Data for Improvement, Data for Accountability

JANET A. WEISS

University of Michigan

Turner and Coburn begin this special issue with the observation that data use has been a strategy for fostering improvement in public schools in the United States. The articles in the issue contribute to understanding the process by which these improvements might happen. They address the complex linkages from the provision of data to professionals, the encouragement and pressure to use the data in making decisions, the institutional contexts that support or discourage the use of data in making decisions, and the resulting practices that lead to educational outcomes.

As I read the articles, I am struck that some of the data systems discussed by the authors are intended principally to improve the performance of school staff, whereas other data systems are intended principally to hold schools and districts accountable for outcomes. Both types are expected to improve educational results, but, although both make use of data as an active ingredient in their strategy, the two types represent different logics of action and intervention (Weiss, 1999).

If data systems are to improve performance, the data systems should be designed to help school staff (especially teachers and principals) do a better job of instruction in order to enhance student learning or achieve other educational goals (such as attainment or skill acquisition). They seek to build the knowledge and skills of teachers and principals so that they can be more effective in their professional work. If, on the other hand, data systems are intended to hold educators accountable, the data systems should be designed to measure and report outcomes in a way that will allow actors outside the schools (often elected officials or other stake-

holder groups) to assess whether the investment of taxpayer resources has produced desired educational outcomes. They seek to put pressure on teachers and principals so that they work harder and smarter to produce better results.

The research described in these articles makes it clear that the design of a data system for improvement requires different choices than if the purpose were accountability. The authors help us to understand this important difference.

DATA FOR IMPROVEMENT

Marsh (2012, this issue) focuses her literature review on efforts to support data use by educators for actionable knowledge in support of improvement in schools or districts. She carefully examines a set of 29 interventions that are mostly intended to improve the quality of instruction by teachers. She observes that the interventions all included "human support" to help teachers to use data. These human supports include on-site technical experts, data coaches, tools for teachers to connect data to curriculum and student progress, and training and professional development in the use of data. These essential supports helped teachers to get access to data, understand the meaning of the data, and connect what they learn to their own practice with the students in their own classrooms.

She also notes that "making data 'safe'" appears to be another prerequisite for facilitating data use. By this she means that teachers feared being evaluated and judged by public reports of the test scores of their own students; they preferred anonymous, aggregated test results that could not be linked to individual teachers. She shows how principals and data experts set up norms, protocols, and language to limit the risks of critiquing instructional practices by not critiquing individual teachers. Trust in relationships within the school led teachers to believe that data would not be used against them, and thus made them more open to attending to data.

Marsh notes that structured social interactions were key ingredients of support for data use, so that teachers and principals could learn by discussing data with one another and had time during the work day to have these discussions. The use of data for improvement is not a one-person activity but is guided and nurtured by conversation and comparison. These collaborations make it more likely that data were used to inform instruction by encouraging teachers and principals to see how their peers were using student data. The research reviewed by Daly (2012) makes the same point—that the social networks of teachers can support their interest in data and their capacity to use the data in their own classrooms.

Whether educators can use data to improve practice depends in part on whether they have control over key elements of their own practice. If teachers do not have "the flexibility or authority to adjust instruction or curriculum according to what is discovered after analyzing the data, educators could not effectively act on the data" (Marsh, 2012, this issue). If the goal is improvement, improvement depends on influencing those actors who have the authority to take appropriate action.

The designers of data systems can encourage the likelihood that teachers and principals will use data for improvement by attending to these features: providing human support to help teachers to translate data into meaningful guidance; making data easily digestible; avoiding fears of evaluation and judgment by reducing the transparency of data reporting; creating respectful and supportive collaborations to support data use by individual teachers; and directing data to teachers and administrators who have direct control over the curricular and instructional choices that lead to student learning.

DATA FOR ACCOUNTABILITY

When data are used for accountability, a different set of issues arises, and quite different implications come to the fore. Data systems focused on accountability tend to direct data toward elected officials and federal and state agencies responsible for the distribution of resources, rather than directing data to teachers. These actors have the authority to set performance goals for teachers and schools and to allocate resources to initiatives and activities within schools. If the purpose of the data system is to measure performance so that policy makers can provide incentives to encourage good performance and discourage poor performance, data should flow to those actors who are responsible for overall performance and have the authority to direct policy and resource allocation within a district or a school. The needs and interests of these higher level administrators become paramount in the design of the data system.

When data are collected for accountability purposes, they must be comparable across schools, districts, or even states. Only if policy makers are measuring the same thing from one school to the next, or one district to the next, can they tell whether some are performing better than others. The result is that the data collected in accountability systems tend to be highly standardized so that they can be measured and compared across settings, even when those settings are heterogeneous in their educational context (large and small districts, rich and poor schools, older students and younger students, high-parent-involvement schools and low-parent involvement, and so on). Although this standardization is critical to allow

comparison and a fair and consistent policy environment, the standardization necessarily does violence to the understanding of performance in particular sites. As Supovitz's (2012) article in this issue makes clear, most kinds of standardized testing are not very helpful to teachers trying to improve the performance of their students. Most standardized test results do not give the teachers much guidance about what their own students understand or need. Global measures like annual yearly progress (AYP) summarize outcomes but do not diagnose difficulties or suggest promising avenues for changing performance. Instead, their purpose is to simplify the complex performance of a district, school, or classroom to capture in a single measure so that the performance of the educators in the district can be assessed and compared.

When the teachers, administrators, elected officials, parents, students, and other stakeholders for a given school or district understand that data are collected for accountability, they all try to enhance their relative standing on the measures that will be used. They especially try to avoid appearing to fail on the measures used. This, of course, is the point of any accountability system; the system is intended to give people incentives to raise their performance. However, the authors of the articles in this issue, in line with much other literature, remind us of the distinction between raising performance and raising the appearance of performance. Educators whose performance is being measured have an incentive to succeed on the measures that will be made public and that will have consequences. Performance on those measures may be different from the incentive to succeed in some larger or more significant sense. Distortions and manipulation to appear successful are inevitable features of data for accountability.

Jennings (2012, this issue) reviews research showing that some schools seem to respond productively and positively to the accountability for performance, whereas other schools focus on steps to appear successful, without necessarily improving the overall performance of the school. For example, in some schools, staff appear to ignore those students (like those who transferred in during the school year) whose results will not be counted in school performance measures rather than make a special effort to boost the learning of those students. In other schools, the staff seem to pay just as much attention to these students even though their test scores don't "count" on school performance measures.

If we expect educators to focus on improving performance on those things that are measured, the choice of what to measure is critical. These design choices will drive the responses of district and school leaders. Jennings reviews the research about the advantages and disadvantages of using measures of proficiency, growth, or a combination of proficiency

and growth to assess student learning. If the measure specifies a minimum level of proficiency that must be achieved to be considered successful (a cut score), school staff can be expected to focus on moving those students who are just below the minimum level to just above the minimum level in order to maximize their overall performance on that measure. If, however, the chosen measure is growth in performance over the school year, school staff can be expected to focus their attention on those students who are most likely to show improvement from one testing period to the next. Jennings's review shows that teachers and principals become highly sophisticated about targeting their efforts so that they can show results to their best advantage on these measures. Marsh also comments on the gaming of data systems that follows when data are used to hold people accountable for performance. Given that the choice of measure seems to produce distributional effects that may differ from what policy makers intended, Jennings's analysis highlights the difficulty of selecting measures that map directly onto the behavior that policy makers wish to promote.

Henig's (2012, this issue) article reviews some of the institutional and theoretical reasons that a focus on data for accountability leads to this result. He reminds us that only some data are collected (whereas others are not) because those data serve the interests and needs of more powerful political actors. School boards, state legislatures, and the federal government adopt data-based policies to hold schools accountable because these are ways for elected officials to demonstrate to voters that they can control the professional activities of educators. The educational agencies at all levels that implement these policies then seek their own control over activities by schools and teachers and jealously guard their expertise and authority over data collection and interpretation. Henig's review reminds us that education data become ammunition in the hands of the various constituencies (elected officials, community leaders, parents, business) seeking control over school activities. When this happens, the teachers and principals whose performance is being measured are often motivated to conceal rather than reveal the challenges they face and the difficulties they encounter. Instead of data becoming a source of guidance for teachers about how to improve student outcomes, data can easily become an arena for teachers to conceal performance so as to protect themselves from the critiques offered by parents, interest groups, elected officials, or higher level administrators.

Henig does point out that some states and some communities have succeeded in creating comprehensive data regimes that protect data collection from becoming an arena for political dispute. In these places, systematic attention to data has moderated the effect of interest group

politics instead of being captured by these politics. Data use for accountability is not necessarily incompatible with the process of constructive data use. However, Henig shows that sustained constructive leadership, careful design, and implementation are crucial to counteract the political forces that are unleashed by high-stakes measurement systems.

The authors of these articles have helpful advice for designers of data systems for accountability. (1) Attend carefully to the selection of measures of performance to recognize that teachers and administrators will seek to perform well on those measures (regardless of whether the measures reflect the full range of performance of interest). (2) Using a broad set of indicators seems in principle to be desirable to capture more accurately the range of educational outcomes of interest to the community. However, do not necessarily expect that strong performance on one measure will imply strong performance on other measures. For example, Jennings notes that schools with high test scores do not necessarily produce strong results in social development or encouraging college attendance. (3) The authors point to the difficulties created for school officials by the simultaneous pressures of local, state, and federal accountability regimes that are not necessarily compatible or coordinated. If local, state, and federal officials coordinate and align their performance expectations, local schools have a much clearer path to follow to improve performance. When expectations are incompatible, none of the accountability regimes is likely to produce the intended effects.

This special issue makes a contribution by illuminating the choices faced by policy makers in figuring out how to use data to get better results in schools. The two major approaches to using data can inform a wide variety of policy choices.

If the goal is to get teachers to use data in classrooms to improve student learning, the premise is that teachers will teach more effectively if they have better information about which students are learning what, how, and why. If the goal is to hold district leaders accountable for school performance, the premise is that district leaders will exert more effective control by using data about school to guide the use of incentives and sanctions. The selection of measures of school performance, the transparency and disclosure of data, the level of aggregation at which data are reported, the training in the use and analysis of data, the frequency of data reporting, and the kinds of support and follow-up are all likely to vary depending on the goal. Because both approaches prominently feature data and data use, they are sometimes considered to be highly similar and guided by the same design concerns. The authors in this special issue are convincing that the lessons of one type of data-based policies (for example, collect data frequently enough to allow teachers to monitor changes in

student learning and make corrections when appropriate) cannot be easily transferred to the other (for example, collect data only when enough time has passed to give educators a reasonable opportunity to show that improvement in outcomes has occurred).

As educational researchers, we too need to be alert to the differences among policy logics. When the policy's primary goal is to help teachers improve their work with students, research questions should explore the ways in which teachers understand and respond to the data collected and reported, and the ways in which consequences flow to teachers themselves from data. When, on the other hand, the policy's goal is to design ways to hold schools accountable, research should focus on the complex causal chains by which decision making at federal and state levels can have consequences for educational practices and student outcomes within schools. Researchers interested in the roles played by data in educational policy and outcomes have made substantial headway in identifying these issues and questions, but we have much to learn as we go forward.

References

Daly, A. J. (2012). Data, dyads, and dynamics: Exploring data use and social networks in educational improvement. *Teachers College Record, 114*(11).

Henig, J. R. (2012). The politics of data use. *Teachers College Record, 114*(11).

Jennings, J. L. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record, 114*(11).

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11).

Supovitz, J. A. (2012). Getting at student understanding—The key to teachers' use of test data. *Teachers College Record, 114*(11).

Turner, E. O., & Coburn, C. E. (2012). Interventions to promote data use: An introduction. *Teachers College Record, 114*(11).

Weiss, J. A. (1999). Theoretical foundations of policy intervention. In by H. G. Frederickson & J. M. Johnston (Eds.), *Public management reform and innovation* (pp. 37–69). Tuscaloosa: University of Alabama Press.

JANET A. WEISS is vice provost and dean of the Rackham Graduate School at the University of Michigan. She was a member of the Network on Learning and Teaching funded by the MacArthur Foundation and is a coauthor of "How School Choice Affects Students Who Do Not Choose" in J. Betts and T. Loveless (Eds.), *Getting Choice Right* (Brookings Institution Press, 2005), and "Toward a Deeper Understanding of the Educational Elephant" in J. Bransford, D. Stipek, N. Vye, L. Gomez, & D. Lam (Eds.), *The Role of Research in Educational Improvement* (Harvard Education Press, 2009).