



Given the primary role of teachers in affecting student achievement, U.S. policy makers and reformers have increasingly focused on monitoring and evaluating teacher effectiveness by emphasizing the links to student learning outcomes. Large-scale international assessments are frequently used as base examples to justify reform. But, relatively little is known about what other countries actually do. We wonder: How do other countries evaluate teachers?

How do **other** countries evaluate teachers?

Countries known to outpace the U.S. in student achievement use a variety of educational and organizational methods, but rarely use the approaches to education reform that we are promoting.

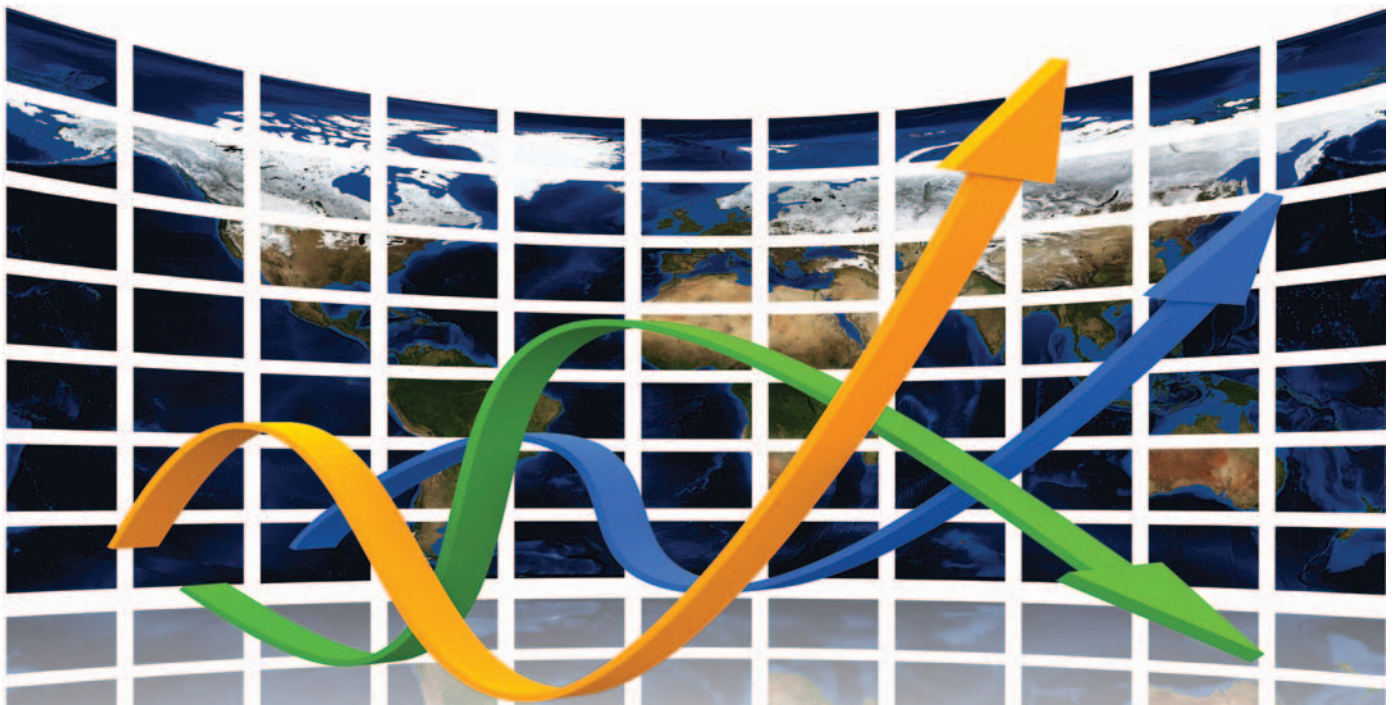
By James H. Williams and Laura C. Engel

We have set out on a broad research effort, looking at whether top-performing countries use educational practices and reform initiatives in vogue in the U.S. We've compared Finland, Korea, Japan, Ontario in Canada, and Singapore, exploring in each system the role of high-stakes testing, policies used to motivate schools and teachers to improve student learning, and the organization of accountability for learning. We use Ontario because Canadian education policy is substantially decentralized to the provinces. In each area, we looked at the role of teachers and systems of teacher evaluation.

Perhaps not surprisingly, we have learned:

1. Teacher evaluation is used for both accountability and instructional improvement in most school systems. However, teacher evaluation systems are organized differently depending on the model of accountability.
2. There is a growing trend to use student test results and metrics to inform accountability for schools, principals, and teachers, instructional improvement in classrooms and schools, and reform at the system level.
3. In particular, standardized testing of students, a primary and growing

JAMES H. WILLIAMS (jhw@gwu.edu) is an associate professor of international education and international affairs, and **LAURA C. ENGEL** (lce@gwu.edu) is an assistant professor of international education and international affairs at The George Washington University, Washington, D.C.



component of teacher evaluation in the U.S., is generally administered and used differently in other countries.

How is teacher evaluation linked with accountability and instructional improvement?

Looking across systems, we see four primary approaches to accountability: professional, organizational, market, and parental/community. Each approach has strong implications for teacher evaluation and its use in instructional improvement.

Professional accountability results from practitioner identification with the profession and a corresponding internalized obligation to uphold, even advance, its standards. Professional accountability is enhanced by social recognition and prestige (Scribner, Cockrell, K., Cockrell, D., & Valentine, 1999). In a professional accountability mode, teacher evaluation is closely linked with professional norms and peer assessments.

Finland's teacher evaluation system is based almost entirely on professional accountability, in which teachers are accountable to each other, the school, the children, and their parents. In the early 1990s, Finland abolished the school inspection system that was in place to evaluate teachers and provide external feedback. Now, teacher evaluation is more group-based, reflective, and participatory, with the aim of creating professional learning communities among teachers and administrators (Sahlberg, 2011). Evaluation is ultimately a consultative and formative process. Principals often use their own knowledge and experience as teachers to assist teachers and help them recognize areas of strengths and improve areas of weakness. Organizational accountability exists, but its primary purpose is to coordinate and lead the professional activities of teachers rather than command and control. Poor performance in relation to professional norms violates the trust that is said to characterize the system.

Another example of professional accountability is Japan's practice of lesson study, in which teachers, new and seasoned, take turns presenting lessons that are practiced and critiqued in a group setting. This system, while certainly not the only mechanism of accountability in Japanese schools, reinforces teachers' accountability to each other according to norms of good teaching.

Organizational accountability refers to the structures, norms, incentives, and sanctions of the formal institution. Organizational accountability is effective to the extent that the organization can compel or motivate individual employees to carry out its wishes. This is easier when the organization's goals are clear and broadly shared — when the means to achieving organizational goals are known and man-

ageable with available resources, and when employees want to do so. These conditions can be difficult to achieve in education because goals are multiple and overlapping, means for achieving goals are imperfectly known, and factors affecting student achievement are only partially under teachers' control.

Organizational accountability is part of most school systems. In some systems, teacher evaluation is informal or norm-based, with details left to individual schools or local education units. Other systems have developed standard external criteria or sophisticated rubrics and instruments to collect data, sometimes including student achievement results. Teacher evaluation may be mandated as in most systems, or voluntary as in Mexico's Carrera Magisterial (CM). CM is a voluntary performance-based salary and bonus system initiated in 1992. Research on CM in Mexico found a statistically significant positive relationship between CM scores and student performance as well as a strong positive relationship between CM score and school socioeconomic status. This means that higher-scoring teachers were more effective, but they also tended to end up in better-off communities. In the context of Mexico's decentralization, the differences across schools and communities and teacher quality grew even more pronounced (Luschei, 2012).

Market accountability is a result of consumer selection among competing services or products (Garn, 2001). Organizations responding to market signals will compete to deliver the most desired services at the lowest cost. Market accountability calls for schools to compete to provide the best services demanded by parents and students. It uses parental choice, voucher programs, and merit-based teacher pay to stimulate performance. Depending on how incentives are structured, market mechanisms may encourage competition among members of a unit, or cooperation among members of a unit competing with other units. Teacher evaluation in a system organized according to market accountability involves clear signals to the "market" about the effectiveness of schools and teachers. Student achievement test scores are often used to provide this signal. Whether this is an appropriate use of scores is a question of validity and dependent on the test.

Two key components of a market-oriented teacher evaluation system are 1) what is done with the results, especially failure and 2) how incentives are organized in terms of teachers' performance. If teachers are at high risk of losing their jobs and schools of being closed, teacher evaluation becomes extremely high stakes. If teachers see something they can do to improve, they're likely to do so. However, if they see nothing they can do, market accountability is likely to reduce teacher performance. Systems that offer

Despite trends to link student achievement and accountability, few countries use student achievement scores as the primary criterion for teacher evaluation.

incentives, second chances, and support are likely to be more successful.

Singapore's annual teacher evaluation system is one example. Each teacher is evaluated annually according to his or her performance against specific targets. Teachers are also evaluated for their Currently Estimated Potential (CEP), which measures their long-term potential and identifies training needs and overall aptitude. Performance ratings from formal evaluation and the CEP are used to determine annual performance bonuses. The system rarely results in teachers being fired, though that is possible. Rather, CEP highlights areas of improvement, particularly for beginning teachers, who are coached intensively by a mentor, and given one year to show improvement. Another key component is the "package" of criteria used to evaluate teacher performance, which includes test scores *and* information on their collaboration with peers and their community (Tucker, 2011).

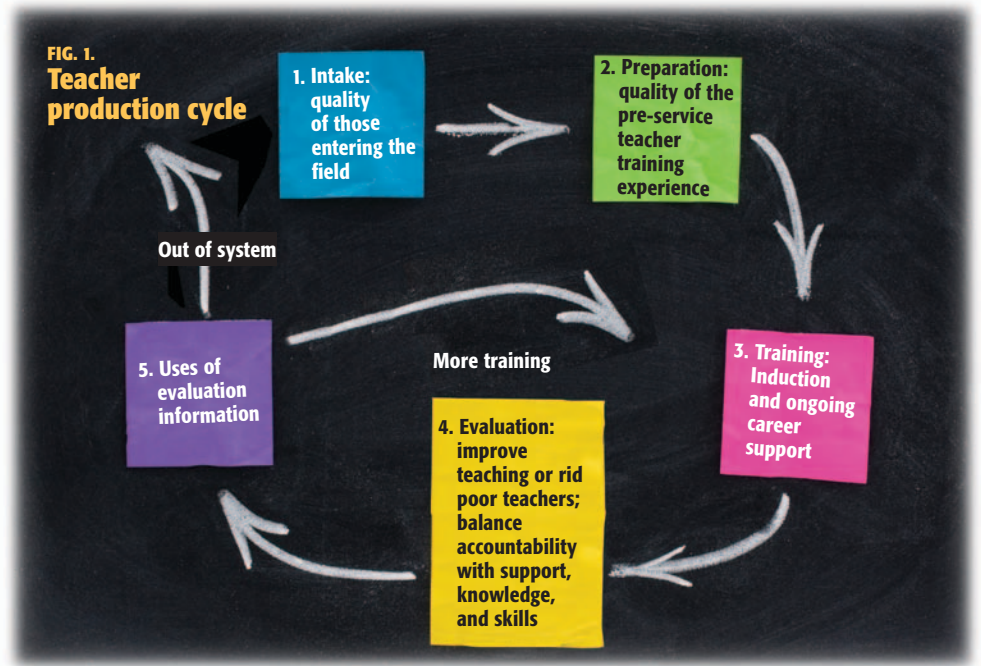
Parental/community accountability is an attempt to describe the informal, bottom-up accountability resulting from parental and societal pressure on schools, teachers, and students to do well. Parental/community accountability is more of a sociocultural phenomenon than purely an economic one. It varies among families and communities. Typically, families with high socioeconomic status or aspirations demand more accountability. It also varies across cultures, with Confucian cultures tending to hold teachers, schools, and especially students highly accountable for student learning.

Parental support may exist independently or in conjunction with more formal modes of accountability. While not formalized, parental support nonetheless is remarkably effective in contexts where it predominates. Tucker (2011) attributes much of Japan's academic success to the support and pressure of parents on teachers and schools to do well. Korea and Japan achieve accountability through high levels of parental support and pressure for education, coupled with strong ministries of education and high professionalism among teachers. There is a great deal of professional development at the school level, and responsibility for effective teaching is both individual and shared. Professionals and education authorities do not always see eye to eye, and there is a history of competition and debate of educational issues. Further, extensive cram school markets exist outside the public

school systems to meet demand from students and parents for additional lessons.

School systems rarely adopt one mode of accountability to the exclusion of others. For example, Ontario employs a mix of organizational, professional, and parental/community accountability. Nonetheless, in general, one mode is favored. Compared to the other five countries, U.S. reform relies most heavily on market accountability, though organizational, professional, and parental accountability are also present.

Teacher accountability should be viewed in the



context of what might be called the "teacher production cycle." Teacher evaluation, for example, whether for accountability or instructional improvement, surely plays a different role in a system where only the top 20% of applicants are admitted to teacher training as compared to one where most applicants are admitted. In the latter, teacher evaluation certainly serves a greater quality control function, perhaps serving to weed out ill-suited teachers.

In the teacher production process, teacher quality is improved, and teacher evaluation is most effective when quality control and enhancement takes place at all stages, rather than relying exclusively on evaluation. Specifically, teacher evaluation is most efficient where it supports and is supported by other parts of the production process: where there are high levels of intake, preparation and induction, and ongoing training and support, and where evaluation serves to increase teachers' commitment, morale, knowledge, and skill. Finland, Japan, Korea, Ontario, and Singapore provide many of these conditions, though in different ways.

If we believe teachers are the single most important element of schooling, then policies, including teacher evaluation policy, ought to generate rather than deplete the nation's teaching capital.

The design of teacher evaluation schemes depends in large part on what their designers know about good teaching and how good teachers are developed. If good teaching is something that any intelligent, academically prepared, and willing individual can do without a lot of training, teacher evaluation might serve primarily as a selection mechanism. If, however, good teaching is a professional skill developed over time with experience and through relationships with other professionals, then teacher evaluation might serve more of a signaling and formative mechanism.

Linking achievement and accountability

Undeniably, use of student achievement tests has grown internationally, and, increasingly, education policy makers have looked to test results to evaluate student learning and schools. Many schools around the world are now required to report on their performance and to use different forms of internal and/or external evaluation data, including student achievement data, to demonstrate performance. Student data are now readily available, precise, and assumed to be objective, neutral, and comparable. (Of course, their validity depends on the test, population, and purpose.) According to this logic, testing gives all stakeholders information about how well teachers

are teaching their students to learn. Disaggregating scores by subgroups allows for checks on the progress of groups whose suboptimal performance is often masked by overall averages. Seen in this way, testing provides the critical linkage in accountability for student learning.

Moreover, ministries of education and international organizations are increasingly emphasizing the importance of effective teachers in student achievement. In 2005, the Organization for Economic Cooperation and Development (OECD) issued a report, *Teachers Matter*, which emphasized the quality of teachers as a key factor affecting student outcomes. The report highlights the need to incentivize teacher practice and advocates for a transparent, fair process of evaluating teachers. The underlying belief: "The interests of students will be better served where teachers achieve employment security by continuing to do a good job, rather than by regulation that effectively guarantees their employment" (OECD, 2005, p. 163). What is interesting is the combination of emphases — doing a good job, employment security, and using a range of data. The OECD report has been influential in reshaping how countries think about teacher evaluation. Korea, for example, recently unveiled a new annual teacher evaluation system influenced by these ideas.

Despite these trends, however, *most countries don't use achievement test scores as a primary means of evaluating teachers*. In some systems, teacher evaluation takes place between the teacher and administrator, with little formalized data or sophisticated rubrics and instruments. When student learning is taken into account, it is often a broader conception of learning than reliance on student results on standardized tests. For example, teacher observations by administrators, formal and informal, are often considered key evidence of good teacher practice and student learning. Some systems have more open and reflective models of teacher evaluation that include peer reviews used mainly to improve practice rather than to evaluate it.

In Finland, high-stakes testing plays no role in teacher evaluation. The only standardized tests are sample-based assessments given at grades 6 and 9 to inform policy and curricular decisions, and the matriculation examination for university admission. Finnish students are tested, of



"I was a teacher. I miss snow days."

course, early and often, but in the form of formative classroom assessments to help teachers improve instruction and identify children needing additional support. In contrast, testing plays an outsized role in Japan and particularly Korea, but it is the university entrance examinations, not standardized student tests given at school, that carry such high stakes. And it is students and parents, especially those aspiring to the highest levels of education, rather than teachers, who are most affected by testing. Although secondary teaching is greatly influenced by preparation for entrance examinations, teachers do not risk losing their jobs if students fail. And while in Singapore, Ontario, and Mexico student test results are used to hold teachers more accountable for student learning, they are but one of several elements in the teacher evaluation package.

Nonetheless, systems are shaped by trends linking student achievement with accountability, and the use of both accountability and data are on the rise. Japan is an interesting case. For many years Japan used no standardized testing of students to evaluate student achievement on the grounds that such testing would lead to competition among students, teachers, and schools. When national tests were used, the tests were sample-based and used to inform system-level decisions. Recently, with a drop in relative ranking on international tests, especially the Program for International Student Assessment (PISA), Japan has reintroduced standardized testing and is said to be thinking about ways to include student performance as part of larger teacher evaluation packages. Despite resistance from its education ministry, there is increasing demand for information on schools' comparative performance.

Developing “teaching capital”

One way to assess teacher policy is to ask whether it builds or erodes what we might call a country's teaching capital — the capacity of a country's teachers to educate all of its children. The concept is almost identical to the idea of professional capital developed in parallel by Hargreaves and Fullan (2012). If we believe that teachers are the single most important element of schooling (OECD, 2005), then it would seem logical to adopt policies, including teacher evaluation policy, to generate rather than deplete the nation's teaching capital.

Which policies might do that depend greatly on how one understands the development of good teaching. If, for example, good teaching is understood as a trait — good teachers are born not made — then developing a nation's teaching capital means finding ways to recruit and identify those who were “born” teachers and discourage or fire those who were not. U.S. reform discourse, though it widely

embraces the idea that all children can learn, seems to view teaching as an individual trait, rather than a professional norm that almost anyone can acquire (Elmore, 1996). We see the same emphasis in the OECD report, which emphasizes that “it is not possible for everyone to be an effective practitioner and to sustain that over the long term” (2005, p. 12).

If one understands good teaching as a matter of will — teachers can teach well if they want to — then building teacher capital involves incentivizing good teaching behavior. This model is partially exemplified in Singapore and Mexico. If one understands teacher production as an ongoing learning process, then building a nation's teaching capital calls for hiring teachers with potentially good teaching traits and providing incentives to teach well. In addition, it should promote a collegial, safe learning environment that encourages trial, allows and corrects for errors, gives teachers professional time, space, and autonomy, and provides social and monetary respect for their profession while generally valuing them as professionals. Finland comes closest to this pure model. Finland has made teaching a prestigious and competitive field to enter, provides ample initial and ongoing learning opportunities for teachers to develop their professional skills, “trusts” them to use that professionalism in teaching, and relies on professional accountability to ensure quality. ■

References

- Elmore, R. (1996). Getting to scale with good educational practice. *Harvard Education Review*, 66 (1), 1-26.
- Garn, G. (2001). Moving from bureaucratic to market accountability: The problem of imperfect information. *Education Administration Quarterly*, 37 (4), 571-599.
- Hargreaves, A. & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. New York, NY: Teachers College Press.
- Luschei, T.F. (2012). In search of good teachers: Patterns of teacher quality in two Mexican states. *Comparative Education Review*, 56 (1), 69-97.
- Organization for Economic Cooperation and Development. (2005). *Teachers matter: Attracting, developing, and retaining effective teachers*. Paris, France: Author.
- Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland*. New York, NY: Teachers College Press.
- Scribner, J.P., Cockrell, K.S., Cockrell, D.H., & Valentine, J.W. (1999). Creating professional communities in schools through organizational learning: An evaluation of a school improvement process. *Educational Administration Quarterly*, 35 (1), 130-160.
- Tucker, M. (2011). *Surpassing Shanghai: An agenda for American education built on the world's leading systems*. Boston, MA: Harvard Education Press.

Copyright of Phi Delta Kappan is the property of Phi Delta Kappa International and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.