Data Use by Teachers:
Productive Improvement or Panacea?

Mary M. Kennedy, Ph.D.
College of Education
Michigan State University
East Lansing, Michigan 48824-1034

The Education Policy Center
at Michigan State University

MICHIGAN STATE
UNIVERSITY

www.epc.msu.edu • epc@msu.edu

Data Use by Teachers:
Productive Improvement or Panacea?

Mary M. Kennedy
Michigan State University
May 20, 2011

Abstract

The idea that teaching practices could be improved through better data systems has captured the imagination of policy makers, researchers and foundations. However, the concept of data use itself is so ill-defined as to suggest that it is more of a panacea than a coherent reform proposal. This paper examines the concept with an eye toward what we would need to know about teacher data use to ensure that it would actually yield productive improvements in teaching and would not turn out to be an ill-thought-through panacea. The paper offers a research agenda organized around two central issues: The difference between psychometric feedback and clinical feedback, and the difference between productive teaching and defensive teaching.

Data Use by Teachers:
Productive Improvement or Panacea?

Mary M. Kennedy

Michigan State University

The idea that teaching practices could be improved through better data systems has captured the imagination of policy makers, researchers and foundations. It started with the idea that students could learn more if they received better information about what they should be learning, what they had done so far, and what was still needed for them to master the material. It spread to the idea that teachers, too, could benefit from better knowledge of how their students were doing, and that schools and districts would benefit from more formative feedback as well.

My task in this paper is to examine research on how teachers use data to understand and respond to their own students, and to outline a research agenda regarding teachers' data use. For purposes of this paper, I assume the data most likely to be of interest would be standardized, system-wide assessments of student progress that may or may not be embedded into the curriculum and that provide *either* teachers *or* students, or both, with information about student progress, what students have learned or what they have failed to learn. To avoid questions about the technical quality of the data themselves, the paper assumes that any such data system provides timely information of reasonable technical quality.

The rhetoric about data use persistently argues that it will foster improvements at all level of education: within individual classrooms, within schools, within districts. But the term also seems to refer to a wide variety of data and a wide variety of uses, thus raising the possibility that the concept itself is more of a panacea, a mythical cure-all, than a specific reform proposal designed for a specific purpose. I therefore begin this paper by examining the character of the proposed innovation itself. The central concern I have is whether the concept of formal data systems might turn out to be the latest in a long line of panaceas that have ultimately been found wanting and have been abandoned.

Data Use as a Panacea

By definition, a panacea is a solution to a wide variety of problems rather than to a single, clearly specified problem. Education policy in the United States has a long history of embracing panaceas, investing in them, and then ultimately abandoning them. Panaceas are difficult to study in part because they tend to be ill-defined. The term data use, for instance, has been used to refer to curriculum-embedded student assessments, management information systems, models of learning progressions, findings from the *What Works* Clearinghouse and other forms of data as well. Some of these systems, such as the *What Works* Clearinghouse, seek to foster a kind of data use that consists of selecting programs that have been demonstrated to be most effective. Other systems, such as district-level test score data bases, seek to foster a

kind of data use that consists of beefing up services or curricula or instruction wherever scores are low. Some seek to foster a kind of data use that consists of making high-stakes accountability decisions about teachers, while others seek to foster a kind of formative improvement cycle for teachers. The fact that the phrase "data use" is so popular, even when it refers to such disparate activities and purposes, adds to the suspicion that the idea might be wishful thinking rather than shrewd strategy.

Panaceas are also difficult to study because researchers themselves can become smitten by them and become advocates more than examiners of these new ideas (e.g., Popham 2006; Rothkopf 2009). Even when researchers are skeptical, panaceas are difficult to study because they are put forward as solutions to numerous problems, rather than to any one specific problem. Consequently there is no obvious alternative with which to compare a panacea, nor any obvious outcome to assess. So research on panaceas tends to focus on how well they are being *implemented* rather than on whether they actually succeed at any particular goal relative to some alternative. In the case of data use, for instance, we see studies describing the things districts and schools have done to install a data system, but that say little about whether the data system, once installed, actually improved any decision making or instructional activities (e.g., Datnow et al. 2007; Kerr et al. 2006).

The first task in establishing a research agenda on data use, then, is to remove the concept from its panacea pedestal and define a specific, researchable reason to care about it. The research question of interest cannot be simple whether data are used by teachers, but rather whether data are used *productively,* to improve instruction and to improve student learning. Below I outline the domain of interest regarding teachers' use of data. In this case, the data of interest are largely test data.

Research on Teachers' Use of Data

Every educational intervention is susceptible to unanticipated consequences and these consequences need to be identified and understood along with the intended consequences. Teachers may respond to new rules and procedures not by becoming more productive, but instead by becoming more rigid and defensive in their practice. They may stick to the letter of a new law, ignoring its spirit. They may learn to alter superficial aspects of their work, or to give an appearance of compliance without really engaging with deeper and more complicated instructional issues. Such defensive responses were found by Heilig and Darling-Hammond (2008) when they studied an accountability system in Texas. Teachers there learned how to create the appearance of rising test scores without really altering their instructional practices. These authors found that, for example, low-achieving students were disproportionately excluded from taking the state's high-stakes test, and that gains on the high stakes test were not reflected on other, lower-stakes tests. This is not the kind of response advocates seek. Rather, they would like to see teachers use test data to re-think their instructional strategies, re-arrange their resources or instructional systems, or make other efforts to improve student learning.

So when we seek greater use of data, we need to ask what kind of use is occurring, and what kind of instructional changes are occurring in response to test data. For purposes of this paper, I refer to the desired uses of test data as *productive* uses and the undesirable uses as

defensive uses. Defensive responses to test data include efforts to "game the system" by removing low-performing students when tests are administered, but they also include responses such as narrowing instructional content to the literal items on the test, with no attention to underlying concepts that make those items relevant, or teaching students to memorize specific facts with no attention to their meaning. Test items are designed to represent broad domains of knowledge and when teachers focus too tightly on the discrete items, the items lose their representativeness and are no longer reliable estimates of student knowledge of that domain. Therefore, examining the kind of instructional changes that follow from test use is an important part of a research agenda on data use.

I would like to also extend these concepts of "productive" and "defensive" to refer to instruction more broadly. Our country has spent the past two decades in the throes of a massive reform designed to improve the overall quality of education. A variety of terms have been used to describe the nature of this reform. Some say it seeks more rigorous or demanding content, some that seeks more attention to big ideas or complex ideas, some that it seeks more intellectual engagement, some that it seeks more attention to reasoning and problem solving, and some that it seeks more authentic learning activities. All of these ideas are discussed as alternatives to the kind of instruction that has traditionally permeated schools in the United States, which is sometimes referred to as lower-level, focused more on recall than on reasoning, more on facts than on ideas, more on discrete bits of knowledge than on integrated knowledge.

There is no agreement on the details of either of these approaches to teaching, nor on what either should be called. I tend to think of the reform approach as "teaching for learning" since it is driven by a sense of what should be learned, and to think of traditional instruction as "teaching for coverage" since it seems to be driven by how much material has been covered rather than by what students are actually learning. For purposes of this paper, though, I prefer to call them productive and defensive approaches to instruction, respectively. Productive instruction is analogous to productive uses of test data in that both have a strong conception of substantive purpose. Defensive instruction is analogous to defensive use of tests in that both are more concerned with coverage, either covering all the pages in the book or ensuring that students can correctly respond to all the test items, regardless of whether students actually understand what they have read or recited.

Since the data of interest to teachers is primarily test data, this matter of whether instructional approaches are productive or defensive is important. There is a good chance that test data will motivate more defensive instructional practices, encouraging teachers to give students worksheets that reinforce a short list of facts needed to do well on the test. If this is the outcome, then the use of test data would work against other educational reforms by discouraging teachers from pursuing deeper understanding and more complex ideas. Moreover, there is a great deal of evidence that teachers prefer more defensive practices because of other pressures in the environment. So even if test data encourage more productive instruction, that influence may be masked by other factors that foster other, more defensive instructional practices. Thus a full understanding of the role of tests in instruction needs to consider not just the influence of the tests themselves, but also the complementary or mitigating influences of other pressures in the environment.

So the central issue is broad. It encompasses how tests influence teachers' instructional decisions, whether they foster or discourage productive instruction, and how their influence complements or competes with other influences in the instructional system.

One remaining complication is this: Teachers already know more than anyone else in the school about their own students. They spend every day directly interacting with students, responding to their questions, listening to their speculations and reading their daily assignments. Teachers are well aware of the differences among their students, and of the nuances of how their students think about subject matter. This knowledge is of a different character than that provided by formal assessment systems, of course, but it is also free and readily available. Thus an important question about the use of test data is whether the data provide enough value to justify its cost *relative to the knowledge teachers already have at their disposal.* Following Meehl (1954), I refer to these two kinds of knowledge as *clinical feedback* and *psychometric feedback* and argue that a research agenda on teachers' use of test data needs to consider what kind of knowledge each form of feedback provides and what benefits or drawbacks each may have.

Table 1 displays the intersection between these two pairs of terms—clinical vs psychometric feedback and productive vs defensive instruction—to depict a domain for research on teachers' use of data. The two rows represent the two main kinds of feedback that teachers have available to them, and the two columns represent the two possible ways that teachers can respond instructionally to that feedback.

Table 1

	Productive Response	Defensive Response
Psychometric Feedback	Cell 1	Cell 2
Clinical Feedback	Cell 3	Cell 4

Viewed in this way, the agenda for research on teachers' use of data would focus on two broad questions:

- Does psychometric feedback provide more or different knowledge than clinical feedback? In other words, is there more or different action in Cell 1 than in Cell 3? We should also ask a cost-benefit question: Does the difference between Cell 1 and Cell 3 justify the cost of the data system?

- Does the introduction of psychometric feedback encourage relatively more productive instruction or relatively more defensive instruction? That is, do we see more action in Cell 1 or in Cell 2 when we introduce a new data system? And since the impact of psychometric feedback can be enhanced or suppressed by other factors in the system, this question must always be addressed in the context of other influences on practice.

Both of these questions highlight the importance of examining teachers' use of test data in the context of their' larger environment: in the context of the clinical feedback they already have and in the context of other institutional forces. Moreover, these contextual factors must be examined from the teachers' point of view, for it is teachers' perceptions of these factors that influence their decisions, not the interpretations of outside observers or the intentions of well-meaning institutions. Years ago, when psychologists were first trying to simulate human problem solving on computers, they realized that they had to distinguish the objective parameters of a problem from the problem solver's own perception of it. Simon and Newell (1971) used the term *problem space* to refer to what the problem solver understood about the situation at any given moment. The notion of a problem space is useful here because it reminds us that what teachers might do with psychometric feedback, or any other resource, depends on what they perceive their needs, options, or constraints to be, not what an external observer might think would be a logical use.

Teachers' problem spaces have three main components: students, goals and constraints. The following three sections of this paper examine each of these. In each case, I first ask how teachers normally perceive and understand this component of their problem space and then ask the questions suggested by Table 1: whether psychometric feedback adds to clinical feedback in helping teachers understand this component of their problem space, and whether psychometric feedback is likely to motivate more productive or more defensive responses to this component.

Students

As Cohen (1988) has argued, teaching is ultimately an attempt to change other human beings. Such enterprises are extremely difficult and also heavily dependent on the willing cooperation of other people. For teachers to be successful, then, students must be participants in the instructional process. Yet school attendance is compulsory in the United States, so teachers' classrooms will always contain some students who have no interest in learning. This means that teachers cannot merely present content; they must either persuade students that this content is worth thinking about or persuade them to at least not interfere with other students who may want to learn.

The second problem that students present is that, even when they are willing to participate, they are still young, restless, and inexperienced at the kind of sustained, concentrated, purposeful work that is required for learning. This is true at every grade level, for at every grade level, teachers demand more sustained attention from students than they did the year before, and students are never prepared for that level of commitment. These two facts, that teachers work with children and adolescents rather than adults, and that these youngsters are not volunteers, has profound implications for teachers: Their task is not merely to teach about school subjects, but also to teach students how to learn and how to cooperate, neither of which comes naturally, and they must persuade students that all of this is worthwhile.

The third problem students present is that they vary, so that instruction on any topic will necessarily be understood differently by different students. Moreover, students' responses to instruction can motivate them to participate relatively more or less. Students who grasp the

material quickly might become bored and begin fiddling, thus distracting other learners, or they might show off their mastery by calling out answers to every question, thus preventing other students from figuring it out for themselves. On the other side are students who have difficulty. If teachers take time to answer their questions and help them think things through, their classmates might become bored and begin entertaining themselves with off-task and disruptive behaviors.

Productive and Defensive Responses to Students

Student resistance to learning can make classroom life look something like an intellectual tug of war, for even as teachers struggle to persuade students to learn, students try to persuade teachers to reduce the intellectual demands placed on them. They try to redefine assignments into less ambiguous, less ambitious projects that make fewer intellectual demands on them. And because teachers need compliant students in order to do any teaching at all, they often concede to student's pleas for relief (Cusick 1983; Doyle 1986; Doyle and Carter 1984; Metz 1993). Sedlak et al. (1986) called this bargaining process a "tacit conspiracy to avoid sustained, rigorous, demanding academic inquiry (pg 5)."

Students' energy level and easy distractibility also motivate teachers to concentrate much of their instructional energy on containing students. They engage in continuous monitoring, much of which is aimed at preventing a lapse in attention, preventing a side-bar conversation, preventing disruptive behavior, or preventing distractions (Kounin 1970). Not only do teachers continuously monitor students, they also convey their awareness to students, publically commenting on which students are behaving appropriately and which are not (Jackson 1968/1990; Kounin 1970). Transcripts of classroom lessons commonly contain teacher comments such as "Gina's point about the slope of this line is a good one. Get back in your seat, Nathan. What else can we say about the slope of this line?"

One important aspect of containment that has received relatively less notice is that teachers also try to contain student *enthusiasm*. Teachers frequently design lessons to foster a minimum level of student interest, but no more than that (Kennedy 2005). They need students to be interested enough to cooperate with the lesson but not so interested that they actually become excited. Over-enthusiasm can lead to too many off-task questions and comments and can slow the group down. One teacher in Kennedy's study said that the first year she worked with a new reform-oriented, problem-solving curriculum, she covered only a third of the text by the end of the year because she had not learned how to contain student discussions and close them out so that the students could move to the next topic.

Because student attention is quickly diverted, teachers also place a high value on maintaining lesson momentum, and their central strategy for attaining momentum is to *keep the entire group moving in tandem* through the curriculum. They have learned that whenever they take time to attend to a single student, they risk losing the attention of the rest of the class. If they divide students into groups and try to work with one group, they must constantly scan the rest of the room to make sure the other students do not lose their focus. A central theme in teachers' responses to their students, then, is to consistently, persistently, and explicitly pull the entire class forward through the lesson.

These aspects of teachers' relationships with students may not seem relevant to questions about data use but they are relevant in two ways. First, they help us understand teachers' widespread reliance on defensive instructional strategies. Moreover, they suggest that one strong hypothesis for how teachers might respond to psychometric feedback is that they will continue using the same defensive strategies they have always used. If this is the case, having more and more detailed information about student progress may not really improve instruction or learning. On the other hand, this portrait of teachers' relations with students also raises an interesting alternative: if psychometric feedback actually helped teachers understand their students' strengths and weaknesses in a way that was instructionally useful, it might also help them develop more productive instructional responses.

Psychometric vs. Clinical Feedback regarding Students

Table 1 suggests that one important question about psychometric feedback is whether or how it might complement, compete with, or supplant the knowledge teachers normally have access to via clinical feedback. Teachers' clinical knowledge of their students is voluminous. Teachers read and evaluate hundreds of daily assignments, field hundreds of questions and listen to hundreds of students' responses to their own questions. They observe conversations among students and see the solutions students write on the board.

Given the pervasiveness of clinical feedback, we should know more about it than we do. But clinical feedback is difficult to study because it is generated in the flurry of classroom life, while teachers are also monitoring students for their interest, attention and conformance with rules. Reynolds et al. (1995) made an effort to help teachers articulate the clues they picked up through classroom interactions. They asked teachers to replay videos repeatedly in order to identify the evidence that they had used to form their assessments. Through this recursive process, teachers eventually identified a wide range of indicators, including the following:

> changes in children's demeanor, extending a concept into new discoveries,
> making spontaneous modifications rather than simple imitations, using a process
> in new contexts, feeling confident in using shortcuts, being able to explain ideas
> to others, and showing high engagement and focus during activities (Reynolds et
> al, 1995, pg 367).

This is not to say that clinical feedback is always beneficial to teachers, nor that it necessarily fosters more productive teaching. In fact, teachers often respond to their clinical feedback, such as bargaining, defensively, becoming more rigid and placing more emphasis on conformance with rules. Moreover, even though clinical feedback is voluminous, it is also both fleeting and ambiguous. Students may seem to grasp a concept on Monday, but to have lost it on Tuesday. Several authors have commented on this problem of uncertainty (Cohen 1988; Floden and Clark 1988; Labaree 2000) and have suggested that it can motivate teachers to narrow their goals to more conservative, more easily achievable ones.

Two studies are especially pertinent to the question of how teachers forge instructional meaning from psychometric data. One study (Goertz et al. 2009) examined teacher's use of an interim assessment system that was designed to foster a cyclical approach to instruction in which teachers teach, then examine where students are, then adjust their teaching accordingly.

These researchers found that the assessments provided information only about whether students could respond correctly to particular test items and did not reveal the reasons for their errors. Consequently, teachers generated their own diagnostic theories to account for student errors and then designed instructional practices based on their self-generated theories. The authors suggested that the data themselves were less important to their instructional responses than the teachers' diagnoses were. Ironically, the diagnoses were probably based on their clinical knowledge of their students, the same knowledge that teachers have always used to decide how to adjust their instructional practices.

The second study was done by a group of researchers whose original purpose was to develop a system of curriculum-embedded, psychometrically-sound assessments that could provide diagnostic feedback to secondary science teachers (Ayala et al. 2008; Brandon et al. 2008; Furtak et al. 2008; Shavelson 2008; Yin et al. 2008). Throughout their development and pilot testing process, the researchers worked with a collaborating teacher who gave them feedback on their efforts. In an early lament about the process, Shavelson (2006) listed numerous difficulties they encountered: The system was too unwieldy; the assessments were too long; the feedback was too slow. Sometimes the researchers generated too much information, or information that conflicted with other available information. Perhaps the most telling problem Shavelson identified was that teachers lacked a framework for interpreting and using the assessment information and in particular didn't know what to do with the achievement patterns they saw. At one point, Shavelson (2006) said, "Teachers did not know how to use this information to close the gap between students' current and desired learning states; nor did we!" (pg 69).

Ultimately, what began as curriculum-embedded, psychometrically sound formative tests evolved into something called "reflective lessons" that teachers themselves implemented, and through which they obtained the kind of immediate feedback normally obtained through clinical feedback. This evolution itself is worth attention, and raises the question whether the final version, the reflective lesson, is closer to its psychometric parents or its clinical sibling in its contribution to teachers' instructional decision making.

Psychometric Feedback and the Problem of Defensive Responses to Students

The second question Table 1 presents is whether psychometric feedback motivates productive changes in instruction. To answer this, we need a way to distinguish those responses that lead to genuine improvements in the education students receive. If psychometric feedback presents nothing more than a list of correct and incorrect responses, rather than a diagnostic interpretation of them, it could foster more defensive responses and fewer productive responses. For example, teachers might benefit from feedback that helped them understand *why* specific errors occurred, or what was not understood and what was misunderstood. Pellegrino (2002) argues that every assessment system is necessarily based in a model of the human mind. Test items are designed to tap into the contents of that mind, but it is the underlying model that tells us what their responses actually mean. Most assessments, he argues, do not acknowledge the fact that human minds hold knowledge in organized structures, so that it is not enough to know

that students were correct on one item and wrong on another. What is needed is knowledge of how the student actually thinks about the entire domain, links ideas together, views causal relations and so forth.

If psychometric feedback were to provide this kind of knowledge to teachers, it would convey to teachers not just which students missed which items, but what concepts students had grasped or misunderstood. It is not clear that many, or even any, available data systems actually provide this kind of instructional insight to teachers. Researchers in the sciences who study learning progressions (e.g., Alonzo and Steedle 2009; Songer et al. 2009) aim to provide this kind of knowledge to teachers, but their assessments are tied to discrete domains of knowledge with clearly specified learning goals and are not sufficiently developed to incorporate into extant year-long school curricula.

Goals

The second important part of the problem space of teachers is their goals. Teachers' goals can be rather complicated, in part because Americans as a society often espouse conflicting purposes for schools and teachers: They want youngsters to learn particular content, but also to be nurtured, developed into good citizens, and motivated to participate productively in society. They want teachers to be role models for moral and ethical behavior but also to be efficient, organized and goal-oriented. They believe all students deserve equal treatment and resources, but they also think some students should receive more. They are divided on whether children should be controlled by external rules with consequences or whether, instead, they should be taught to regulate themselves. They want to socialize students to accommodate the prevailing cultural norms, yet also want students to become critical thinkers; and so forth. These different ideas wax and wane in their social popularity and strain the education system. They have been examined by several writers (Berlak and Berlak 1981; Cremin 1990; Egan 1997; Egan 2001; Tyack and Cuban 1995). But it is teachers, as street-level bureaucrats, who ultimately decide which goals to work on during each curriculum unit, each lesson, and each fleeting interaction with each child.

Even putting aside social and cultural goals, goals for content knowledge can be ambiguous as well. Districts typically provide at least three types of guides for teachers: Curriculum standards, textbooks and their associated teachers' manuals, and one or more standardized achievement tests. These three types of guides may or may not be coherent, internally consistent, or aligned with each other, and they have each separately been the object of substantial criticism. Standards have been criticized because they do not provide instructional guidance but instead simply stipulate broad outcomes (Cohen and Moffitt 2009; Shattuck 2005). They also vary greatly across the states and often are not very well aligned with any of the national professional standards (Porter et al. 2009). Textbooks have been famously characterized as a mile wide and an inch deep (Schmidt et al. 1999). They have also been criticized for reciting information without offering any hint as to its value or meaning. In part because states differ in their standards, American textbooks are crowded with pictures and "factlets," offering something for every constituency. These texts could easily discourage the very kind of thoughtful instruction we claim we want.

Tests provide the third leg of the curricular-guidance stool, and their role has changed substantially in the past two decades. There was a time when students might be tested once every three or six years, but now they are tested every year and often every quarter; many such tests are attached to textbooks. If such a test accompanies a textbook that is a mile wide and an inch deep, it could even further encourage teachers to focus on discrete facts at the expense of thoughtful examination of deeper concepts.

Productive and Defensive Responses to a Complicated Goal Structure

Many people assume that teachers' instructional plans are based on rational choices — that is, teachers begin with a curricular goal and then select the most effective techniques for achieving that goal, where techniques may include demonstrations, recitations, learning activities or other events. But teachers think more like designers who face a number of constraints and who invent solutions to fit within them. Architecture might be an appropriate analogy. An architect might be asked to build a house with three bedrooms and a south-facing living room, that will sit on the side of a hill, and whose construction must not disturb a particular tree. The constraints are constant, but there are an infinite number of possible ways to satisfy them all.

One way that teachers differ from architects, though, is that teachers have very little time to plan out their lessons and so they do not aim for perfect or elegant solutions. Instead, they aim for something that *will do*. Simon and Newell (1971) called this kind of thinking "satisficing," a term that blends the concepts of "satisfy" and "suffice" and is intended to remind us that the goal is not perfection, but rather adequacy. These authors argued that, in many situations, rational problem solving is not possible because we have too little information, we have multiple needs, or goals, and we can only plan for short purposive episodes, not for longer ones. They did not mean to imply that satisficing is a bad thing; rather, it is the best we can hope for in ambiguous, multidimensional situations.

Like architectural planning, teaching necessarily involves numerous simultaneous considerations. Lessons are designed so that teachers can simultaneously manage student behavior, motivate students and teach students; ensure civil discourse; ensure that work is completed at precisely 11:35; ensure that materials are available, distributed efficiently, and collected efficiently at the end of the period or day, and so forth (Fenwick 1998; Kennedy 2005). When Kennedy (2005) asked teachers what motivated particular actions during a lesson, teachers frequently offered two or three different considerations. For instance, a teacher might attribute an action to (a) noticing that a particular student was confused, (b) realizing that the lesson needed to be finished in ten minutes; and (c) realizing that one of the table groups had moved off task.

These aspects of teacher planning are also relevant to psychometric feedback. Much of the advocacy for institutionalized testing systems is based on an assumption that this kind of feedback will help teachers engage in some form of rational planning. If that assumption is wrong – that is, if teacher plans necessarily rest on satisficing rather than rational planning – then the way teachers draw on psychometric feedback for instructional planning may never match our expectations for it.

Psychometric vs. Clinical Feedback regarding Goals

Following Table 1, we should ask whether psychometric feedback complements, competes with, or supplants the knowledge teachers normally gain from clinical feedback, specifically with respect to instructional goals. These two forms of feedback inform teachers about very different types of goals: psychometric feedback tends to be very narrowly focused on specific curricular goals, whereas clinical feedback may help teachers see whether students are interested or bored, cooperating with one another or fighting, contributing to the social order of the classroom as a whole, taking personal responsibility for their work, and so forth.

Clinical feedback is likely the *only* source of information about the many non-content-matter goals teachers have for their students. But psychometric feedback can offer something teachers cannot obtain from clinical feedback, for in addition to informing them about what students have learned, it can also inform them about what the substantive goals should be. An early study of the Vermont state summative assessment (Koretz et al. 1994) provides a useful example. For this assessment, the state used student products that were generated in the course of regular instruction, and then taught teachers a scoring rubric for evaluating these products. Teachers found that the scoring rubric itself was helpful because it helped them understand *what good student work should look like,* and in that sense it helped them engage in more focused, goal-oriented instruction.

Psychometric Feedback and the Problem of Defensive Responses to Goals

I have found no studies that engaged in serious content analyses of psychometric feedback systems to see how well they aligned with other types of curricular guidance. This is unfortunate, for any study that aims to understand the way teachers interpret and respond to test data must also ascertain what the messages are within those data and how those messages mesh with messages in textbooks and other guidelines. There are at least two important questions here.

First, do teachers respond to psychometric feedback differently depending on how well the data align with other types of curricular guidelines? For instance, if psychometric feedback is not aligned with the standards and textbooks, the disparities might increase defensive responses to psychometric feedback because teachers would not have the curricular resources they need to generate productive instruction in response. On the other side, of course, if teachers have access to a conceptually coherent, integrated and aligned set of curricular guidelines, we would like to see whether such a system fosters more productive instructional responses to feedback.

But there is another question that can be raised here as well: In this more ideal situation, where teachers have coherent and aligned guidance about goals and curricula, does psychometric feedback provide any additional benefit? This question is important because test data are likely to be the most expensive part of a curriculum guidance system, and we need to know whether it adds enough unique information to justify its cost. In fact, many studies that purport to examine the use of data systems are really examining curricular guidance systems more broadly. For example, Datnow, Park & Wohlstetter (2007) studied four districts that had

been identified as leaders in data-driven decision making. These districts had invested in coherent curricula as well as data management systems and had provided numerous tools—forms, checklists and the like—to help school groups study themselves. They also provided more time for teachers to examine data and to meet and discuss alternatives, and they provided training to develop capacity and to help form a culture of data use. Similarly, Young (2006) studied grade-level planning teams in four schools and found numerous "agenda-setting" supports for teachers; Marsh, McCombs, & Martorell (2009) looked at schools that combined data systems with instructional coaches.

In all these studies, districts made a variety of changes designed to provide teachers with more clear and coherent guidance regarding student learning goals. Suppose they had provided all of these goal-clarifying supports—the professional development, planning time, curriculum materials, coaches, team meetings, and so forth, *without also providing data*. And suppose they still have achieved beneficial results. Systems such as these provide teachers with a host of instructional supports intended to clarify both goals and alternative means for achieving those goals. So far, researchers have assumed that whatever benefits they saw came from the data system, not from the organized set of goal-clarifying guidance. We need to know if these systems provide benefits by themselves for two reasons. First, districts have not provided clear and integrated goal-oriented support systems in the past, so the effect of such systems is worth sorting out. And second, data systems are the most expensive component of such systems. If the rest of the system is useful to teachers by itself, it might enable them to make better use of their clinical feedback and remove the need for psychometric feedback altogether.

One example of such an alternative guidance system is described by Ma (1999), who noticed that teachers in China appeared to continually increase their understanding of the content they taught. On further investigation, Ma discovered that Chinese teachers' teaching manuals explained the content of each unit's mathematics to the teachers and also explained the kind of misconceptions and confusions students were likely to have. Chinese teachers also have more planning time than US teachers and have access to peers who are teaching the same content, thus increasing their opportunities to think about how best to teach this content, what kind of examples to use, and so forth. Such guidance can help teachers focus on learning goals and also help them make more sense of their clinical feedback, so that they may not need expensive psychometric feedback.

I asked at the beginning of this paper if the concept of data use might be a panacea because, by itself, data use lacks any instructional meaning. Advocates for data use assume that instructional planning is a rational process organized around a well-understood learning goal, and that psychometric feedback regarding progress toward that goal would help teachers plan and revise instruction. But if teachers' planning is based on satisficing rather than a rational process of weighing of alternatives, then the central premise behind arguments for data systems is faulty. Moreover, there is a potential for psychometric feedback to splinter content knowledge in such a way that it may actually discourage the kind of content-rich and intellectually-challenging instruction that students need. If this occurs, psychometric feedback could move our education system further away from the very outcomes its advocates claim it supports.

Constraints

The third component of teachers' problem space is constraints. Schools and school districts structure teachers' work in important ways. They schedule teachers' time, provide them physical space and with instructional materials and resources, allocate students to them, and impose various demands on them. Teachers tend not to think explicitly about the constraints that structure their work, but instead to accept them as givens. However, three constraints are particularly relevant to their ability to teach productively and to respond productively to psychometric feedback.

Distractions. Institutions disrupt teachers' plans by scheduling field trips, assemblies and tests, by making announcements over the public address system, putting telephones into classrooms which ring during class, and by sending central office staff into teachers' classrooms during class time to get signatures on paperwork (Kennedy, 2005). Institutions also frequently move students in and out of classrooms for special programs, and the schedules of these movements are not necessarily coordinated with teachers' own lesson schedules. In their cross national comparison of teaching practices, Stigler and Hiebert (1999) noticed that American teachers' lessons were more frequently interrupted than those of teachers in other countries.

Institutions also revise policies from year to year and regularly introduce reforms, requiring teachers to adjust their own work to accommodate these changes. In fact, the zeal for reform itself has reached such a fevered pitch that teachers can be confronted with one or more new reform initiatives every year. Each new superintendant and each new building principal feels obligated to introduce a new initiative. They shift to block scheduling, then away from it, into a zero-tolerance policy, then away from that, into a new textbook or a new curriculum framework, then to something else. Yet each initiative requires teachers to revise their own routines and strategies. Every new policy, from zero tolerance to team teaching, pulls teachers' attention away from their teaching and toward a new logistical problem of how to accommodate this latest innovation.

The net result of these distractions is that teachers must stop thinking about how to engage students with curriculum content and think instead about how to revise their procedures, schedules, or organizational strategies to accommodate the newest helpful idea.

Planning Time. Teachers typically are allocated one planning hour per day. During this single hour, they are expected to plan all of their daily lessons, read and respond to student work, assign grades to student products, and diagnose their learning progress. Much of the student work they read provides clinical feedback that can influence their further plans, but the volume of such work can be extensive.

In the US it is assumed that teachers contribute additional time at home in the evenings. When they do, the ratio of planning time to instructional time might be closer to 2:5—that is, two hours of planning time for five hours of instructional time. This expanded ratio is still far less than the ratios of teachers in China and Japan, which are closer to 2:1 (Kennedy in press). Moreover, it is not uncommon for American institutions to invade teachers' planning time with other agendas. Institutions convene teachers to review test results, engage in professional

development, plan assemblies and other special events, revise discipline policies or paperwork procedures, or engage in other institutional tasks.

It should not be surprising, then, that Kennedy (2005) found teachers rushing to class without having read the day's lesson in the textbook, forgetting to bring materials they planned to use, or discovering mid-lesson unforeseen problems with their plans or materials. Their limitation in planning time means that teachers rely heavily on their textbooks for direction, examples and homework assignments, and rarely spend much time thinking through the intellectual nuances of the content that they are teaching.

Instructional Time. Teachers are also constrained in the time available for instruction. Both the length of the school year and the schedule of breaks for vacations and holidays are fixed in advance. One of the first things teachers do as they plan out a new school year is to arrange the curriculum content along the contours of the calendar so that they can accommodate breaks, testing schedules or other interruptions. The primary goal of this long-range planning is to ensure that all the content can get covered but also to ensure that school breaks don't come up in the middle of any major curricular units. Though teachers can adjust these schedules somewhat, they cannot deviate too far or they risk not finishing the curriculum by the end of the year. If students fail to fully grasp a particular unit, teachers face a conflict between the need to spend more time on that unit and the need to move on to other material that must also be taught before the year is finished.

Notice that a new data system is highly likely to be added to the list of constraints within which teachers must plan and organize their work. They will need to reserve more instructional days for testing, reserve more planning days for data analysis, and perhaps spend more time coordinating their work with that of their peers.

Productive and Defensive Responses to Constraints

Institutional constraints can have important effects on instruction and can increase the likelihood of defensive instruction. They may force teachers to hastily throw their lessons together and interruptions during class can cause teachers to lose their train of thought, thus throwing the lesson off kilter. These effects are sporadic, but there is a more enduring effect as well: Constraints can weaken teachers' sense of efficacy and control over their work. Ingersoll (2003) argues that teachers have been given a role similar to foremen in a shop: they have very little control over their situation, but a great deal of responsibility. The sense of helplessness that can arise from these institutional constraints could, in turn, foster more defensive instructional practices. Teachers themselves have few opportunities to give concentrated thought to any intellectual concepts, let alone provide their students with such opportunities.

Psychometric vs. Clinical Feedback regarding Constraints

In order to think about how psychometric feedback might complement, compete with, or supplant clinical feedback in the context of constraints, let's extend the term "clinical feedback" to refer to things teachers learn through their daily interactions with everyone in the school. That is, much of teachers' understandings of their school policies and procedures is likely gathered through clinical sources rather than through formal top-down communications. Of

interest, then, is whether teachers might learn something from psychometric feedback that is pertinent to their perception of, or response to, their constraints and that might add to or differ from what they normally learn from clinical feedback alone.

There is a good chance that psychometric feedback could provide such knowledge. Every school has a status hierarchy, and Ingersoll (2003) has shown that there are visible patterns of favoritism that affect the allocation of space and time within schools. Teachers probably know, through their clinical feedback, which teachers are more or less favored by the administration. But psychometric feedback may point to a different hierarchy, one based on a merit criterion. Over time, allocations of coveted resources could change so that they are more aligned with student achievement patterns and less aligned with friendship patterns.

Notice, though, that even if psychometric feedback alters the allocation of rewards and perks, it may not alter anyone's understanding of *why* student achievement differs from one classroom to the next, for teachers rarely have opportunities to observe one another's instruction.

Psychometric Feedback and the Problem of Defensive Responses to Constraints

Probably the single most important feature of psychometric feedback is that it comes to teachers through a large and complex centralized system. As such, it adds to teachers' constraints in at least four ways. First, the assessments themselves consume some amount of teachers' instructional time, leaving less of it for teaching content. Second, the data presentations consume some amount of teachers' planning time, as teachers struggle to make sense of the feedback. Third, the system as a whole imposes a set of instructional goals on teachers that may limit their flexibility to respond to other instructional demands. Finally, the implementation of the system may be accompanied by institutional pressures to demonstrate growth which may frighten teachers. All four of these features of data systems could motivate more defensive instructional responses. Hence, any research on the introduction of such systems needs to examine the nature and extent of these constraints as well as how teachers respond to these constraints.

Another important feature of data systems, though, is that they alter the locus of instructional planning, generally encouraging more group-level planning and analysis and less isolated teacher-by-teacher analysis and planning. Though this may be welcomed by some teachers, it may also be considered a constraint in the sense that it structures teachers' work in different ways. In fact, even if teachers adapt to and embrace group planning, ultimately they still teach alone. An important question for researchers here is how school-wide and grade-level planning conversations affect an individual teacher's actual instructional activities.

Very few studies have actually examined the systemic influence of data systems, in part because very few districts actually have wholesale, functioning data systems that could be used to collectively plan instruction. From their national survey of teachers, Means et al. (2007) found that teachers frequently had access to data on attendance, demographics, course enrollment histories and the like, but fewer had access to student test data, and only 19% had

access to data on the students they were currently serving.

As new systems are introduced, it will be important to examine them in the context of existing constraints governing teaching practices. For example, in a school fraught with reforms and distractions, a new data system might be viewed as simply one more thing, this year's reform, this year's nuisance. But in a more stable environment, teachers might respond more productively.

There are also questions about how psychometric feedback interacts with other constraints, especially instructional time. Suppose a teacher completes a curriculum unit and then obtains psychometric feedback indicating that students did not learn as much as expected. The teacher is then faced with two options, neither of which is ideal: Take additional time to review the material, thus getting behind schedule, or move on to the next unit, thus maintaining the schedule at the expense of students' grasp of the content. The psychometric feedback does not tell teachers which is the preferable choice and the constraint of time is not negotiable.

A similar dilemma applies to differentiation. One recent study (Wayman and Stringfield 2006) examined teachers' responses to patterns of test data in high data-use schools. These authors asked the publishers who produce data systems to nominate their most successful schools for study, and then interviewed teachers in three of these schools. Not surprisingly, teachers in these schools were enthusiastic about the data systems, saying that they felt they had better knowledge of their students and a greater sense of efficacy. They reported more collaboration as well and said that the data were important to school-wide discussions and planning processes. Less was said about what teachers actually did differently within their own classrooms in response to the feedback. But teachers did say that the data enabled them to better differentiate among their students and to respond to students' individual needs:

> For instance, a pair of teachers in School B described a complex system of team teaching where the two teachers combined classes for certain lessons, then formed subgroups of students to split out for other lessons or small-group help. A group of teachers in School A described a process they use to identify individual students for enrichment or remediation, based on frequent local assessments. In all examples, the teachers reported a data system to be the primary support. [Wayman & Stringfield, 2006, pg 564]

While this finding is encouraging for the prospect of "data use" per se, it does not reveal much about the underlying relationship between the feedback and instruction. If teachers allow some students to review an earlier unit while other students move on, they may ultimately increase the variability among their students. At some point the school year will end and, at that time, the students who repeated earlier units will finish the year having learned less material than their peers. Differentiation is often listed as a benefit of psychometric feedback systems, but the problem of how differentiated instruction is balanced against the constraints of the school calendar is rarely considered.

The Value Added by Psychometric Feedback

My aim in this paper has been to examine the actual or potential value of data systems for teachers' instructional planning and decision making. Because these systems are new and

are not pervasive, much of the literature so far focuses on how the systems are being implemented more than on how they influence practice. My central concern has been that psychometric feedback systems are being introduced into school systems before their instructional value has been ascertained. The concept of data use has become a panacea, an ill-defined concept that is expected to solve a wide range of educational problems.

To "deconstruct" this concept, I propose a research agenda organized around two broad questions: Does psychometric feedback offer a value for teachers over and above that provided by their normal clinical feedback, and does psychometric feedback foster more productive or more defensive instructional responses?

In considering these two questions, I have also shown that teachers face numerous problems and constraints in their work: Their students may not be attentive or motivated, their curricular guidance can be confusing, inconsistent or vague, and changes in district policies and procedures can distract them from their instructional goals and limit the time they have for teaching and for planning. These aspects of teachers problem spaces raise a third question, which is whether, given the variety of problem teachers face, a new data system is the best investment for scarce district resources.

All instructional supports—professional development, curriculum materials, staff meetings, or data systems—have the potential to foster productive improvements in teaching or to foster defensive instructional responses. Productive responses are goal oriented and focus on student learning, while defensive responses consist of mindless coverage of textbook or test content. A central concern in this paper has been the question of whether data use will lead to productive instructional improvements or whether, instead, it will turn out to be a panacea. If the latter, the effects of data systems could be to spend large sums of money and to exacerbate the defensive practices we currently see in classrooms. Certainly there is much to be learned about the panoply of institutional forces that shape teaching practices, only one of which is centralized data systems.

References

Alonzo, Alicia C., and Jeffrey T. Steedle. 2009 "Developing and Assessing a Force and Motion Learning Progression." *Science Education* 93: 389-421.

Ayala, Carlos C., Richard J. Shavelson, Maria Araceli Ruiz-Primo, Paul R. Brandon, Yue Yin, Erin Marie Furtak, Donald B. Young, and Miki K. Tomita. 2008 "From Formal Embedded Assessments to Reflective Lessons: The Development of Formative Assessment Studies." *Applied measurement in education* 21 (4): 1-34.

Berlak, Ann , and Harold Berlak. 1981. *Dilemmas of Schooling: Teaching and Social Change*. London: Methuen.

Brandon, Paul R., Donald B. Young, Richard J. Shavelson, Rachael Jones, Carlos C. Ayala, Maria Araceli Ruiz-Primo, Yue Yin, Miki K. Tomita, and Erin Marie Furtak. 2008 "Lessons Learned from the Process of Curriculum Developers' and Assessment Developers' Collaboration on the Development of Embedded Formative Assessments." *Applied measurement in education* 21: 391-402.

Clark, Christopher M. , and Penelope L. Peterson. 1986."Teachers' Thought Processes." In *Handbook of Research on Teaching*, edited by Merlin C. Wittrock, 255-96. New York: Macmillan.

Cohen, David K. 1988."Teaching Practice: Plus Que Ca Change. . ." In *Contributing to Educational Change: Perspectives on Research and Practice*, edited by P. Jackson. Berkeley CA: McCutchan.

Cohen, David K., and Susan Moffitt. 2009. *The Ordeal of Equality: Did Federal Regulations Fix the Schools?* Cambridge MA: Harvard University Press.

Cremin, Lawrence A. 1990. *Popular Education and Its Discontents*. New York: Harper and Row.

Cusick, Philip A. 1983. The Egalitarian Ideal and the American High School: Studies of Three Schools. New York: Longman.

Datnow, Ananda, Vicki Park, and Priscilla Wohlstetter. 2007 "Achieving with Data: How High-Performing School Systems Use Data to Improve Instruction for Elementary Students." Los Angeles: University of Southern California Center for Educational Governamce.

Doyle, Walter. 1986 "Content Representation in Teachers' Definitions of Academic Work." *Journal of Curriculum Studies* 18 (4): 365-79.

Doyle, Walter, and Kathy Carter. 1984 "Academic Tasks in Classrooms." *Curriculum Inquiry* 14: 129-49.

Egan, Kieran. 1997. The Educated Mind: How Cognitive Tools Shape Our Understanding. Chicago: University of Chicago Press.

———. 2001 "Why Education Is So Difficult and Contentious." *Teachers College Record* 103 (6): 923-41.

Fenwick, D. Tara. 1998 "Managing Space, Energy, and Self: Junior High Teachers' Experiences of Classroom Management." *Teaching and Teacher Education* 14 (6): 619-31.

Floden, Robert E. , and Christopher M. Clark. 1988 "Preparing Teachers for Uncertainty." *Teachers College Record* 89: 505-24.

Furtak, Erin Marie, Maria Araceli Ruiz-Primo, Jonathan T. Shemwell, Carlos C. Ayala, Paul R. Brandon, Richard J. Shavelson, and Yue Yin. 2008 "On the Fidelity of Implementing Embedded Formative Assessments and Its Relation to Student Learning." *Applied measurement in education* 21: 360-89.

Goertz, Margaret E., Leslie Nabors Oláh, and Matthew Riggan. 2009 "From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction." Philadelphia PA: Consortium for Policy Research in Education.

Heilig, Julian Vasquez, and Linda Darling-Hammond. 2008 "Accountability Texas-Style: The Progress and Learning of Urban Minority Students in a High-Stakes Testing Context." *Educational Evaluation and Policy Analysis* 30 (2): 75-11-.

Ingersoll, Richard M. 2003. Who Controls Teachers' Work? Power and Accountability in America's Schools. Cambridge MA: Harvard University Press.

Jackson, Philip W. 1968/1990. *Life in Classrooms. Reissued with a New Introduction*. New York: Teachers College Press.

Kennedy, Mary M. in press. "Attribution Error and the Quest for Teacher Quality." *Educational Researcher*.

Kennedy, Mary M. 2005. *Inside Teaching: How Classroom Life Undermines Reform*. Cambridge MA: Harvard University Press.

Kerr, Kerri A., Julie A. Marsh, Gina S. Ikemoto, Hilary Darilek, and Heather Barney. 2006 "Strategies to Promote Data Use for Instructional Improvement: Actions, Outcomes, and Lessons from Three Urban Districts." *American Journal of Education* 112: 496-520.

Koretz, Daniel, Brian Stecher, Stephen Klein, and Daniel McCaffrey. 1994 "The Vermont Portfolio Assessment." *Educational Measurement: Issues and Practice* 13 (30): 5-16.

Kounin, Jacob. 1970. *Discipline and Group Management in Classrooms*. New York: Holt, Rhinehart and Winston.

Labaree, David F. 2000 "On the Nature of Teaching and Teacher Education: Difficult Practices That Look Easy." *Journal of Teacher Education* 51 (3): 228-33.

Ma, Liping. 1999. Knowing and Teaching Elementary Mathematics: Teachers' Understanding of Fundamental Mathematics in China and the United States. Mahwah NJ: Lawrence Erlbaum.

Means, Barbara, Lawrence Gallagher, and Christine Padilla. 2007 "Teachers' Use of Student Data Systems to Improve Instruction." Menlo Park, CA: SRI International.

Meehl, Paul E. 1954. Clinical Versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence. Minneapolis, MN: University of Minnesota Press.

Metz, Mary.   1993."Teachers' Ultimate Dependence on Their Students." In *Teachers' Work: Individuals, Colleagues and Context*, edited by J. Little and M. McLaughlin. New York: Teachers College Press.

Pellegrino, James W. "Knowing What Students Know." *Issues in Science and Technology* (2002), http://www.issues.org/19.2/pellegrino.htm.

Popham, W. J. 2006 "Defining and Enhancing Formative Assessment." Los Angeles.

Porter, Andrew C., Morgan S. Polikoff, and John Smithson. 2009 "Is There a De Facto National Intended Curriculum? Evidence from State Content Standards." *Educational Evaluation and Policy Analysis* 31 (3): 238-68.

Reynolds, Sharon , Kathleen Martin, and Judy Groulx. 1995 "Patterns of Understanding." *Educational Assessment* 3 (4): 363-71.

Rothkopf, Ernst Z. 2009 "An Immodest Proposal: Pedagogic Information Support for Teachers." *Teachers College Record* 111 (1): 164-79.

Schmidt, William H., C. C. McKnight, Lee S. Cogan, P. M. Jakwerth, and Richard T. Houang. 1999. *Facing the Consequences: Using Timss for a Closer Look at U. S. Mathematics and Science Education*. Boston: Klewer Academic Publishers.

Shattuck, Roger. 2005 "The Shame of the Schools." *New York Review of Books* 52 (6).

Shavelson, R. J. 2008 "Guest Editor's Introduction." *Applied measurement in education* 21: 293-94.

Shavelson, Richard J. 2006."On the Integration of Formative Assessment in Teaching and Learning: Implications for New Pathways in Teacher Education." In *Competency Oriented Teacher Training: Old Research Demands and New Pathways*, edited by Fritz K. Oser, Frank Achtenhagen and Ursula Renold, 63-78: Sense Publishers.

Simon, Herbert A., and Allen Newell. 1971 "Human Problem Solving: The State of the Theory in 1970." *American Psychologist* 26 (2): 145-59.

Songer, Nancy Butler, Ben Kelcey, and Amelia W. Gotwals. 2009 "How and When Does Complex Reasoning Occur? Empirically Driven Development of a Learning Progression Focused on Complex Reasoning About Biodiversity." In *American Educational Reseach Journal*. San Diego.

Stigler, James W., and James Hiebert. 1999. The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom. New York: The Free Press.

Tyack, D., and L Cuban. 1995. *Tinkering toward Utopia: A Century of Public School Reform*. Cambridge: Harvard University Press.

Wayman, Jeffrey C., and Sam Stringfield. 2006 "Technology-Supported Involvement of Entire Faculties in Examination of Student Data for Instructional Improvement." *American Journal of Education* 112 (4): 549-71.

Yin, Yue, Richard J. Shavelson, Carlos C. Ayala, Maria Araceli Ruiz-Primo, Paul R. Brandon, Erin Marie Furtak, Miki K. Tomita, and Donald B. Young. 2008 "On the Impact of Formative Assessment on Student Motivation, Achievement, and Conceptual Change." *Applied measurement in education* 21: 335-59.