Expanding the Scope of Implementation Research in Education to Inform Design

William R. Penuel

SRI International

Kenneth A. Frank

Michigan State University

Barry J. Fishman

University of Michigan

Nora Sabelli

Britte Cheng

SRI International

This material is based in part upon work supported by the National Science Foundation under Grant Numbers SBE-0354453, and BCS-0624307. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Abstract

In this paper, we present a vision for implementation research in education that can inform all stages of program development. We present examples of implementation research that have performed three potential functions can serve to have a greater impact on practice: (1) identify problems of practice that can become targets of design, (2) bridge gaps between current system capacity and ambitious visions for change reflected in curricular reforms, and (3) test experimentally contrasting models of implementation support. These functions expand the scope for implementation research in education, since the functions rarely are central in sociological analyses of implementation or in current approaches to measuring and analyzing fidelity of implementation advocated by proponents of experimental research on program efficacy. In addition, by taking both curricular interventions and contexts of implementation as objects of design and study, this kind of forward-looking implementation research can inform the process of system-level changes in education in ways that improve implementation of curricular interventions.

Expanding the Scope of Implementation Research in Education to Inform Design

The results of recent large-scale, experimental studies to identify effective programs and curriculum materials have been disheartening to policymakers and researchers alike. The findings from large studies of federally-funded programs completed in the last five years in reading (Gamse et al., 2008), mathematics (Agodini et al., 2009), educational technology (Dynarski et al., 2007), and afterschool programming (James-Burdumy, Dynarski, & Deke, 2007) are all similar, in that all have all found either very small or no positive impacts on student achievement. In response to these findings, critics have raised questions about the quality and depth with which programs were implemented (Bissell et al., 2003; Mahoney & Zigler, 2006). How can researchers conclude, these critics argue, that programs do not work, if they have not been implemented well or consistently under different conditions?

Some researchers argue that answers to such questions can be developed within the context of large-scale experiments, but to date, the field has developed no consensus about methods for interpreting implementation results. Researchers can and do measure implementation within experimental studies, as was done in each of the studies cited above. In addition, researchers can and do model how implementation processes are related to outcomes in the contexts of experimental study, which can help inform refinements to the design of programs (Judd & Kenny, 1981; Krull & MacKinnon, 1999; MacKinnon & Dwyer, 1993; O'Donnell, 2008). But estimates of the strength of associations between implementation and outcomes in experiments in some cases may be difficult to interpret, and deriving causal inferences from those results is tricky and may

be contested (Angrist, Imbens, & Rubin, 1996; Werner, 2004). Implementation processes are endogenous to experimental studies; programs' effectiveness is inextricably intertwined with the ease of implementation, the ways that programs are enacted, and the contexts of implementation, much in the way that the effectiveness of a particular diet is bound up with how easy it is for people to follow (cf., Dansinger, Gleason, Griffith, Selker, & Schaefer, 2005). Further, if programs are difficult for teachers to implement well and have not already been adapted to a range of contexts and shown to be reliably usable, researchers may not uncover any significant associations between implementation quality and outcomes (McDonald, 2009).

An alternative approach is to foster more research on implementation and its contexts at all stages of program development, not just at the point at which scale up begins. In developing a new program, for example, developers could begin by investigating the contexts where they plan to try out the program, so they have a better sense of new capacities local schools and districts will need to develop to implement the new program. As programs transition from either the laboratory to the classroom, or from 1-2 classrooms to many, implementation research can help identify gaps in program guidelines and specifications, a more refined sense of professional development needs, and likely variations in implementation that could be associated with differences in program effectiveness. Finally, at the efficacy and scale-up phases of development, analyses of how implementation mediates impacts can identify ways to strengthen programs and suggest designs for future experimental studies.

The idea that implementation research should be more integral to the design and refinement of programs or infrastructures needed to support implementation is not new

(see, especially Elmore, 1980), but few programs use implementation research in this way, and there is little accumulation of knowledge of implementation processes across programs. Strong disciplinary boundaries exist that separate policy research from scholars engaged in early-stage research and development efforts, who are more focused on using such efforts to generate new insights into the science of learning. These latter scholars' own research and development efforts often entail grapping with issues of how to promote learning in real classroom environments (e.g., Barab & Luehmann, 2003), but the insights they develop from their research about implementation are only infrequently a focus in their publications. One solution for researchers to adopt is to rely on different kinds of experts for each stage of program design and development, as has been done in other fields (Sloane, 2008). But such approach limits the possibility that implementation research could develop useful knowledge of basic processes, such as how teachers adapt programs to their local contexts, which are implicated across different stages of program development.

In this paper, we present three new functions for implementation research at different stages of program development that can inform the design of effective programs that improve student learning. These functions emphasize ways that implementation research can inform the earliest stages of program design, program refinement, and the identification of effective models for scaling-up programs. By expanding the range of implementation research, we argue, implementation research can advance the science of learning as the study of linked systems of curriculum materials, learning supports for teachers and school leaders, and organizational forms and processes needed to support enactment.

Past and Current Implementation Research in Education

Problems of program implementation have been a focus of education research for decades. In the late 1950s, when the National Science Foundation first funded the design of instructional materials for schools, curriculum developers became frustrated by what they saw as teachers' failure to enact curricula in ways that reflected an understanding of the structure of scientific disciplines (Bruner, 1960). Later, in the 1970s, policy researchers suggested that adaptations teachers make to curriculum materials are necessary and always occur, to meet the needs of students and demands of local contexts (Berman & McLaughlin, 1975; McLaughlin, 1976). Since that time, researchers have remained concerned with the degree to which teachers' adaptations are congruent with designers' intentions, seeking to distinguish "creative transformations" of curriculum materials from "lethal mutations" in teachers' enactments (A. L. Brown & Campione, 1996; M. W. Brown & Edelson, 2001). Researchers have also investigated the extent to which poor implementation quality can diminish the strength of an intervention, making it less likely that investigators will be able to detect significant effects of programs (Cordray & Pion, 2006).

For most of that time, implementation studies were conducted principally by sociologists and political scientists in education, and their research has focused developing explanations for variability in implementation informed by theories from those disciplines. A recent example is Rowan and Miller's (2007) study of the efficacy of three different school reform models' approaches to supporting changes in teaching. The study used agency theory from sociology and political science (Eisenhardt, 1989; Emirbayer & Mische, 1998) as a lens for exploring strategies program developers,

policymakers and other educational leaders have for resolving so-called "agency dilemmas" that derive from the fact that in education, the agents who develop policies and programs do not always share goals with but are dependent on teachers as agents to implement them. The Rowan and Miller study identified some types of controls on teachers' implementation—essentially different strategies for addressing agency dilemmas"—as more likely to yield reliable patterns of implementation than others. Their conclusion that relying on collegial influence (professional controls) and implementation specification and monitoring (procedural controls) has potential, if not yet realized, implications for the design not only of programs and policies but also of mechanisms to support their implementation.

In the past decade, researchers engaged in curriculum development have become more involved in implementation research. Spurred by calls for more rigorous research to identify effective programs and conditions of their effectiveness (National Research Council, 2002; President's Committee of Advisors on Science and Technology Panel on Educational Technology, 1997) and supported by new funding streams (e.g., NSF's Interagency Education Research Initiative and the U.S. Department of Education's NCER field-initiated grant programs), program and curriculum developers have sought primarily to collect and analyze data on implementation fidelity. The focus on fidelity, which can be defined as the degree to which teachers in enacting a curriculum adhere to its sequence of activities in ways that have integrity to the principles of designers (O'Donnell, 2008), is in the service of defining conditions under which programs may be more or less effective. Researchers hypothesize that with lower treatment integrity, the difference between treatment and control groups' performance in a randomized control trial may be

diminished (Cordray & Pion, 2006). Thus, in the context of experimental research on curricular effectiveness, both measuring and promoting implementation fidelity are likely to be critical aspects of research and development efforts.

Both the sociological approaches to implementation research and stage model of research adapted from NIH provide clear guidance to the field about how such research can inform design, especially in the early stages of development. Until recently, sociological approaches have focused mainly on organizational and institutional processes affecting education but not on processes that take place inside classrooms (see, Spillane & Jennings, 1997, for critiques). With the development and validation of new measures of instruction (Correnti, 2007; Rowan, Camburn, & Correnti, 2004; Rowan, Harrison, & Hayes, 2004), sociological analyses such as the one cited above by Rowan and Miller (2007) have been able to relate organizational processes to instruction, but these measures have mostly been applied to programs that exist at scale. The Institute of Education Sciences at the U.S. Department of Education's guidance about the kind of implementation research needed for projects seeking development phase (Goal 2) funding recognizes that development is likely to take place in a small number of classrooms, but the guidance is vague about appropriate methods for studying implementation and silent about the kinds of organizational supports that developers may need to consider creating as part of their interventions:

> Feasibility of implementation might be addressed, for example, with evidence demonstrating that the intervention can be implemented with fidelity in a few authentic education delivery settings that represent the type of settings (e.g., classrooms) for which the intervention is intended.

> Feasibility should be demonstrated on a small sample of users (e.g., teachers, students) who are like those for whom the product is intended and should show that they can utilize or implement the intervention in the way that the developer intends the intervention to be implemented. (Institute of Education Sciences, 2009, p. 64)

Missing also particularly from the guidance provided by the Institute of Education Sciences is a recognition that to succeed, most ambitious curricula that require teachers make significant changes to their practice imply the need for new supports for teacher learning and new organizational forms that build schools' and districts' capacity for change. Ignoring the learning needs of teachers and current capacity of schools and districts in developing interventions is likely to lead to the development of innovations that are neither usable nor useful to teachers (Blumenfeld, Fishman, Krajcik, Marx, & Soloway, 2000; Fishman & Krajcik, 2003). Conversely, designing only programs that fit within the current organizational capacities of schools and districts and institutional arrangements of schooling is not likely to lead to identifying powerful programs and practices that can improve the achievement of all students. Needed, then are models for conducting implementation research that can inform the design not just of curriculum materials but also the learning supports for teachers and organizational reforms that may be requires to enact materials with integrity to designers' intentions.

Three New Functions for Implementation Research in Education

We envision three new functions that implementation research can play in education to inform the design of curricular interventions and the supports required for their enactment. The first function is in identifying opportunities within current systems of

educational practice where new designs may have a potential to improve teaching and learning. This form of implementation research takes place at the earliest stages of design, when developers need to develop an understanding of the contexts for which they are designing interventions. The second function is in informing program refinement, where researchers may be engaged in refining the curriculum and/or aspects of implementation support on the basis of identified gaps between the current capacity of systems and intentions of ambitious curricular reforms. The third function is testing models of implementation support through experimental research. This last function fits within the NIH-inspired stage model employed by the Institute of Education Sciences for studying leadership and teacher quality, but we highlight it here since there is no explicit focus in those funding streams on the relationship between professional development and curriculum implementation. Our example highlights how the two can be studied in the context of an efficacy study comparing different approaches to professional development.

In this section of the paper, we describe these functions of a forward-looking approach to implementation research in greater detail. We present examples of educational research that is carrying out each function, with attention to the context for the research, focus of design efforts, methods and findings of implementation research, and how research informed the design or refinement of curricular interventions and systems of support for those interventions.

Identifying Problems of Practice as Basis for Design

A fundamental challenge to scaling curricular interventions developed in a laboratory or in so-called "hothouses" is that teachers do not perceive them to be useful or usable (Blumenfeld et al., 2000; Fishman & Krajcik, 2003). The *usability* of curricular

intervention depends in large part on teachers' perceptions of the fit of curriculum to their students and the goals for learning set by their districts and states (Penuel, Fishman, Yamaguchi, & Gallagher, 2007). Such judgments can also be influenced by normative pressure from colleagues (Coburn & Russell, 2008; Frank, Zhao, & Borman, 2004) and by the level of resources provided to support implementation, including professional development and time to plan for implementation, which signal to teachers the salience and importance of implementation to leaders and policymakers (Penuel, Fishman, Gallagher, Korbak, & Lopez-Prado, 2009). The *usability* of particular curricular interventions can be thought of largely as a function of the fit between the local capacity to support a curricular intervention (e.g., by providing needed resources, access to expertise useful for implementation) and the requirements of the curricular intervention itself. The greater the gap between the current capacity of a school or district and curricular demands, the less usable the design will be for that school or district (Blumenfeld et al., 2000).

Models for conducting implementation research at the early stages of design to improve the likely usefulness and usability of interventions can be found within the field of software engineering. In that field, implementation research has influenced design through a family of practices that go by various names: contextual design (Beyer & Holtzblatt, 1997), rapid prototyping (Gorden & Bieman, 1995), and rapid ethnography (Millen, 2000). Although the specifics of each practice differs, this family of practices all employ social science methods "up front" to identify the dilemmas and problems of practice that could inform the design and development of products. In the past decade, the influence of social science methods has increased tremendously, particularly as a result of

the success of the design processes employed by such firms as IDEO, which draw extensively on ethnographic methods (Hargadon & Sutton, 1997), and which have led to designs that have helped their clients sell many products (Taylor, 2005). The social science methods adopted by IDEO and other firms would typically not yield findings that are publishable in anthropological journals, but that is not their aim. Rather their chief purpose is to inform processes of design with sufficient details about ways contexts and users differ in ways that would not be possible without "up front" research. IDEOs' and other firms that draw from social science practices draw on a long-standing tradition of participatory design from Scandinavia, a practice institutionalized in labor laws in those countries that mandate participation by users in the design process (Ehn, 1992).

What software engineering has to offer implementation research in education are methods of design and research for the early stages of developing a program or curriculum. Design methods that are participatory in that they include end-users (often teachers) in the process have the potential to increase the usability and usefulness of educational materials (Penuel, Roschelle, & Shechtman, 2007; Shrader, Williams, Lachance-Whitcomb, Finn, & Gomez, 2001; Slotta & Peters, 2008). In addition, specific tools from participatory design, such as the use of rapid prototyping and the development of use-cases and scenarios, can be used to make design more efficient and tailored to different users' contexts (Tripp & Bichelmeyer, 1990). Rapid-ethnographic studies can develop knowledge of the design constraints and requirements for a particular setting, to the extent that it identifies the institutional and organizational capacities, goals, and conflicts in that setting (Roschelle, Penuel, Yarnall, & Shechtman, 2005). In addition, these methods can surface critical work practices that already exist at scale and that can

be augmented by new technologies, either by making them more efficient or by improving them (Penuel, Lynn, & Berger, 2006).

An illustration of the potential of such an approach to inform the early stages of design of an intervention in education is Wireless Generation's mCLASS (mobile classroom assessment) system. The system for early literacy assessment consists of two components. Teachers first use handheld computers to conduct one-on-one observational assessments using the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) with individual students. They then transfer the resulting data through a cable to a centralized database, where the collected information is compiled automatically into reports on individual students, classrooms, schools, and districts, and on longitudinal student progress. Teachers, administrators, and other stakeholders can access these results through a secure Web site as soon as the data are uploaded; the current system also includes specific recommendations for teachers related to instructional groupings based on student data. In supporting the rapid collection, processing, and interpretation of data on student skills, the assessments and reports are designed to serve a number of purposes, including screening students for risks, adjusting and individualizing instruction, and to evaluate individuals' and schools' progress towards standards.

Both the scale (over 1 million users) and evidence collected from studies of the use of the mCLASS assessment system (Hupert & Heinze, 2006; Hupert, Heinze, Gunn, & Stewart, 2007) suggest it is an extraordinarily useful and usable system for teachers. Using the system increased the efficiency of conducting assessment and accuracy on timed components of the assessment, and use decreased administrative tasks linked to managing and reporting data. In addition, teachers reported the data were more relevant,

because the data collected reported more current levels of student understanding than did end-of-year examinations of student performance.

The research that the company did early in the development of the mCLASS system was critical to producing these kinds of efficiencies and in ensuring that teachers would perceive the tool to be useful. At the suggestion of researchers at EDC's Center for Technology in Learning, where the company was incubated, company founders spent time observing classrooms before undertaking design work. Their visits identified a common practice among teachers in the early elementary grades, making running records of students' reading fluency, which they felt was burdensome and could be improved with supporting handheld computer software. At that time, they observed teachers taking hours not to collect the data but to manage and enter data so that it could be useful in the ways that designers had intended. The research they conducted was not formal ethnographic research, but by observing problems of practice and being attuned to the affordances of the technological tools at their disposal, the company founders were able to turn an observation into a usable, useful design idea.

The usefulness of the tool was further aided by factors outside the control of the company, but to which the company successfully adapted. The development of the mCLASS system in 2001 intersected with the passage of the *No Child Left Behind Act of 2001,* and Title I, Part B of the federal legislation requires teachers to conduct regular screening, diagnostic, and progress monitoring assessments of K-3 students' reading skills, and to use the results to assign students to instructional programs. Subsequent regulations recommend specific assessments for these purposes (Kame'enui, 2002). This legislation increased the need for literacy assessments that can be administered quickly

and that provide results that teachers can easily interpret and use. Although the original running records the mCLASS system was designed to support did not become part of the new literacy assessment regime, an alternate system (DIBELS) did, and Wireless Generation shifted its assessment content to match the new requirements, ensuring its continued relevance to a broad number of teachers.

The mCLASS handheld-to-Web system is designed to meet these needs by streamlining assessment administration, by automating scoring and reporting, and by producing reports for teachers that suggest instructional groups or tactics to improve achievement. Reports for administrators monitor the effectiveness of teaching, learning, and organizational practices in their schools. The design process for the handheld assessments includes observation of the paper-based assessment administration process, and identification of opportunities to simplify administration while retaining the assessment's content, validity, and structure. For example, several handheld assessments automate a process of diagnostic branching, so that the teacher can focus on observing and recording student responses instead of on referring to a manual for the next question. Scoring results by computer, instead of by hand, increases the efficiency of administering the assessments. Preliminary studies show, for example, that mCLASS assessments cut administration, scoring, and reporting time in half as compared to their paper counterparts (Lynn, 2005; Texas Education Agency, 2003).

Bridging Gaps in System Capacity to Support Ambitious Curricular Change

A challenge for implementation research is to study change efforts in which the goal is not simply to fit innovations within an existing context, but to be able to design and study new contexts as well. Implementation research is needed to support the design of

different models for intervening at the level of the system, to increase the capacity of schools and districts to pursue ambitious reforms for improving teaching and learning (Burkhardt & Schoenfeld, 2003; Cohen, Fuhrman, & Mosher, 2007). Instead of always trying to make curricular interventions fit within the contexts of implementation, developers can sometimes develop supports for interventions and study the effects of those supports on the feasibility of implementation.

Researchers in curriculum, the learning sciences, and the social contexts of teaching who study teacher learning with curriculum are among those researchers most deeply engaged in this form of implementation research in education. Researchers in these fields have been engaged, for example, in analyses of how and when teachers use supports for their own learning about content embedded within curriculum materials (Remillard, 2000; Schneider & Krajcik, 2002), teachers' interactions with colleagues as resources for learning about curriculum and implementation of reforms (Bidwell & Yasumoto, 1997; Frank et al., 2004; Grossman, Wineburg, & Woolworth, 2001; Lieberman & Miller, 2001; Little, 2001), and the aspects of professional development that are strongly associated with curriculum implementation (Penuel, Fishman et al., 2007). For the most part, however, these studies have been conducted with curriculum materials that have already been developed; their findings could inform revisions to those materials, but the studies were not framed explicitly to do so.

A good example of implementation research aimed to inform the revision of supports for teacher professional development is a study conducted by Fishman and colleagues (Fishman, Marx, Best, & Tal, 2003). The study focused on teachers' implementation of a curriculum unit in science developed as part of the Center for Learning Technologies in

Urban Schools (LeTUS) project, and its aim was to conduct research that could improve the quality of the professional development for teachers such that their enactments led to improved student learning. The study focused on a single concept, students' map reading skills necessary to interpret maps of watersheds, and used items from the team's proximal assessments of student learning as the basis for making judgments about where to focus improvements on professional development and about the success of those improvements.

The research reported in the Fishman et al. (2003) article is actually a sequence of multi-method studies. The researchers collected pre- and post- student learning data from their assessments, observation data, surveys, and interviews with teachers during one enactment. Next, the team made refinements to the professional development to give teachers practice with the aspects of the curriculum related to map reading skills and to engage them in analyzing student responses to the items and in brainstorming strategies for developing students' map skills. Then, the research team conducted a second study of the enactment, to determine whether or not the revised professional development had produced improvements in the enactment and in student learning. Results for the second enactment, were significantly better for students, suggesting the promise of their iterative approach to studying implementation and refining professional development on the basis of the research.

Other research teams are working to bridge gaps at the level of the school districts, in educational systems that have adopted ambitious visions for transforming teaching in learning in a particular subject matter domain but that may not have the current capacity to enact those visions. Two separate but related projects led by researchers at the Institute

for Learning at the University of Pittsburgh (Besterfield-Sacre, Resnick, Mehalik, Sherer, & Halverson, 2007; Resnick, Besterfield-Sacre, Mehalik, Sherer, & Halverson, 2007) and at Vanderbilt University (Cobb & Smith, 2008, 2009) are engaging large urban districts in creating systems of support for teacher learning to enact curriculum materials that reflect ambitious new visions for teaching and learning in mathematics. Both projects have introduced to districts the use of validated measures of quality instruction (Matsumura, Garnier, Pascal, & Valdes, 2002), developed by researchers at the University of Pittsburgh, as a means to judging the extent to which the vision the districts have adopted for improving mathematics teaching has been successfully enacted. In addition to collecting data on instruction, the projects are helping the districts gain insight into the professional development systems that exist to support teachers learning about curriculum and instruction and the adequacy of these systems for promoting significant teacher change. Both projects examine teacher knowledge critically, from the perspective of what is needed to enact the particular visions of their district, assuming that this knowledge may not be accessible to those who most need it. The success of these endeavors is being evaluated as part of ongoing research efforts by both teams, but both projects represent attempts to identify feedback loops that employ implementation data as potential bases for district and school leaders to make adjustments to reform strategies they are pursuing.

Testing the Efficacy of Different Models of Implementation Support

Once fully developed, the kinds of systems designed by the LeTUS researchers and by researchers at the University of Pittsburgh and Vanderbilt University in collaboration with school districts can themselves become objects of impact studies. Testing

contrasting models of support for implementation with randomized experiments is essential, if as a field we want to make causal claims about what kinds of support are necessary for teachers to implement particular curricular interventions. At the same time, such studies are likely to be expensive, since they must have adequate power to detect impacts not only on student learning but also on teachers and their practices (Wayne, Yoon, Zhu, Cronen, & Garet, 2008). If researchers also want to conduct correlational analyses within such experiments to develop hypotheses about what supports may be most closely related to changes in teaching and learning, studies may have to be even larger if estimated relationships between support and outcome variables are small. For this reason, this type of implementation research may be most useful for curricular interventions that have proven efficacious in earlier-stage research, where the investment in a large experiment can be justified by the promise of the intervention.

Examples of this kind of implementation research can readily be found in the fields of medicine and public health, fields with strong commitments to supporting evidence-based practice that are similar to those in current educational policy. In medicine, for example, researchers have used small-scale experiments to test the efficacy of workshops aimed at improving medical practitioners' use of evidence-based approaches in their practice (Cochrane Collaboration, 2005) and large-scale cluster randomized controlled trials comparing the efficacy of different dissemination strategies (Watson et al., 2002). Public health researchers have also conducted experiments comparing the efficacy of different approaches to implementation support for public health providers. For example, Kelly and colleagues (Kelly et al., 2000) studied three different approaches to professional development to support the implementation of AIDS prevention programs.

In their study, they randomly assigned participants to one of the programs and then compared the impacts of the programs on rates of program adoption and implementation.

A study completed recently by Penuel and colleagues (Penuel et al., in press; Penuel & Gallagher, in press) illustrates one application of this approach to experimentation that compares different models of implementation support. Their study focused on conditions under which teachers' adaptations of curriculum might support, rather than hinder, making improvements to both teaching and learning. Instead of viewing adaptation as a problem to be solved, the study put at the heart of its inquiry a central question in policy debates today: Should we prepare teachers to adopt, adapt, or create curriculum materials for students? The study did not set out to resolve this debate, but rather to inform it by a study of what happens when we randomly assign teachers to different support conditions that correspond to these alternatives in one subject area in a single school district.

This efficacy trial compared the impacts of three different programs for preparing teachers to teach for deep understanding of Earth science concepts, following the Understanding by Design (Wiggins & McTighe, 1998) model for unit creation. All three programs tested in the study reflected research-based principles for professional development (e.g., were of a significant duration, involved teachers in active learning strategies), but they differed with respect to the role they gave to teachers in curriculum. In the Adopt program, teachers learned how to *adopt* high-quality curriculum materials developed by experts in Earth science and curriculum design. In the Design program, teachers learned how to *design* curriculum experiences aligned to local standards using available materials and lessons they developed themselves. In the Principled Adaptation program, teachers learned how to *adapt* expert-developed materials in a principled way to

align to local standards. To test the efficacy of the program, teachers who volunteered for the study were randomly assigned to one of the three programs or to a "business-as-usual" control group, and changes to teaching and learning were documented using a combination of surveys, observation, analyses of lesson plans, and standards-aligned tests of student learning. Teachers in all four groups came from a district seeking to promote the Understanding by Design model, but the district had not yet made significant investments in professional development to support implementation.

Results of the analyses of teacher survey and observations at the end of the first year of implementation (Penuel et al., in press) indicated all three programs affected teachers' instructional planning and practice in Earth science, though effects differed by program, and the program was not as effective in some areas as in others. After a year, teachers in the Design and Principled Adaptation programs reported significant changes to their unit planning process, a finding that is also consistent with intent of the professional development designs for those conditions. In particular, teachers reported that the programs had affected both the process by which they planned and the content of their units. Consistent with the idea that all adoption involves some adaptation, teachers assigned to the Adopt program also reported making some changes to how they planned units of instruction. Qualitative data from the implementation survey revealed the nature of their changes was different from that of the teachers assigned to the Design and Principled Adaptation programs, in that their changes were largely limited to creating pacing guides to go with the curriculum materials they were expected to adopt. Observational data showed that all three programs produced students who could provide explanations for why their teacher had them engage in particular activities with reference

to a big idea in the unit. At the same time, none of the designs had an impact on the probability that students would be observed engaging in explanation or application. Further, in contrast to the data on instructional planning that would suggest a greater attention to assessments, observers did not find teachers making use of preconceptions in their instruction.[1]

For us, this study of adaptation illustrates the potential of experimental research to advance scientific understanding of how best to support teacher enactment of curriculum. This study explicitly compares different models of teacher support, and despite the study's reliance on a volunteer population, the employment of a random assignment design helps reduce bias associated with teacher selection (Shadish, Cook, & Campbell, 2002). In contrast to experiments that are focused mainly on the effects of particular programs and curricula on student learning, this study devoted considerably more resources to teacher professional development and its effects on teaching and learning. As such, the study provides some preliminary evidence in answer to the question of how to productively support teacher adaptation.

Realizing the Promise: Cultivating Implementation Research That Can Advance the Science of Learning

For implementation research to have an impact in the way it has not yet had on the field of educational practice, some reframing of the goals, funding streams for, and participants in research will be necessary. We recognize we are not the first to suggest that implementation research ought to be more forward-looking, and our calls come

[1] Results of the study of the programs' impacts on student learning also showed significant impacts of two of the programs. Since those results have not appeared yet in peer-reviewed publications, however, we do not report them here.

amidst others' calls for a more "design-focused" research on supports for implementation. Thus, it is not likely within the current system that implementation research will be able to have an impact on the science of learning without some important changes to how we do research.

One change we see as necessary to impact the science of learning is to change the goal of intervention research from a search for "programs" and "curricula" to what have been called "curricular activity systems" (Roschelle, Knudsen, & Hegedus, in press). Curricular activity systems refer to linked, coherent sets of print and other media resources for student and teacher learning that facilitate tasks that develop student knowledge in a skill in a domain. We prefer this term to the terms "program," "curriculum" or "intervention," because such terms can imply a narrow focus on materials for students and deflect attention both from the ways that activities mediate students' and teachers' interact with materials and from what materials are needed to support teacher learning with respect to guiding these activities. Studies of patterns of the implementation of curricular activity systems thus require descriptions and analyses of classroom processes, as well as analysis of professional development materials and organizational routines and processes that are designed to support individual teacher learning.

It is this simultaneous focus on curricular activity systems and organizational and institutional contexts that distinguishes implementation research as we are defining it from traditional evaluation research and from most learning sciences research and sociological studies in education. Evaluation studies often focus on program implementation, yielding useful insights as to the breadth, depth, and quality of

implementation. But in such studies, the programs always occupy the "figure," while the context is the serves as the "ground" of research. The same focus on programs is evident for most learning sciences research, where researchers' attentions have often been on creating so-called "hot houses" of innovation that enable them to foster and study new learning environments that diverge significantly from common practice (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004). By contrast, sociological studies in education bring organizational and institutional contexts into focus, but such studies rarely attend to the ways in which what matters about the context may differ, depending upon the particular curricular activity systems being enacted.

We also see a need for new funding streams to support implementation research that expand what researchers may propose in field-initiated grant competitions. Specifically, we call for inter-agency support for a new program of research focused specifically on studying implementation processes in different organizational and institutional contexts. This new research program needs to be collaborative between the National Science Foundation (NSF) and U.S. Department of Education (ED), since these two agencies historically have had complementary goals with respect to education. The NSF has pursued innovations and research in learning and pedagogy in its Education and Human Resources Directorate and within its Directorate for Social, Behavioral, and Economic Sciences has promoted the study of organizational learning. ED, through grants from both the Institute of Education Sciences and Office of Innovation and Improvement, has developed educational interventions and supported schools and districts in doing what works. A new program could be similar to the earlier Interagency Education Research Initiative (IERI) in its focus on scaling, but it should be distinct in its focus on

implementation research not just for well-developed programs but also on programs under development or that address pressing policy concerns, such as closing equity gaps in student opportunities to learn.

A distinct program of the magnitude of IERI is needed because resources are needed specifically to advance new interdisciplinary and integrated methods of educational research that answer specific questions about what is needed to scale up programs and policies and to sustain improvements over time. The typical 3 to 5 year research project is not sufficient to achieve, understand and document the changes desired, nor the study of their replication at other sites. This is in part because the process of change takes time, but also because collaboration among researchers and between researchers and districts to define, study, and develop usable findings from implementation research takes longer than the timeframe of a typical research project. Further, what is learned from individual projects may not help us understand "what works" across a variety of contexts, in terms of scaling and sustaining change. In addition, curriculum developers need to develop with scale in mind, but current funding streams provide almost no incentive for planning for scale in the context of a development-focused grant. Similarly, there is not a separate funding stream for researchers to develop, test, and compare the efficacy of different approaches to scaling.

Finally, we call for the inclusion of more diverse voices in early stages of development and implementation research efforts. Many research teams led by curriculum developers will benefit from the inclusion of scholars from other disciplines, policymakers, practitioners, students, parents and community members. It will not be sufficient to advance the science of learning if implementation researchers develop new

methods in isolation from one another and from the concerns of policy makers, educational leaders, teachers, and parents and community members. New forums for building community not only across disciplinary boundaries but also across institutional ones are necessary to achieve a broader impact from implementation research. We see the need for more enduring researcher-community partnerships that develop in tandem with interdisciplinary research partnerships and that focus on problems of practice as experienced today by districts and schools and as anticipated for tomorrow. We also see the need to develop models of co-design and co-development of programs and policies and for research on those models. We have cited some examples above that represent early attempts to formulate such models, but the research on those models is still underdeveloped. We need to know what models are effective and under what conditions, since goals and available resources are likely to vary from project to project.

Conclusion

We have described in this articles three functions of implementation research that go far beyond what is envisioned by traditional "scaling up studies." In the examples of research that illustrate the functions—identifying problems of practice that can become the focus of design, bridging gaps between current capacity and visions for change, and testing models of implementation support—we have shown the potential of such research to inform design in ways that result in more usable and useful innovations for teachers and, in some cases, have resulted in significant changes to teaching and learning. At present, these examples of implementation research are far too few to have a broad impact in the field, however. Researchers need to take an expanded view of the object of research and development and to adopt new ways of doing research that are supported by

federal funding streams for field-initiated research. It is our hope and our belief that by integrating implementation research more fully into the cycle of program and policy development in the ways we have exemplified, the pathways to improving teaching and learning that have eluded us for decades can be identified and pursued.

References

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., Novak, T., Murphy, R. F., et al. (2009). *Achievement effects of four early elementary school math curricula: Findings from first graders in 39 schools*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444-455.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is--or might be--the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 26*, 6-8, 14.

Barab, S., & Luehmann, A. L. (2003). Building sustainable science curriculum: Acknowledging and accommodating local adaptation. *Science Education, 87*(4), 454-467.

Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change, Volume 4: The findings in review*. Santa Monica, CA: RAND.

Besterfield-Sacre, M., Resnick, L. B., Mehalik, M., Sherer, J. Z., & Halverson, E. (2007, February). *A framework for effective management of school system performance.* Paper presented at the Conference on Human and Social Capital in Learning Systems, Pittsburgh, PA.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: A customer-centered approach to systems designs*. San Francisco, CA: Morgan Kaufmann.

Bidwell, C. E., & Yasumoto, J. Y. (1997). The collegial focus: Teaching fields, colleague relationships, and instructional practice in American high schools. *Sociology of Education, 72*(4), 234-256.

Bissell, J. S., Cross, C. T., Mapp, K., Reisner, E. R., Vandell, D. L., Warren, C., et al. (2003). Statement released by members of the scientific advisory board for the 21st Century Community Learning Center evaluation [Electronic Version]. Retrieved February 14, 2007 from www.gse.harvard.edu/hfrp/content/projects/afterschool/resources/21stcclc_statement.doc

Blumenfeld, P., Fishman, B. J., Krajcik, J., Marx, R. W., & Soloway, E. (2000). Creating usable innovations in systemic reform: Scaling up technology-embedded project-based science in urban schools. *Educational Psychologist, 35*(3), 149-164.

Brown, A. L., & Campione, J. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (pp. 289-325). Hillsdale, NJ: Erlbaum.

Brown, M. W., & Edelson, D. C. (2001, April). *Teaching by design: Curriculum design as a lens on instructional practice.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Bruner, J. S. (1960). *The process of education*. Cambridge, MA: Harvard University Press.

Burkhardt, H., & Schoenfeld, A. H. (2003). Improving educational research: Toward a more useful, more influential, and better-funded enterprise. *Educational Researcher, 32*(9), 3-14.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively Guided Instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3-20.

Cobb, P. A., & Smith, T. M. (2008, March). *Designing learning organizations for instructional improvement in mathematics.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Cobb, P. A., & Smith, T. M. (2009, April). *Instructional improvement at scale.* Paper presented at the National Council of Teachers of Mathematics Research Presession, Washington, DC.

Coburn, C. E., & Russell, J. L. (2008). District policy and teachers' social networks. *Educational Evaluation and Policy Analysis, 30*(3), 203-235.

Cochrane Collaboration. (2005). Continuing education meetings and workshops: Effects on professional practice and health care outcomes (Cochrane review). *Journal of Continuing Education in the Health Professions, 21*(3), 187-188.

Cohen, D. K., Fuhrman, S. H., & Mosher, F. (2007). Conclusion: A review of policy and research in education. In S. H. Furhrman, D. K. Cohen & F. Mosher (Eds.), *The state of education policy research* (pp. 349-382). New York: Routledge.

Cordray, D., & Pion, G. M. (2006). Treatment strength and integrity: Models and methods. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research*

methodology: Psychological measurement and evaluation (pp. 103-124). Washington, DC: American Psychological Association.

Correnti, R. (2007). An empirical investigation of professional development effects on literacy instruction using daily logs. *Educational Evaluation and Policy Analysis, 29*(4), 262-295.

Dansinger, M. L., Gleason, J. A., Griffith, J. L., Selker, H. P., & Schaefer, E. J. (2005). Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction. *The Journal of the American Medical Association, 293*, 43-53.

Davis, E. A., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher, 34*(3), 3-14.

Dynarski, M., Agodini, R., Heaviside, S., Carey, N., Campuzano, L., Means, B., et al. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Ehn, P. (1992). Scandinavian design: On participation and skill. In P. S. Adler & T. A. Winograd (Eds.), *Usability: Turning technologies into tools* (pp. 96-132). New York: Oxford University Press.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review, 14*, 57-74.

Elmore, R. F. (1980). Backward mapping: Implementation research and policy decisions. *Political Science Quarterly, 94*(4), 601-616.

Emirbayer, M., & Mische, A. (1998). What is agency? *The American Journal of Sociology, 103*(4), 962-1023.

Fishman, B. J., & Krajcik, J. (2003). What does it mean to create sustainable science curriculum innovations? A commentary. *Science Education, 87*(4), 564-573.

Fishman, B. J., Marx, R. W., Best, S., & Tal, R. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education, 19*(6), 643-658.

Fishman, B. J., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *The Journal of the Learning Sciences, 13*(1), 43-76.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: Application to the implementation of computer technology in schools. *Sociology of Education, 77*(2), 148-171.

Gamoran, A., Gunter, R., & Williams, T. (2005). Professional community by design: Building social capital through teacher professional development. In L. V. Hedges & B. Schneider (Eds.), *The social organization of schooling* (pp. 111-126). New York: Russell Sage.

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., Unlu, F., Bozzi, L., et al. (2008). *Reading First impact study: Final report*. Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Gorden, V. S., & Bieman, J. M. (1995). Rapid prototyping: Lessons learned. *Software, IEEE, 12*(1), 85-95.

Grossman, P., Wineburg, S. S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record, 103*(6), 942-1012.

Halverson, R. R. (2003). Systems of practice: How leaders use artifacts to create professional community in schools. *Education Policy Analysis Archives, 11*(37).

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly, 42*(4), 716-749.

Hupert, N., & Heinze, J. (2006). Results in the palms of their hands: Using handheld computers for data-driven decision making in the classroom. In M. van't Hooft & K. Swan (Eds.), *Ubiquitous computing in education: Invisible technology, visible impact* (pp. 211-229). Mahwah, NJ: Erlbaum.

Hupert, N., Heinze, J., Gunn, G., & Stewart, J. (2007). Using technology-assisted progress monitoring to drive improved student outcomes. In E. B. Mandinach & M. Honey (Eds.), *Data-driven school improvement: Linking data and learning* (pp. 130-150). New York: Teachers College Press.

Institute of Education Sciences. (2009). *CFDA 84.305A Education research grants*. Washington, DC: U.S. Department of Education.

James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the National Evaluation of the 21st Century Community Learning Centers program. *Educational Evaluation and Policy Analysis, 29*(4), 296-318.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5*, 602-619.

Kame'enui, E. (2002). *Final report: An analysis of reading assessment instruments for K-3.* Eugene, OR: Institute for the Development of Educational Achievement.

Kelly, J. A., Somlai, A. M., DiFranceisco, W. J., Otto-Salaj, L. L., McAuliffe, T. L., Hackl, K. L., et al. (2000). Bridging the gap between the science and service of HIV prevention: Transferring effective research-based HIV prevention interventions to community AIDS service providers. . *American Journal of Public Health, 90*(7), 1082-1099.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23*(4), 418-444.

Kruse, S. (2001). Creating communities of reform: Continuous improvement planning teams. *Journal of Educational Administration, 39*(4), 359-383.

Lieberman, A., & Miller, L. C. (2001). *Teachers caught in the action: professional development that matters*. New York: Teachers College Press.

Little, J. W. (2001). Professional development in pursuit of school reform. In A. Lieberman & L. Miller (Eds.), *Teachers caught in the action: Professional development that matters* (pp. 23-43). New York: Teachers College Press.

Lynn, E. (2005). *mCLASS(tm):DIBELS® Customer Feedback Survey*. New York: Wireless Generation, Inc.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*, 144-158.

Mahoney, J. L., & Zigler, E. F. (2006). Translating science to policy under the No Child Left Behind Act of 2001: Lessons from the national evaluation of the 21st

Century Community Learning Centers. *Journal of Applied Developmental Psychology, 27*, 282-294.

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

McDonald, S.-K. (2009). Scale-up as a framework for intervention, program, and policy evaluation research. In G. Sykes, B. Schneider & D. N. Plank (Eds.), *Handbook of education policy research* (pp. 191-208). Washington, DC: American Educational Research Association.

McLaughlin, M. W. (1976). Implementation as mutual adaptation: Change in classroom organization. *Teachers College Record, 77*(3), 339-351.

Millen, D. R. (2000). Rapid ethnography: Time deepening strategies for HCI field research. In *Proceedings of the conference on designing interactive systems: Processes, practices, methods, and techniques* (pp. 280-286). New York, NY: ACM Press.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33-84.

Penuel, W. R., Benbow, A., Mably, C., McWilliams, H., McAuliffe, C., & Hayden, M. M. (in press). Teaching for understanding in Earth science: Comparing impacts on

planning and instruction in three professional development designs for middle school science teachers. *Journal of Science Teacher Education*.

Penuel, W. R., Fishman, B. J., Gallagher, L. P., Korbak, C., & Lopez-Prado, B. (2009). Is alignment enough? Investigating the effects of state policies and professional development on science curriculum implementation. *Science Education, 93*(4), 656-677.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921-958.

Penuel, W. R., & Gallagher, L. P. (in press). Comparing three approaches to preparing teachers to teach for deep understanding in Earth science: Short-term impacts on teachers and teaching practice. *The Journal of the Learning Sciences*.

Penuel, W. R., Lynn, E., & Berger, L. (2006). Classroom assessment with handheld computers. In M. van 't Hooft & K. Swan (Eds.), *Ubiquitous computing in education: Invisible technology, visible impact* (pp. 103-125). Mahwah, NJ: Erlbaum.

Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). The WHIRL co-design process: Participant experiences. *Research and Practice in Technology Enhanced Learning, 2*(1), 51-74.

President's Committee of Advisors on Science and Technology Panel on Educational Technology. (1997). *Report to the President on the use of technology to strengthen K-12 education in the U.S.* Washington, DC: Author.

Remillard, J. T. (2000). Can curriculum materials support teachers' learning? Two fourth-grade teachers' use of a new mathematics text. *Elementary School Journal, 100*(4), 331-350.

Resnick, L. B., Besterfield-Sacre, M., Mehalik, M. M., Sherer, J. Z., & Halverson, E. R. (2007). A framework for effective management of school system performance. In P. A. Moss (Ed.), *National Society for the Study of Education (NSSE) Yearbook on evidence and decision making* (Vol. 106, pp. 155-185). New York, NY: Teachers College Press.

Roschelle, J., Knudsen, J., & Hegedus, S. J. (in press). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In M. J. Jacobson & P. Reimann (Eds.), *Designs for learning environments of the future: International perspectives from the learning sciences*. New York: Springer.

Roschelle, J., Penuel, W. R., Yarnall, L., & Shechtman, N. (2005). Handheld tools that "informate" assessment of student learning in science: A requirements analysis. *Journal of Computer Assisted Learning, 21*(3), 190-203.

Rowan, B., Camburn, E., & Correnti, R. (2004). Using teacher logs to measure the enacted curriculum: A study of literacy teaching in third-grade classrooms. *Elementary School Journal, 105*(1), 75-102.

Rowan, B., Harrison, D. M., & Hayes, A. (2004). Using instructional logs to study mathematics curriculum and teaching in the early grades. *Elementary School Journal, 105*(1), 103-127.

Rowan, B., & Miller, R. J. (2007). Organizational strategies for promoting instructional change: Implementation dynamics in schools working with comprehensive school reform providers. *American Educational Research Journal, 44*(2), 252-297.

Schneider, R. M., & Krajcik, J. (2002). Supporting science teacher learning: The role of educative curriculum materials. *Journal of Science Teacher Education, 13*(3), 221-245.

Shrader, G., Williams, K., Lachance-Whitcomb, J., Finn, L.-E., & Gomez, L. (2001, April). *Participatory design of science curricula: The case for research for practice.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Singer, J. E., Krajcik, J., Marx, R. W., & Clay-Chambers, J. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist, 35*(3), 165-179.

Sloane, F. C. (2008). Through the looking glass: Experiments, quasi-experiments, and the medical model. *Educational Researcher, 37*(1), 41-46.

Slotta, J., & Peters, V. (2008, June). *A blended model for knowledge communities: Embedded scaffolded inquiry.* Paper presented at the 8th International Conference of the Learning Sciences, Utrecht, the Netherlands.

Spillane, J., & Jennings, N. (1997). Aligned instructional policy and ambitious pedagogy: Exploring instructional reform from the classroom perspective. *The Teachers College Record, 98*(3), 449-481.

Taylor, C. (2005, March 6). School of bright ideas. *TIME, 165*.

Texas Education Agency. (2003). *A report to the 78th Texas Legislature from the Texas Education Agency: Interim Report on Ed Tech Pilots.* Austin, TX: Texas Education Agency.

Tripp, S. D., & Bichelmeyer, B. (1990). Rapid prototyping: An alternative instructional design strategy. *Educational Technology, Research, and Development, 38*(1), 31-44.

Watson, M. C., Bond, C. M., Grimshaw, J. M., Mollison, J., Ludbrook, A., & Walker, A. E. (2002). Educational strategies to promote evidence-based community pharmacy practice: A cluster randomized controlled trial (RCT). *Family Practice, 19*(5), 529-536.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher, 37*(8), 479-479.

Werner, A. (2004). *A guide to implementation research*. Washington, DC: The Urban Institute Press.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.