful Data

Many state and local education agencies have begun their development of information technology systems by developing the infrastructure to collect, store, and access information, without first determining which data elements should be generated and who will be able to analyze the resulting rich information base. We went about it the

opposite way. We know that the identification of useful and meaningful indicators, coupled with the development of analytical skills among practitioners, is the basis of an effective system of accountability.

Choosing which indicators to measure raises immediate problems. On the one hand, there is the temptation to use a multitude of indicators in order to support extremely discrete, complex analyses. On the other hand, state agencies may want to select a parsimonious set of indicators to ensure that they are as accessible and intuitively graspable as possible. Education should not become a "black box" endeavor controlled by the few researchers and analysts able to make sense of a bewildering sea of information. What is needed is to develop the inherent analyst in every educator and to bring analysis and differentiation to every classroom.

The identification of useful and meaningful indicators, coupled with the development of analytical skills among practitioners, is the basis of an effective system of accountability.

The potential number of indicators in any one information system is enormous. It is important to define the range of indicators needed to track the development of any system element as the smallest number of data required to make sound decisions. The ability to track changes over time and to represent these changes in easy-to-understand chart form are also important.

An effective accountability system requires information from multiple sources to inform an analysis of the many aspects of education systems. Relevant data may be effectively grouped into four categories:[1]

- *Contextual* and *demographic* data, including student characteristics such as family income status, mobility, race/ethnicity, gender, limited English proficiency (LEP) status, truancy, attendance, discipline referrals, and graduation rate, as well as school characteristics such as school enrollment, structure, scheduling, class size, parent participation, teacher certification, and teacher assignments.
- *Inputs* or *resources,* such as curriculum, student interventions, professional development and mentoring supports, and common planning activities.
- *Process measures* or *instructional practices,* including perceptual or evaluative evidence of standards-based instruction provided by self-study, surveys, learning support indicators, mission statements, observations, and structured school visits.
- *Outcomes* or *measures of student learning,* consisting primarily of state assessment results, local assessments, student work, and teachers' observations.

The combination of quantitative and qualitative information from these four sources of data allows for detailed analysis of the impact of instructional practices and support mechanisms on student learning. The multiplicity of these measures is the key to an effective comprehensive education informa-

1 Margaret Votta, senior research associate at the Annenberg Institute for School Reform, contributed significantly to my knowledge of this categorization, which is adapted from Victoria Bernhardt (1999) and Madhabi Chatterji (2002).

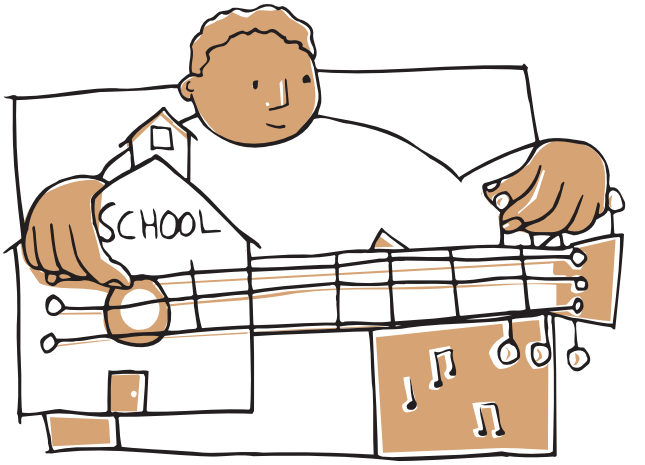
tion system. Developing a system of measurement that incorporates all four areas involves a number of challenges.

Contextual and Demographic Data

Most states have already identified contextual and demographic indicators, especially the states that have adopted comprehensive standards such as those developed by the National Center for Education Statistics. In response to federal reporting requirements, existing district information systems now contain information they need to compare schools or districts.

Inputs

Many districts have identified the inputs associated with high performance, but they may not be adequately measured. For example, many districts place a high value on the use of articulated curricula aligned to state standards. But alignment is difficult to measure. Measuring a district's performance is also challenging because many resources are directed at the school level, often creating disparities in application across a district. There is a clear need for state education agencies to tackle these issues related to measurement of inputs in educational systems to enable schools and districts to make coherent links between reform efforts and resultant student performance.



Process measures

Measuring the effectiveness of education processes is perhaps the most difficult challenge in bringing measurement to education accountability. Rhode Island approaches this issue through three complementary tools within its SALT initiative. The first is the SALT Survey, a comprehensive survey with different sets of inquiries for teachers, administrators, students, and parents. This tool provides invaluable

The ability to draw accurate inferences from a complex body of evidence is not an area of competence we have traditionally required of education professionals.

insight into the beliefs and opinions of these constituencies with regard to a wide variety of topics, including instructional practices, school-improvement efforts, professional development, school safety, parent involvement, student engagement, and faculty culture.

The SALT Survey is complemented by the SALT visit, which is conducted on a multiyear schedule. Based in part on the British "Inspectorate" model, the four- or five-day visit by peer reviewers from other districts and the state education agency provides an outside, structured perspective on student learning, teaching, and school climate and organization. At the conclusion of the visit, a report is issued on the findings of the team. The report is meant to be a reality check to see whether the school's perceptions match those of the outside peer observers, who share the same goals for student and school performance and understand the concerns and challenges of the public school environment.

Finally, state officials collect and monitor the content of school-improvement plans, district strategic plans, financial data, and consolidated grant applications. These reviews reveal where a district is committing resources and how it outlines its priorities. They also provide insight into what is not yet on the district agenda. Due to the complexity of program evaluation and other measures of educational practices, state education agencies bear a significant responsibility in the development of these largely qualitative information systems needed to measure processes and practices.

Outcomes

Measurement of outcomes presents two distinct difficulties. First, districts are just beginning to see the value of generating a rich and immediate picture of student achievement that can only be generated by the use of local assessments. Designing and implementing an integrated system of local assessments that provides immediate feedback to the classroom teacher and that informs specific practices is extremely difficult. Second, creating discrete measures of the effectiveness of individual inputs or practices has bedeviled an educational system intent on relying on a scientific research base, although some emergent work has been done in this area.

Precise evaluation of distinct program and practice efforts is largely beyond the capacity of most school districts and even of most state education agencies. Program and practice evaluation techniques remain the purview of skilled consultants and research institutions. Widespread use of accurate measures of educational subsystem outcomes remains in its infancy, where, without concerted regional or national initiatives to bring program evaluation to our schools and districts, it will remain.

A Comprehensive Education Information System

Statewide accountability for student achievement requires the integration of measures from all four of these categories of data. For diverse stakeholders to share diverse information and use it to inform decision making, a state needs a comprehensive education infor-

mation system. Teachers and administrators must be able to use information systems to access a broad range of data to analyze student results, explore successful practices, and continuously adapt and improve instruction.

The use of information to make decisions about the distribution of resources, adoption of programs, implementation of specific instructional practices, and assignment of personnel requires a heightened level of commitment to information technology and data analysis. Bringing such a system to bear in an educational environment is a considerable and ongoing challenge.

A comprehensive education information system consists of six core elements: identification, generation, collection, storage, access, and analysis of relevant data. Information must be current, accurate, and reliable, and should include both quantitative and qualitative measures. Generating and providing access to relevant information presents considerable logistical demands, but is only an initial step in moving towards an inquiry-based system of accountability. Every state currently generates a tremendous amount of relevant information about its schools and school systems. However, the ability to categorize, access, and analyze that information at the school, district, and statewide levels is generally insufficient or cannot be completed within time frames that make the data useful.

Effective analysis is by far the most important element of the information system. The ability to draw accurate inferences from a complex body of evidence is not an area of competence we have traditionally required of education professionals. It is a skill more often identified with research scientists than with classroom teachers and administrators. However, it is this ability to measure current performance against clear expectations that holds the most promise for creating educational systems that truly address the needs of all students.

Looking to the Future

Analysis of a school's demographic, achievement, instructional, and systems-evaluation data against specific reform goals is the key to ongoing self-study that accurately measures the effectiveness of individual action plans. State agencies will continue to work with school districts to develop their capacity to gauge the nature and extent of support required to implement new instructional programs.

This evaluative work is the key to moving from "pockets of excellence" to systemic improvement through longitudinal studies, cohort comparisons, and evaluation of specific reform strategies. Only by bringing measurement and analysis to our ongoing efforts to improve instructional practice will we be able to meet the demands of bringing all students to high standards.

References

Bernhardt, V. 1999. *The School Portfolio: A Comprehensive Framework for School Improvement.* Larchmont, NY: Eye on Education.

Chatterji, Madhabi. 2002. "Models and Methods for Examining Standards-Based Reforms and Accountability Initiatives: Have the Tools of Inquiry Answered Pressing Questions on Improving Schools?" *Review of Educational Research* 72, no. 3 (Fall):345–386.

State of Rhode Island and Providence Plantations, Rhode Island Board of Regents for Elementary and Secondary Education, and the Rhode Island Department of Elementary and Secondary Education. 2001. *Rhode Island's New Common Core of Learning.* Providence, RI: State of RI, RI Bd. of Regents, and RIDE.

When Districts Use Evidence to Improve Instruction: What Do We Know and Where Do We Go from Here?

Meredith I. Honig and Cynthia E. Coburn

How do central office administrators use evidence in their practice? A research review revealed that district leaders use evidence far more often than is commonly assumed, although there is little concrete evidence so far that this has led to improved school outcomes. Districts need ways to translate evidence into usable forms.

In recent years, school district central offices have faced unprecedented demands to use evidence in their decision making. For example, the federal No Child Left Behind Act requires that all programs funded under this initiative stem from "scientifically based research" and student-performance data – a requirement that potentially affects decisions throughout central offices, from the selection of professional development approaches to debates about the inclusion of the arts in the curriculum. The U.S. Department of Education's (2002b) *Strategic Plan for 2002–2007* explicitly calls for the "transform[ation of] education into an evidence-based field" by strengthening the quality and use of educational research.

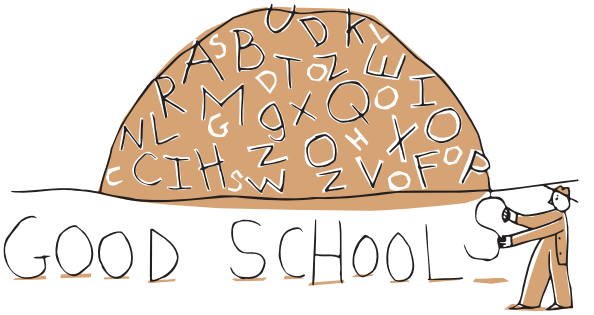
Much attention has been paid to central offices as supporters of schools' use of evidence, but what do we know about the use of evidence in district central office administrators' own decision making? To address this question, we conducted a comprehensive review of research on district central office decision making and evidence use. We cast a broad net for empirical studies on how central office administrators use various forms of evidence, including student data, research-based practice, evaluation data, and feedback from teachers, school principals, and community members. In all, we reviewed thirty-nine empirical studies, literature reviews, and descriptions of evidence use published since 1975 and various other studies on district central office decision making.[1]

Ironically, we found that pressures on district central offices to use evidence are not themselves based on substantial evidence that evidence-based decision making in central offices leads to improvements in academic, managerial, or other kinds of outcomes. These studies are few and far between and have many shortcomings. Often, studies refer to "district" use of evidence without distinguishing among different individuals throughout central offices who may or may not use evidence for a range of purposes, from choosing reading curricula to deploying school buses. Studies also tend not to distinguish among types of information in use (or not in use), even though research in the business sector and elsewhere reveals

Meredith I. Honig is an assistant professor of education at the University of Maryland College of Education. Cynthia E. Coburn is an assistant professor at the University of California, Berkeley, Graduate School of Education.

1 For a full discussion of our methods, please see Coburn, Honig & Stein, forthcoming.

that using student-performance data, for example, should pose very different challenges and opportunities than using an academic research article or a research-based school reform model. Despite these limitations of the research, we were able to glean several key lessons about what evidence use by district central offices involves and what it takes.



Lessons from Research about District Use of Evidence

The body of research we reviewed revealed that central office evidence use is a far more frequent and complex process than typically represented in federal and state mandates. Several lessons emerged from our study for central offices to keep in mind as they use evidence for decision making. We conclude that evidence use can be strengthened and expanded, in part, by increased collaboration within district central offices and stronger roles for external support providers in helping central office administrators translate evidence into usable forms.

District central offices are no strangers to evidence use.
Despite some policy claims to the contrary, superintendents and other district central office administrators have long used a variety of locally and externally generated evidence in all sorts of decision-making processes. Local evidence – community assessments, a district's evaluations of its own programs, and student and family surveys – grounds many central office decisions related to resource allocation and policy development (Honig 2001, 2003; Marsh 2002; Massell 2001). District central offices use standardized-test scores and school-improvement plans as a regular part of strategic planning processes (Massell 2001; Massell & Goertz 2002).

District central offices draw on external research studies to inform or justify decisions to choose or abandon instructional programs and school reform models (Corcoran, Fuhrman & Belcher 2001; Kean 1983; Newman, Brown & Rivers 1983; Robinson 1988). They also seek out research and research-based "best practices" through professional conferences (Datnow, Hubbard & Mehan 2002; Osheka & Campagne 1989) and from visiting researchers (Boeckx 1994), local universities (Nafziger, Griffith & Goren 1985), community-based organizations (Honig 2004a), research and development organizations (Corcoran & Rouk 1985), and voluntary associations such as the National School Boards Association and the American Association of School Administrators, among others.

Evidence use may provide light in the dark, but it's more like striking a match than turning on the floodlights.
The word *evidence,* derived from the same root as *evident* – meaning, literally, conspicuous, apparent, or obvious – may create the impression that the information being used as evidence somehow

speaks for itself. But central office administrators report that information is usually riddled with ambiguity about which information to collect, what it means, and how to use it.

The ambiguity stems, in part, from the form evidence typically takes. Research papers that emphasize the abstract rather than the concrete, that use technical language, and that are very long are more likely to be perceived as ambiguous by district administrators (Kean 1981, 1983; Roberts & Smith 1982; West & Rhoton 1994). Ambiguity also stems from the sheer volume of evidence available at any given time (Massell & Goertz 2002). Ultimately, though, implications for action are often unavoidably ambiguous, especially in complex systems like district central offices (Honig 2001; O'Day 2002). Central office administrators report that social science research and evaluation findings, in particular, often fail to provide direct guides for action (Corcoran, Fuhrman & Belcher 2001; Fullan 1980).

Ambiguity can curtail the consensus sometimes essential to central office decision making. For example, Kennedy (1982a) has revealed how researchers' attempts to balance positive and negative findings can allow central office administrators with different views to "freely infer what they wanted" about the degree to which the findings supported their position and thereby frustrate consensus (p. 82). But the ambiguity of evidence is not inherently problematic. For example, ambiguity of information about how best to meet the needs of students can open opportunities for deliberation and for tailoring reform strategies to individual student needs (e.g., Honig 2001).

Often district central office administrators use evidence in a range of political decisions that may help or hinder instructional improvement.

The connection between evidence used by central office and instructional improvement isn't always direct or predictable. Sometimes central office administrators use evidence for purposes that seem directly related to instructional improvement, such as allocating resources based on data about student needs. However, perhaps more often district central office administrators use evidence in a range of other, largely political decisions. These political decisions – not the evidence, per se – may help or hinder instructional improvement. For example, district central office administrators use evidence to bolster their arguments at school board meetings and various community events to increase board and community support for particular education reform strategies (Corcoran, Fuhrman & Belcher 2001; Marsh 2002). One superintendent recounted how he used research to "stabilize the environment" within his central office among his own staff to advance an improvement strategy: "When confronted with research, our teachers and administrators began to 'buy in' to the program" he was trying to implement (Boeckx 1994, 24).

Occasionally, research grounds central office presentations of particular reform approaches at school board meetings as a way to influence school board opinions, even if that research was not used in the development, selection, or implementation of those programs (Robinson 1988). In these ways, evidence is used to influence pub-

lic opinion or group consensus, which, in turn, impacts decision making and, other conditions permitting, improvement (Englert, Kean & Scribner 1977; Kennedy 1982a, 1982b).

If it's "useful," they will use it. Despite policies promoting evidence use and the sheer quantity of information available from a variety of sources, district central office administrators frequently report limited access to evidence they consider relevant to their most pressing concerns (Corcoran, Fuhrman & Belcher 2001). Available evidence tends to come in non-user-friendly forms. Central office administrators seem to want concise research syntheses (Corcoran, Fuhrman & Belcher 2001) based on up-to-date studies (Corcoran, Fuhrman & Belcher 2001; Roberts & Smith 1982; West & Rhoton 1994). Student-outcome data may be difficult to obtain from state educational agencies in usable form (Massell 2001).

Some central offices lack the technological infrastructure to use their own data to answer pressing questions (Reichardt 2000). Time pressures also curtail the collection of relevant evidence; in particular, district personnel cannot always wait for the results from evaluation studies or pilot programs before they take action, either because they need to react to an immediate need or because they feel pressured to appear decisive (Bickel & Cooley 1985; Corcoran, Fuhrman & Belcher 2001; Englert, Kean & Scribner 1977; Kean 1981, 1983).

How evidence is used depends on what administrators already know, can do, and need to do. Central office administrators are hardly passive recipients of evidence. They actively search for it and grapple with how to incorporate it into their decision making. Which evidence they find and bring back to the central office and how they choose to use it depends largely on their prior knowledge. For example, Mary Kennedy (1982a) has shown that when central office administrators select and use evidence, they filter or screen it through their beliefs, assumptions, interests, and experiences. According to Kennedy, "When people say they have used evidence, what they really mean is that they have rendered it meaningful by connecting it to a prevailing and usually very powerful point of view. Having done so, they can claim the evidence is relevant, timely, and compelling" (p. 101).

The process of rendering evidence meaningful typically involves the translation of evidence into forms that central office administrators consider clear and "actionable."

The process of rendering evidence meaningful typically involves the translation of evidence into forms that central office administrators consider clear and "actionable." For example,

studies show that, especially when faced with large volumes of evidence, central office administrators tend to gravitate towards evidence that is congruent with preexisting beliefs and pay less attention to that which challenges their experiences or assumptions (Coburn & Talbert, forthcoming; Kennedy 1982a, 1982b; Spillane, Reiser & Reimer 2002). They also tend to break large, complex studies into smaller pieces that they consider more manageable (Hannaway 1989). These patterns appear particularly prevalent when consequences for poor performance are high and when available evidence is complex or ambiguous (Honig 2001).

Evidence is not just for the research and evaluation unit any more; broad participation bolsters evidence use. By participating in evidence gathering and translation, central office staff become more familiar with the evidence, develop more confidence that they understand it and know how to use it, and strengthen their belief that they should use it. Broad participation also is essential because different central office administrators tend to be skilled at different aspects of using evidence. For example, certain administrators are skilled at conducting research or otherwise collecting evidence. These are not always the same people who are able interpreters of the large volumes of data that central offices are now routinely required to manage. Nor are these administrator-researchers always the same people who have the authority to decide how evidence should be used to guide central office operations (Honig 2003; Reichardt 2000). Central offices that are well organized to use research seem to distribute various evidence-related functions across staff and to have high degrees of coordination.

Collaboration helps central office administrators create common beliefs and understandings essential to making sense of evidence. Through collaboration, central office staff may increase their social capital – in this case, trusting relationships between those who have evidence and those who will use it – that may help to effectively use evidence. For example, central office administrators may have evidence of school difficulties that could either help direct new resources for school improvement or increase threats of district sanctions. Without trust that the information will be used for support rather than punishment, such evidence may not see the light of day (Honig 2003; Marsh 2002).

Send time and models, not just money. Central offices seem to have access to new funding for data systems and other computer technologies, but they often lack other resources, namely structured time and models of professional practice, that support their use

of evidence. For example, central office administrators often face multiple goals, demands, and priorities that divide their attention (Hannaway 1989; Holley 1980; Peterson 1998) and thus may report that they have little time to "consult the evidence" (Holley 1980).

To use research and data to drive decision making, district administrators must play new and sometimes unfamiliar roles. Demands to use data as part of accountability requirements, for example, call for a shift in orientation in collecting data away from compliance reporting to the government and toward making data accessible to inform ongoing decision making. However, central office administrators don't always have access to models of professional practice that reflect these roles (Burch & Thiem 2004; Reichardt 2000).

External supports seem essential. Various organizations outside central office jurisdictions – including professional associations, reform support providers, intermediary organizations, and research and development agencies, among others – play essential roles in supporting district central office evidence use. These organizations often form "natural channels" through which information flows, because they have credibility with school and central office personnel and the ability to integrate research knowledge with an awareness of local needs and conditions (Corcoran & Rouk 1985; Datnow, Hubbard & Mehan 2002; Honig 2004a, 2004b; Kean, 1981, 1983; Osheka & Campagne 1989; Roberts & Smith 1982). James Spillane (1998) showed that information garnered through ties to professional associations shaped the assumptions and beliefs that district personnel use to interpret information and that such ties were more salient than state policy in shaping districts' instructional agendas.

By contrast, federal and state agencies have mainly mandated the use of evidence, sanctioned the use of specific evidence, invested in particular forms of research, and tied penalties to the failure of district central offices to use evidence. While focusing increased attention on evidence, these steps have not necessarily increased capacity or led to substantive use in district central offices. As Diane Massell (2001) found, state policy provides the conditions to encourage data use, but whether districts embrace the approach depends on *district* conditions, such as whether district staff view outcomes and performance goals as important, relevant, and attainable.

Evidence use and improvement do not operate in a one-to-one relationship. Evidence rarely points to an unambiguous path to improvement.

Selected Implications for Practice

The lessons suggested by the research have important implications for the way central offices use evidence, leading to a series of recommendations.

Be realistic about what evidence offers and how it functions in decision making. District central offices and the organizations that support them should understand that evidence use and improvement do not operate in a one-to-one relationship and that evidence rarely points to an unambiguous path to improvement. For central office administrators to make productive use of evidence, they must have opportunities to interpret and translate evidence.

Improve access to "useful" evidence. Evidence is useful when it clearly relates to pressing central office matters, is available in a timely manner, and comes in relatively straightforward forms. However, evidence translators should take care to ensure that their reformulations of evidence are simple enough that they will be used – but not so simple that they strip their original source of potentially valuable information. Educational researchers may have particularly important roles to play in this process by asking research questions that relate more closely to central office practice and by translating (or collaborating with someone to translate) their research into more accessible formats.

Support nontechnical capacity: professional models and collaboration. Improving evidence use isn't simply a matter of building a better management information system. Central office staff administrators need models of professional practice that include evidence use as part of their day-to-day routines – especially those who may perceive evidence use as a job for "experts." These routines include substantive collaboration among central office administrators overall and, in particular, among those who may be designated to acquire evidence and those charged with incorporating it into central office decisions.

Note: This paper is based on a book chapter, written by the authors and Mary Kay Stein, forthcoming in *Conference Volume on Evidence Use* [provisional title]. The authors thank the MacArthur Network for Teaching and Learning (funded by the John D. and Catherine T. MacArthur Foundation) for their support of this research. The authors also thank Carolina Belalcazar, Mika Yamashita, Marlene Mullin, and Scott Mergl for their various contributions to this article. Any opinions expressed herein are those of the authors and do not necessarily represent the views of the MacArthur Foundation.

References

Bickel, W. E., and W. W. Cooley. 1985. "Decision-Oriented Educational Research in School Districts: The Role of Dissemination Processes," *Studies in Educational Evaluation* 11:183–203.

Boeckx, J. K. 1994. "Developing Community Support for Outcome-Based Education," *School Administrator* (September), 24–25.

Burch, P. E., and C. H. Thiem. 2004. "Private Organizations, School Districts, and the Enterprise of High-Stakes Accountability." Unpublished manuscript, Educational Policy Studies Department, University of Wisconsin–Madison.

Coburn, C. E., M. I. Honig, and M. Stein. Forthcoming. "What's the Evidence on District's Use of Evidence?" *In Conference Volume on Evidence Use* [provisional title]. Chicago: MacArthur Network on Teaching and Learning.

Coburn, C. E., and J. E. Talbert. Forthcoming. "Conceptions of Evidence-Based Practice in School Districts: Mapping the Terrain." Paper to be presented at the annual meeting of the American Educational Research Association, Montreal (April 2005).

Corcoran, T., S. H. Fuhrman, and C. L. Belcher. 2001. "The District Role in Instructional Improvement," *Phi Delta Kappan* 83, no. 1:78–84.

Corcoran, T., and U. Rouk. 1985. *Using Natural Channels for School Improvement: A Report on Four Years of Urban Development Program.* Philadelphia: Research for Better Schools.

Datnow, A., L. Hubbard, and H. Mehan. 2002. *Extending Educational Reform: From One School to Many*. London: RoutledgeFalmer.

Englert, R. M., M. H. Kean, and J. D. Scribner. 1977. "Politics of Program Evaluation in Large City School Districts," *Education and Urban Society* 9, no. 4:429–450.

Fullan, M. 1980. *The Role of Human Agents Internal to School Districts in Knowledge Utilization.* Ontario, Canada: The Ontario Institute for Studies in Education.

Hannaway, J. 1989. *Managers Managing: The Workings of an Administrative System.* New York: Oxford University Press.

Holley, F. 1980. "Evaluation Utilization: Is It Easier to Move a Mountain Than a Molehill?" Paper presented at the annual meeting of the American Educational Research Association.

Honig, M. I. 2001. "Managing Ambiguity: The Implementation of Complex Education Policy." Unpublished doctoral dissertation, Stanford University.

Honig, M. I. 2003. "Building Policy from Practice: Central Office Administrators' Roles and Capacity in Collaborative Policy Implementation," *Educational Administration Quarterly* 39, no. 3:292–338.

Honig, M. I. 2004a. "District Central Office–Community Partnerships: From Contracts to Collaboration to Control." In *Theory and Research in Educational Administration,* vol. 3, edited by W. Hoy and C. Miskel. Greenwich, CT: Information Age Publishing.

Honig, M. I. 2004b. "The New Middle Management: The Role of Intermediary Organizations in Complex Education Policy Implementation," *Educational Evaluation and Policy Analysis* 8, no. 4:527–561.

Kean, M. H. 1981. "Institutionalizing Educational Productivity." Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Kean, M. H. 1983. "Translating Research and Evaluation into Educational Policy," *Urban Review* 15, no. 3:187–200.

Kennedy, M. M. 1982a. "Evidence and Decision." In *Working Knowledge and Other Essays,* edited by M. M. Kennedy, pp. 59–103. Cambridge, MA: Huron Institute.

Kennedy, M. M. 1982b. "Working Knowledge." In *Working Knowledge and Other Essays,* edited by M. M. Kennedy, pp. 1–28. Cambridge, MA: Huron Institute.

Marsh, J. A. 2002. "Democratic Dilemmas: Joint Work, Education Politics, and Community." Unpublished doctoral dissertation, Stanford University.

Massell, D. 2001. "The Theory and Practice of Using Data to Build Capacity: State and Local Strategies and Their Effects." In *From the Capitol to the Classroom: Standards-Based Reform in the States,* edited by S. H. Fuhrman, pp. 148–169. One hundredth yearbook of the National Society for the Study of Education. Chicago: NSSE and University of Chicago Press.

Massell, D., and M. E. Goertz. 2002. "District Strategies for Building Instructional Capacity." In *School Districts and Instructional Renewal,* edited by A. M. Hightower, M. S. Knapp, J. A. Marsh, and M. W. McLaughlin, pp. 43–60. New York: Teachers College Press.

Nafziger, D., J. Griffith, and P. Goren. 1985. "San Diego City Schools' Research Cooperative: A Marriage of Practitioners, Researchers, and Community Members." Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Newman, D. L., R. D. Brown, and L. S. Rivers. 1983. "Locus of Control and Evaluation Use: Does Sense of Control Affect Information Needs and Decision Making?" *Studies in Educational Evaluation* 9:77–88.

O'Day, J. A. 2002. "Complexity, Accountability, and School Improvement," *Harvard Educational Review* 72:1–31.

Osheka, J. R., and D. W. Campagne. 1989. "Power, Responsibility, Control, and Accountability: A Case Study of Decision Making in an Implementation Effort in a City School District." Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Peterson, G. J. 1998. "Demonstrated Actions of Instructional Leaders: A Case Study of Five Superintendents." Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Reichardt, R. 2000. *The State's Role in Supporting Data Driven Decision-Making: A View of Wyoming.* Aurora, CO: Mid-Continent Research for Education and Learning.

Roberts, J. M. E., and S. C. Smith. 1982. *Instructional Improvement: A System-wide Approach.* Philadelphia, PA: Research for Better Schools.

Robinson, C. M. 1988. "Improving Education through the Application of Measurement and Research: A Practitioner's Perspective," *Applied Measurement in Education* 1, no. 1:53–65.

Spillane, J. P. 1998. "State Policy and the Non-monolithic Nature of the Local School District: Organizational and Professional Considerations," *American Educational Research Journal* 35, no. 1:33–63.

Spillane, J. P., B. J. Reiser, and T. Reimer. 2002. "Policy Implementation and Cognition: Reframing and Refocusing Implementation Research," *Review of Educational Research* 72, no. 3:387–431.

U.S. Department of Education. 2002a. *No Child Left Behind: A Desktop Reference.* Washington, DC: U.S. GPO. Available online at <www.ed.gov/admins/lead/account/nclbreference/index.html>.

U.S. Department of Education. 2002b. *U.S. Department of Education Strategic Plan 2002–2007.* Washington, DC: U.S. GPO. Available online at: <www.ed.gov/about/reports/strat/plan2002-07/index.html>

West, R. F., and C. Rhoton. 1994. "School District Administrators' Perceptions of Educational Research and Barriers to Research Utilization," *ERS Spectrum,* 23–30.

Data, Observations, and Tough Questions: A Journalist's Evidence of School Quality

Dale Mezzacappa

Using research and their own techniques for gathering information, journalists are becoming sophisticated in determining whether school practices are effective. And they are providing readers with a better understanding of what "best practice" is – and what it isn't.

The first time I stepped into a classroom as an education reporter was in late 1986, when I had just been appointed to cover the Philadelphia schools. I was writing a story about the district's new policy of giving report cards to kindergarten students.

It was my first story on the beat. I hadn't had time to read research studies on education so I could get ideas for my vast assignment – covering a major urban school district with some 200,000 students and 250 schools.

I was, however, the mother of a three-year-old. And the report-card policy was causing a buzz in preschool circles. I plunged into the story asking the questions a mother would. How will this affect what students do all day? What will they be evaluated on? Will it put pressure on them? Does this signal a more academic focus?

I trusted my instincts to know what a good kindergarten classroom looked like: activity corners; pictures, letters, words, and shapes everywhere; plenty of books; and no desks in rows. I had my ideas, as well, about what a kindergarten report card should consist of: lots of commentary, developmental checklists, and, most definitely, no letter grades.

As it turned out, before ever taking the assignment, I had absorbed the consensus of the best research on early childhood education.

Eighteen years later, asked to reflect on how education reporters use research to inform our reporting on best practices in schools, I find it somewhat difficult to give a straightforward answer. We use research constantly, but one could also argue that we don't use it enough. Sometimes we use it well, and sometimes we don't. And we do our work in a climate that disparages much education research as politically slanted or not sufficiently rigorous.

Because we write for the general reader, education journalists help shape opinion and understanding about the real issues and obstacles in education reform. Bad teaching? Misguided curriculum? Inequitable distribution of resources? Concentrated poverty? Our own conclusions about these and other issues are shaped by our understanding of the professional research, yes, but also by our life experiences and our reporting. Combined, these three sources of evidence give us a pretty good sense of what works.

Dale Mezzacappa is an education reporter for the Philadelphia Inquirer.

Using the Tools of the Trade

At the *Philadelphia Inquirer,* where I work, the editors and reporters on the SMASH (science, medicine, aging, and social health) desk meet weekly to go over journal articles and plan coverage that is based on what the articles say. The education reporters, who have gone back and forth from working for regional and city editors to being organized under a single education editor, don't do that.

This is partly because education research doesn't have the weightiness of a *New England Journal of Medicine* or *Journal of the American Medical Association.* (It remains to be seen whether the current effort to make educational research more "scientific" through the Institute of Education Sciences – the research arm of the U.S. Department of Education – will succeed in creating an authoritative voice for research or devolve into politics, and how this change will affect journalists' work.) Our relationship to and use of research is much more ad hoc, partially because it is driven by the practices and policies of the school districts we cover. But it is multifaceted and always evolving and gaining sophistication.

Education writers learn pretty quickly that for every study that points to the value of an approach or school characteristic there is another study that reaches a different conclusion.

We decide what reports to publicize and write news stories to bring these reports' findings to a wider audience. We evaluate the research, trying to cull the good from the bad. We ask school districts for research to back up the effectiveness of programs on which they are spending money. We mine good research for story ideas, keeping in close touch with researchers whose work is reliable, complete, and relevant to the districts we cover.

We must sift through studies that often reach opposite conclusions. Education writers learn pretty quickly that for every study that points to the value of an approach or school characteristic – tracking and ability grouping, phonics, bilingual education, cooperative learning, the correlation between spending and results, and vouchers, to name a few crucial questions – there is another study that reaches a different conclusion. But in deciding what research to highlight and how to interpret it, we rely on the journalistic skills we would bring to any story – knowing how to judge the credibility of our sources and the point of view they represent.

Most reporters and editors are generalists – they're not trained in social

science or in the fine points of statistical analysis. (Although organizations like the Education Writers Association, the Poynter Institute, and the Hechinger Institute on Education and the Media work very hard to fill those gaps.) Many education reporters cycle through the beat and move on quickly; others make a career of it.

Journalists are, however, quick studies who learn how to dig out information that will get to the crux of the issue. And, contrary to public perception, we have no axe to grind. We have no personal stake in whether a program is successful or not – unlike, for instance, the district that spent millions to put it in place or the educator who has invested a good deal or his or her life in promoting a particular approach to schooling. And, also contrary to the popular view, we don't have an interest in a specific policy or ideological position. We don't, for example, have a stake in whether charter schools are better able to educate low-income children or are bad public policy. But we *are* interested if a district is spending millions on a curriculum or a program that isn't getting results – or if a state or nation is pursuing an agenda that is counterproductive.

We are in a good position, then, to provide a clear eye on research studies and present their findings in understandable ways to the general public. One place we try to begin is by looking at what the studies actually say. For instance, when California decided to invest heavily in reducing class size to twenty students, reporters scrambled to summarize the reports saying that smaller class size improves student achievement. But the definitive Tennessee STAR study (HEROS 2003) said that reduced class size makes a difference when the cutoff is seventeen; the larger classes had twenty-two and twenty-three students. So, it is up to us to make sure the general reader understands this. When a school district we cover invests in an after-school program, and we find a report saying that after-school doesn't impact academic achievement, we have to know whether the particular program described in the study is comparable to the one being put in place. Some reporters do that, and others don't.

There are important, pitched battles in research that it is our obligation to try to sort out. Today, many of those go to the heart of the premise and implementation of No Child Left Behind. Did the "Texas Miracle" occur at the expense of high drop-out rates? The *New York Times*, following up on data collected by researchers in Texas and Boston, wrote about how schools fudged their drop-out data and found individual dropouts who were not counted as such.

More recently, the *Dallas Morning News* found evidence of cheating in dozens of Texas schools by doing an

analysis of publicly available state test-score results that the state itself declined to do. Reporters Josh Benton and Holly Hacker didn't assume that high-poverty schools were incapable of dramatic improvement. But, by careful analysis of data, using the actual scaled scores rather than proficiency rates, they found schools with suspicious patterns. In one, for instance, fourth-graders ranked in the bottom 2 percent in the state, while nearly all fifth-graders – the grade that counted for the purposes of determining whether a school made adequate yearly progress – not only reached proficiency, but got perfect scores. The same students finished below the national average in the Stanford Achievement Test that they had taken eight weeks previously.

Beyond Statistics: Good Stories

As these examples show, education journalists have come a long way in how we analyze and present data about schools. In 1997, well before No Child Left Behind, the *Philadelphia Inquirer* became one of the first newspapers in the country to publish a yearly "School Report Card," a separate section that compiled statistical information about education in the region. That first effort listed the test scores for every school in the area, along with other data that included teacher salaries, average class size, per pupil spending, racial breakdown, and the percentage of students in poverty. To help parents make sense of the numbers, the schools were divided into quintiles, with the highest-scoring schools at the top and the lowest-scoring schools at the bottom. The Philadelphia schools were ranked separately from those in the suburbs, under the assumption that it was not fair or useful to compare the two.

By the next year, we had a more advanced system for rating schools that acknowledged the correlation of poverty and achievement, allowing readers to see how a school fared compared to all schools in the region and compared with demographically similar schools. Readers could not only see that there were high-poverty schools doing well, but that some schools in well-off areas performed poorly.

While it is becoming more important to analyze statistics, reporters also go behind the statistics to make them come alive. The lifeblood of the newspaper reporter is the good story – the compelling narrative that will draw the general reader into what can be a difficult subject. That is, in many ways, what we do best.

While it is becoming more important to analyze statistics, reporters also go behind the statistics to make them come alive. The lifeblood of the newspaper reporter is the good story.

When the mayor and chancellor of the New York City public schools decided to hold back any third-grade student who couldn't read on grade level, *New York Times* reporter Jacques Steinberg decided to spend a year in one third-grade classroom and chronicled the efforts of the teacher and students to reach the goal.

Steinberg didn't specifically cite research in his series. He didn't give long explanations of the conflict over whole language versus phonics. But the entire series was informed by his

Increasingly, with more interactive databases and computer-assisted reporting capabilities, reporters do research themselves that often breaks ground or pushes public policy.

knowledge of the research. He pointed out that the class had thirty children, most with time-consuming needs. He visited the home of a student with few books and fewer places where an eight-year-old could get peace and quiet to read – but also caught her reading flawlessly to her mother from a chapter book, which helped the teacher probe and advance the abilities of the shy child.

Throughout the nine stories, which won the Education Writers Association grand prize that year, Steinberg consistently focused in on issues and scenes that illuminated what a complex, private, and varied process learning to read is for each child. He even recounted the childhood struggles of the teacher himself, who, by his own description, didn't become an independent reader until high school. Readers grabbed by such details are likely to learn more about the difficulties faced by a dedicated teacher working in a crowded classroom with needy children and about how there are no foolproof templates teachers can simply follow to teach all children to read.

Another top-notch education journalist, Jay Mathews of the *Washington Post,* took a slightly different approach when school officials in Washington, D.C., decreed that students who didn't pass a single test wouldn't get promoted. He found a young girl who repeated a grade after missing the cutoff by one question and spent time following her educational journey. Again, readers are more likely to get the point about the danger in relying too heavily on tests from getting to know a child who was so drastically affected by the policy.

More recently, reporter Stephanie Banchero of the *Chicago Tribune* followed one little girl whose mother decided to exercise her right under No Child Left Behind to transfer her from the struggling neighborhood school to a better one across town.

Banchero described the home of the student, where two sisters and their four children lived with their grandmother. She explained with sensitivity that, while the mother wanted something better for her children, her daily travails prevented her from getting the children to school every day. The reporter sat in the classroom with the child and recounted how hard she worked and how she made progress with a skilled teacher in a smaller, more diverse class.

Ultimately, though, the two-hour commute proved too difficult and the mother took the children out and transferred them to a school closer to home. Sadly, she thought that this school, too, was one of the better ones receiving students from the underperforming schools. But she was mistaken. The series raised questions about whether the "choice" portion of the law can work as intended without more resources and help for families.

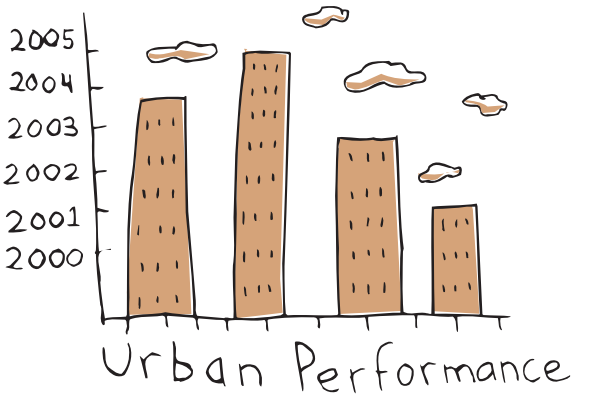
Such anecdotal accounts, of course, aren't the final word on the success or failure of a policy, but they are crucial to understanding it. In the best use of research, we combine anecdotes with statistical analysis.

Increasingly Sophisticated Reporting

Fortunately, education reporters are in a much better position now to undertake sophisticated analyses of school quality and effectiveness than I was when I began on the beat in 1986. Thanks to No Child Left Behind and similar policies, journalists, along with everyone else, have far more information and statistics available to us. And, increasingly, with more interactive databases and computer-assisted reporting capabilities, we do research ourselves that often breaks ground or pushes public policy.

At the *Inquirer,* we are combining statistical analysis with in-depth reporting to provide our readers with an expanded notion of what constitutes a top-quality education. This year, as part of our annual School Report Card, we used test-score information to find a Title I Philadelphia school in which more than 80 percent of students, white and black, are proficient in math and reading. We are planning to profile that school, asking a question that was inconceivable that first year, when city and suburban schools were not compared to each other: Could this city school be better than those in the suburbs, where more than 90 percent of students reach proficiency, but where the racial achievement gap is wide?

We've written about this school before. It is characterized by a stable staff; a veteran principal who was able to choose many of her teachers, regardless of seniority; a constant focus on instruction; small classes; collaboration



among teachers; an active parent community; respect among the adults; and respect between the adults and students. Expectations are uniformly high.

Not surprisingly, these are all the characteristics that, as the Consortium on Chicago School Research found, support student learning: leadership, quality instruction, parent and community partnerships, student-centered learning, and professional capacity.

What I've learned from covering schools is that virtually all research about whether a program is succeeding or not or about whether a school is helping students make progress or not boils down to one issue. To borrow a Clinton-era phrase, "It's the teachers, stupid."

That's why, for the last several years, I've tried to spend a lot of time focusing on issues relating to teacher quality. For that, I've worked with researchers at the University of Pennsylvania to analyze data on Philadelphia teachers, pinpoint schools with revolving-door staffs, and describe what the learning environment is like for the average student. Other newspapers are doing similar projects.

Sarasota's *Herald Tribune*, in a recent series, obtained data on teacher certification test scores from the state of Florida under the Freedom of Information Act and analyzed twenty years' worth of exams. The data showed that half a million children sit in classes with teachers who have failed basic skills tests at least once, and that the overwhelming majority of those children are in poor, minority communities. The newspaper took the state to task for not pursuing policies that would lead to a more equitable distribution of teacher quality and attract more academically prepared people to teaching.

The reporters were careful to say that some people who have trouble with such tests can be good teachers and quoted many administrators saying so. But, ultimately, the newspaper's work forced the state of Florida to analyze whether there is a correlation between the achievement of students and their teachers' history with certification tests. It found that those taught by teachers who failed certification tests more than three times suffered academically.

Of course, journalists are not researchers. But, as the Sarasota *Herald Tribune* series shows, journalists are becoming more adept at using statistical data, better at observing and describing classroom realities, and more willing to ask pointed questions about long-accepted policies and practices. And, in doing so, we are providing our readers with a better understanding of what "best practice" is – and what it isn't.

Reference

Health & Education Research Operative Services, Inc. 2003. *Project STAR: Class Size Research.* Lebanon, Tennessee: HEROS. Available online at <www.heros-inc.org/classsizeresearch.htm>.

Annenberg
Institute for
School Reform

AT BROWN UNIVERSITY

Annenberg Institute for School Reform
Brown University
Box 1985
Providence, Rhode Island 02912

Non Profit Org.
US Postage
PAID
Providence, RI
Permit #202