

**Teacher Evaluation:
Global Perspectives and Lessons for Pakistan**
Gulab Khan*, Asif Khan**, Sajjad Hussain *** & Nasir Shaheen ****

Abstract

Schools strive to raise student achievement by improving teacher quality. In turn, teacher quality can be improved by evaluating teachers, assessing their needs for professional development and by making them answerable for quality in teaching learning processes in schools. It is for this reason that National Education Policy (2009) of Pakistan puts a clear emphasis on judging teachers for the quality of their performance. With a review of this policy underway and the new policy may coming into effect anytime soon, it becomes paramount that teacher evaluation is looked at in global perspectives to draw lessons from. This article summarizes teacher evaluation practices in select OECD and non-OECD countries with the primary objective of informing policy options around teacher evaluation as stipulated in NEP 2009. The article also presents a snapshot of empirical evidence around developmental and high-stakes methods and approaches of evaluating teachers including a synthesis of evidence on Value Added Models. The article posits that finding an appropriate blend of the developmental and high-stakes means with a more focus on the former in any teacher evaluation regime should be based on rigorous research in indigenous settings.

Keywords: Evaluation, Monitoring, Accountability, Student' achievement, Teacher quality

Introduction

Raising achievement for all students is a central goal in modern democracies and states and schools pursue this goal by identifying

* Dr. Gulab Khan, Assistant Professor, School of Electrical Engineering and Computer Science (SEECs), National University of Sciences and Technology, Email: gulab.khan@seecs.edu.pk

** Dr. Asif Khan, Assistant Professor, Department of Educational Development, Karakorum International University, Gilgit.

*** Sajjad Hussain, Lecturer, Center for Education and Staff Training, University of Swat.

**** Nasir Shaheen, Assistant Professor, Centre for Management & Commerce, University of Swat.

and actively working around the factors that are significant in this regard. Among the many factors that have a close association with student achievement, teacher quality comes forth as critical factor in schools. Therefore, schools are being increasingly asked to evaluate teachers so as to improve the quality of teaching and hence student achievement. The purposes of these evaluations are to identify professional development needs of teachers and also to make them answerable for what they invest in their efforts to improve teaching and learning quality in schools.¹

In this backdrop, it is increasingly believed that all students can experience an enhanced quality of education by adopting teacher evaluation as a significant policy device. For such perspectives from around the world and to be at par with other nations in relation to quality of education, Pakistan's National Education Policy (NEP) 2009 lays out a clear emphasis on institutionalizing monitoring and evaluation across the board in the education systems. In this context this article reviews relevant literature to offer a comparative view of the different approaches to assessing and judging teacher effectiveness in schools. The article explores teacher evaluation in cross-national perspectives with the primary purpose of identifying nuances of practices and purposes of judging teachers so as to inform education policy in teacher evaluation in Pakistan.

Teacher Monitoring and Evaluation in Pakistan's National Education Policy

Since its inception, Pakistan has been struggling with the perennial issues in public education including quality, access and governance. Lamenting on the low confidence that the public poses in public institutions, including education, National Educational Policy (NEP) states:

*"To revive confidence in public sector education system by raising the quality of education provided in government owned institutions through setting standards for educational inputs, processes and outputs and institutionalizing the process of monitoring and evaluation from the lowest to the highest levels."*²

However, contrary to the above policy action, teacher evaluation in public schools is still a rare phenomenon in Pakistan. The modus operandi of assessing teacher performance is still the decades old ACR (Annual Confidential Report) that has almost no value with regard to raising teacher quality. Therefore, NEP's above

standpoint seems valid with regard to institutionalizing monitoring and evaluation in schools. In this connection, this paper explores the question: What lessons Pakistan can learn from experiences and evidence from around the world to develop a viable route to teacher evaluation. The paper reviews relevant empirical evidence from around the world on the link between teachers and student achievement. In so doing, it gives a comprehensive overview of how teachers are evaluated in OECD and non-OECD countries that Pakistan can learn from.

Why and How of Teacher Evaluation

Teacher evaluation, used here interchangeably with teacher appraisal, are the reviews of performance that various personnel carry out in schools against set plans and expectations. Such reviews may have formative purpose of identifying specific professional development needs or high-stakes and summative intentions associated with decisions such as promotion, rewards or penalties for teachers³. This essentially gives forth to two broad aims for teacher evaluation i.e., developmental or formative purposes and accountability or high-stakes purposes. High-stakes teacher evaluations aim at making teachers accountable for their performance in teaching whereas developmental purposes have the primary focus on identifying professional development needs of teachers.⁴

Teacher Evaluation in Cross-National Perspectives: A Synthesis of Evidence

Various studies and reports show that practices and purposes of evaluating teachers vary both within and across countries.⁵ A survey conducted in 2002-04 by the OECD showed that different countries had in place a variety of mechanisms with varying tools, objectives and outcomes of evaluating teachers. About 50% of the countries evaluated teachers on a regular basis. Twenty three percent countries had teacher evaluations with no developmental focus whereas 35% of the countries had professional development links in their teacher evaluation systems. Teacher evaluation in the remaining countries varied in how they evaluated and incentivized teachers. The survey showed Chile as a very active user of teacher evaluation systems with about 2/3 of the evaluation systems having professional development as an objective. It was also seen in the survey that the countries with professional development as one of the aspects of teacher evaluation, also had consequences for teachers who were ineffective. Teacher evaluation in such

countries led to creation and implementation of a plan for improvement, withholding of promotion or forfeit of salary.

In the 2009 PISA survey by the OECD, school principals in sampled countries responded to a variety of questions on teacher evaluation in schools. Items directly related to accountability and evaluation of teachers sought information on how achievement data was used for judging effectiveness of teachers and for the purposes of monitoring teachers. The findings show that countries significantly differed in the purposes of uses of assessment data to evaluate teachers. On average, 59% students in the OECD countries were enrolled in schools that monitored teachers using student achievement data.⁶ In Austria, Mexico, United Kingdom, Poland, Turkey, United States, and Israel 80% students studied in schools that practiced using student achievement to monitor and evaluate teachers. There were countries that used appraisals by peers, senior staff, external evaluators and principals in conjunction with student achievement. As well, there were countries that had negligible focus on external evaluations. For example, in Finland, a meager 2% students studied in schools that had any external evaluations whereas internal monitoring and evaluation had a much less focus in these schools.

Other measures to evaluate teachers and make them accountable were linked with team or school evaluations. These measures included high-stakes purposes such as public accountability, tracking of assessment of students administratively, giving parents information about their children's performance, comparing schools and benchmarking at the levels of district and nation. OECD-2010 also reported that 37% students, on average, were enrolled in schools that used public accountability in their high stakes teacher evaluation systems. Schools would make teachers accountable to public by posting student achievement data for public consumption through such channels as online media including websites as well in print media channels. However, there were countries that had a lot less focus on public accountability. Such countries included Austria, Spain, Belgium, Finland, and Switzerland. United Kingdom and United States had a predominance of public accountability as more than 80% students were enrolled in schools that used this approach to ensure quality in teaching and learning in schools.

A more detailed and comprehensive picture of teacher evaluation comes forth in the Teaching and Learning International Survey (TALIS) by the OECD. According to OECD⁷, over 50% teachers said that they had not experienced external evaluations in

the last five years suggesting a largely internal evaluation scenario in the sampled countries. It was also shown that over 50% of the teacher appraisal was based on the student test scores in most of the countries included in the TALIS 2008.

A significant element of a teacher evaluation regime is often the impact it has on various aspects of teachers' lives and schooling. In this regard, TALIS survey in 2008 extracted information on how teacher appraisal and feedback impacted teachers and their daily lives in schools. Teachers reported that an average of 41% of the emphasis was accorded to pushing up student test scores in their evaluations. Many countries including Italy, Bulgaria, Australia, and others emphasized greatly on making student test scores part of teacher appraisal systems. Along with test scores, other areas of teacher appraisal systems that impacted their daily practice were the plans for professional development, teaching and learning practices, and how teachers managed their classrooms. OECD found that teachers' approaches to managing classrooms in Belgium, Norway, Mexico, Slovenia, Hungary, and Australia were most impacted by teacher appraisal practices.⁸

In the TALIS 2008 survey, OECD reported that fewer teachers experienced outcomes relating to financial and career aspects of their lives through teacher appraisal practices. In Belgium (Fl), for example, a very low of only 0.4% teachers experienced a change in their monetary benefits including salary as a result of their appraisals. Malaysian teachers (33%), on the other hand, said that teacher appraisal lead to a medium to higher change in their salaries. In Slovenia, Lithuania, Poland, and Bulgaria, teacher appraisal impacted professional development of teachers to a great extent. Similarly, in Lithuania, Bulgaria, Mexico, Brazil, and Poland, teachers reported that their appraisal had a high effect on how they taught in classes. Parallel to this finding, in these countries, a higher number of teachers reported having a focus on improving student test scores as a focus in their appraisals.

Empirical Link on Teacher Evaluation

Isoré notes that developmental and high stakes purposes of teacher evaluation work in close proximity and are not mutually exclusive. In this article, however, an arbitrary division with the objective of creating a simpler structure to the available literature has been made by classifying research around three broad streams. The research stream that has often used such approaches as classroom observations and performance rubrics has been termed as standards

based teacher evaluation with the developmental objectives in focus. The second stream in empirical research often makes use of student test scores as essential metric for measuring teacher quality and performance. Third, a significant body of more recent evidence points to an extended and more advanced approaches known widely as Value Added Models (VAMs) that come with their own challenges.⁹

Wenglinsky¹⁰ posits that quantitative approaches to assessing and evaluating teacher quality have often lagged behind in capitalizing on the importance of within-classroom processes that are significant with regard to explaining variation in performance among students and classrooms. In this sense, Wenglinsky's study seems to have paved ways for various latter studies that uncovered within-classroom factors that were significant with regard to raising student achievement. To a similar effect, Holtzapple¹¹ carried out analyses on 16 standards instead of 22. Findings showed that the TES was successful in predicting performance at the extremities (unsatisfactory and distinguished) of performance ratings. However, the system was not as effective in predicting performance at the proficient and basic level (the middle of rating scale). Holtzapple found that a rating of "basic" and "unsatisfactory" in the domain of "Teaching and Learning," affected student achievement by lowering of student test score in comparison to the predicted score that was based on the achievement in prior year. On the contrary, a teacher rated as "distinguished" showed that students taught by such teacher achieved as expected.

Other studies further explored the link between standards-based teacher evaluation approaches and student achievement and found mixed results. Like Holtzapple and Kimball, White, Milanowski, and Borman and others, Milanowski carried out a study using Danielson Teaching Framework. Their results were at best moderate yet he concluded them as significant since standards-based approaches using rubrics are often noisy owing to various confounding factors that may operate at multiple levels. In their analyses of three sites, Milanowski, Kimball, and White¹² found that students of those teachers who were evaluated showed gains in their achievement. Similarly, in their multilevel analyses that explored associations between scores from teacher evaluation in standards-based approaches with student achievement, Kimball, et al.¹³ found positive relationships in all grades and the tested subjects. However, except in reading in fourth grade and in each

subject in grade 5, the coefficients were statistically insignificant across-the-board.

In 2011, Sartain et al. studied a pilot program rolled out in 2008 in Chicago named as Chicago's Excellence in Teaching to refurbish a viable and effective system of evaluating teachers. As seen in the studies discussed above, the Pilot program included developmental approaches to teacher evaluation wherein evaluators such as principals and personnel external to school engaged in observing teachers and providing them feedback through conferences. It was found that student achievement showed significant yields for those teachers who were observed and evaluated.¹⁴ According to teachers, the program effectively provided opportunities to them to reflect on their practice and thereby make continuous improvements in their pedagogical skills which ultimately reflected in improved student achievement.

The importance of classroom observations as important tools to improve teaching and learning conditions were explored in other studies as well. In their study, Tyler, Taylor, Kane, and Wooten¹⁵ found convincing evidence to date on the strong link between classroom observation measures and student achievement. They found that a shift, for example, of Teacher Evaluation System (TES) rating from "Basic" to "Proficient" and further on from "Proficient" to "Distinguished" (p. 259) related to a positive change of about 0.17 to 0.2 of a standard deviation. They delve deep into predictive ability and dynamics of classroom observations that explained achievement in mathematics and reading. In their analysis, they found that a teacher who managed and maintained a conducive environment witnessed an increase in student performance in mathematics by 0.25 standard deviation (SD) and in reading by 0.15 of an SD. They also found that teachers who adopted inquiry approaches in their teaching produced higher student achievement in mathematics compared to a teacher who only focused on delivering the specified content. However, this finding was not true in the case of reading. It could be argued based on these findings that teachers adjusted instructional objectives and their pedagogical practices such that they could address the objectives set forth in the developmental approaches to teacher evaluation in the TES. Furthermore, Taylor and Tyler¹⁶ also showed that a student who was taught by a teacher who was evaluated showed an increased score by a margin of 10% of an SD in mathematics compared to a similar student whose teacher was not evaluated. They, however, could not confirm the underlying dynamics that lied at the heart of such relations

between teacher evaluation and student achievement. Similarly, Rockoff and Speroni¹⁷ studied subjective and objective measures in the teacher evaluations that yielded significant relations with student achievement. Their target evaluation system included professional mentors who worked with the novice teachers.

Alongside the developmental purposes, high-stakes approaches and purposes are also constituent parts of teacher evaluation systems around the globe. In a long line of research exploring the efficacy of student test scores in high stakes evaluations of teachers, Bingham, Heywood, and White's¹⁸ study is one of the first. They studied performance of grade 5 students to assess if test scores were a good measure of teacher effectiveness in teacher evaluations. Based on their findings they were able to posit that student test scores are an effective measure in high stakes evaluations that can be used to rate teachers on the value that they added to student achievement. It was for such studies that subsequent exploration of teacher effects on student achievement gained a momentum.

In 1997, Wright, Horn, and Sanders studied how teacher factors affected performance of students to address the issue of non-random assignment of students. Wright, Horn, and Sanders¹⁹ proposed student achievement data as an appropriate measure to assess teachers. They further stated that an adequate standardized testing regime and careful research methods involving longitudinal structure of data could be appropriate to identify the good and bad teachers. Taking this line of research forward, Wößmann et al.²⁰ used one of the most robust survey data available to date, the PISA 2003 survey to reconfirm findings coming from earlier studies.²¹ They used multi-level modeling techniques wherein they found that external exit exams related positively to student test scores. Furthermore, Goldhaber and Hansen²² used administrative data in grades 4-5 and teachers to assess the use of test scores as viable criteria in high stakes teacher evaluations involving award of tenure. They found that such approaches in teacher evaluation associated significantly and positively with student achievement in these grades. They analyzed data on those teachers for whom information was available before and after the award of tenure. They found that teachers who could not secure a tenure had student achievement 11% of a standard deviation less than those students whose teachers were awarded tenure. Therefore, they concluded student test scores were viable measures to assess teacher effectiveness.

One of the latest pieces of evidence on the use of student test scores in high stakes teacher evaluation comes from OECD-2010 that found that schools using information from external but standards-based examinations registered a 16 points or higher increase in student achievement compared with schools where such a practice was not observed. OECD-2010 further states that internal standardized tests conducted by the school showed no observable link with student achievement. Student achievement has also been used in making high stakes decisions involving comparison of teacher performance across schools. Measures involving accountability of teams of teachers or schools with comparisons across districts or at national level resulted in positive associations with achievement.²³ Therefore, OECD-2010 proposed to schools that they can maneuver their mechanisms of accountability to explore viable teacher evaluation and accountability systems for improving learning outcomes for students.

Teacher evaluation, especially involving high-stakes, has not always been seen having positive relationships with or effects on student achievement. High-stakes teacher evaluation, for example public accountability, has been seen having mixed effects²⁴ or in some cases having negative consequences.²⁵ Last but not the least, using student achievement data for public accountability (e.g., by publishing student test scores in media), informing parents, allocating resources, or administrative tracking of student achievement data delivered mixed effects on student performance as shown in OECD report of 2010.

Are Value-Added Models the Way Forward?

When it comes to evaluating teachers based on measures of student outcome such as test scores in internal or external examinations, the biggest stumbling block that comes forth is the association of student performance with individual students. As a solution to this issue, Value Added Models have been considered as the way forward.²⁶ However, even though they are an advanced statistical form of the teacher evaluation mechanisms, there are criticisms on their efficiency in teasing out individual teacher effects on the students whom they taught²⁷.

Using data that they developed for a period of five years, Winter and Cowen²⁸ in their Valued Added Modeling of grade 4 and 5 students in Florida show that student performance was lower in classrooms that were taught by teachers who should have been dismissed under the applicable practice of teacher evaluation based

on student test scores. The authors further suggest that the VAMs can provide ample evidence on the quality of teachers. However, they posit, making decisions based on such evidence remains a human thing. Grossman, Loeb, Cohen, and Wyckoff²⁹ show that teachers who made to the top quartile in the VAMs displayed better instructional profiles than those who scored in lower quartile suggesting some efficiency of the VAMs in identifying quality teachers. In other words, VAMs that they used in their study could identify effective teachers who had a better control and command in instructional skills compared with the teachers who were less effective.

While Winter and Cowen and Grossman, Loeb, Cohen, and Wyckoff propose that VAMs can provide reliable evidence on teacher quality, evidence from elsewhere shows that the type of information that a VAM can give depends more on the statistical modeling and the nature of the data used. For example, Goldhaber, Goldschmidt, and Tseng³⁰ examined relationships that existed between different estimates of teacher effects in VAMs involving different assumptions on learning outcomes. Their analysis shows that the type of VAM used determines the nature of prediction with regard to teacher quality. For these reasons and others, while taking high-stakes decisions, especially involving decisions relating to service and tenure, information from VAMs should be used with utmost caution.

Conclusion

Current debates and efforts in raising teacher quality and hence student attainment has gained increased traction across the world. Educational scene in Pakistan is not immune to such discourses and developments. It is for this reason that Pakistan's National Educational Policy is increasingly focusing on evaluating teachers and rewarding them based on their performance. However, as can be seen in the above review of literature, the challenge of how best to evaluate teachers so as to improve their quality and learning outcomes for all students has transpired to a varied and dynamic landscape of policy analysis in global perspectives. The landscape becomes even more nuanced in the face of contrasting evidence on high stakes and developmental approaches to evaluate teachers.

Evidence on an effective and universal teacher evaluation system is at best mixed and evolving. However, there is consensus among scholars from a wide spectrum of the debate on policy around teacher evaluation that suggests that high-stakes teacher evaluation serves a very narrow objective of making teachers

accountable with a limited opportunity to help teachers improve their professional practice. Furthermore, measuring teacher effectiveness based only on student test scores offers many challenges. It has been pointed out by many that summative measures such as student test score and attaching high-stakes with these measures in teacher evaluations gives rise to what is commonly known as “teaching to the test” and “narrowing of curriculum” in addition to many other negative and unintended consequences involving detrimental effects on teacher collaboration and overall school culture³¹. Therefore, using summative evidence in the form of student test scores is suggested as only one measure in addition to many others in a given teacher evaluation system to assess teacher effectiveness³². For a country like Pakistan where “teaching to the test” is already a widespread phenomenon across almost all spectra of education, a lack of convincing evidence on sustainably raising student achievement through high-stakes evaluations will offer greater challenges. Value Added Models could be put in place as an alternative along with other measures, availability of relevant longitudinal data will be a challenge to begin with. Where relevant data is available, VAMs can be made a sub-part of a larger regime of teacher evaluation involving multiple measures of teacher performance and effectiveness. Other measures along with VAMs could be authentic classroom observations, rubrics, and artefacts of teacher work. As much as possible, teachers need to be made part of a process that aims at developing a viable teacher evaluation system for the country. Furthermore, for national policy around teacher evaluation to be effective and sustainable in the long run, the process must not become an end in itself. The process should serve to assist teachers develop their competencies and hence enhanced learning outcomes for all students.

Notes & References

- ¹ Marlène Isoré, “Teacher Evaluation: Current Practices in OECD Countries and a Literature Review”, OECD Education Working Papers, No. 23. *OECD Publishing (NJI)* (2009).
- ² Government of Pakistan, “National Education Policy 2009”. Retrieved, March, 13, 2017 from; <http://unesco.org.pk/education/teachereducation/files/National%20Education%20Policy.pdf>
- ³ Janet Looney, “Developing High-Quality Teachers: teacher evaluation for improvement”, *European Journal of Education* 46, no. 4 (2011): 440-455.
- ⁴ Donald L. Haefele, “Evaluating teachers: A call for change”, *Journal of Personnel Evaluation in Education* 7, no. 1 (1993): 21-31; See also: G. Latham and K. Wexley, *Increasing productivity through performance appraisal* (Monterey, CA: Brooks/Cole, 1982); Michael Scriven, “Summative teacher evaluation”, *Handbook of teacher evaluation* (1981), 244-271.
- ⁵ Eva Delvaux, Jan Vanhoof, Melissa Tuytens, Eva Vekeman, Geert Devos and Peter Van Petegem, “How may teacher evaluation have an impact on professional development? A multilevel analysis” *Teaching and Teacher Education* 36 (2013): 1-11; See also: Kerry Elliott, “Teacher Performance Appraisal: More about Performance or Development?” *Australian Journal of teacher education* 40, no. 9 (2015): 9; Organisation for Economic Co-operation and Development, Paulo Santiago, and Source OECD. *Teachers matter: Attracting, developing and retaining effective teachers* (2005); OECD, *Evaluating and rewarding the quality of teachers: International practices*, (2009). Available at: <http://dx.doi.org/10.1787/9789264034358-en>; OECD, *PISA 2009 results: What makes a school successful? Resources, policies and practices (Volume IV)*, 2010. Available at: <http://dx.doi.org/10.1787/9789264091559-en>; OECD, “Education policy outlook 2015: Making reforms happen”. Available at: https://www.oecd.org/edu/EPO%202015_Highlights.pdf.
- ⁶ OECD, “Education policy outlook 2015: Making reforms happen”. Available at: https://www.oecd.org/edu/EPO%202015_Highlights.pdf.
- ⁷ OECD, “Creating effective teaching and learning environments: First results from TALIS”. Available at: <http://www.oecd.org/edu/school/43023606.pdf>
- ⁸ Ibid.
- ⁹ Pam Grossman, Susanna Loeb, Julie Cohen, and James Wyckoff, “Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers’ value-added scores”, *American Journal of Education* 119, no. 3 (2013): 445-470; See also: Mark Paige, “Using VAM in high-stakes employment decisions”, *Phi Delta Kappan* 94, no. 3 (2012): 29-32.

¹⁰ Harold Wenglinsky, "The link between teacher classroom practices and student academic performance", *Education policy analysis archives* 10 (2002): 12.

¹¹ Elizabeth Holtzapple, "Criterion-related validity evidence for a standards-based teacher evaluation system" *Journal of Personnel Evaluation in Education* 17, no. 3 (2003): 207-219.

¹² Anthony Milanowski, "The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati", *Peabody Journal of Education* 79, no. 4 (2004): 33-53.

¹³ Steven M. Kimball, Brad White, Anthony T. Milanowski, and Geoffrey Borman, "Examining the relationship between teacher evaluation and student assessment results in Washoe County", *Peabody Journal of Education* 79, no. 4 (2004): 54-78.

¹⁴ Lauren Sartain, Sara Ray Stoelinga, and Eric R. Brown, "Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation", *Consortium on Chicago School Research* (2011).

¹⁵ John H. Tyler, Eric S. Taylor, Thomas J. Kane, and Amy L. Wooten, "Using student performance data to identify effective classroom practices", *The American Economic Review* 100, no. 2 (2010): 256-260.

¹⁶ Eric S. Taylor and John H. Tyler, "The effect of evaluation on performance: Evidence from longitudinal student achievement data of mid-career teachers", No. w16877, *National Bureau of Economic Research* (2011).

¹⁷ Jonah E. Rockoff, "The impact of individual teachers on student achievement: Evidence from panel data", *The American Economic Review* 94, no. 2 (2004): 247-252.

¹⁸ Richard D. Bingham, John S. Heywood, and Sammis B. White, "Evaluating schools and teachers based on student performance: Testing an alternative methodology", *Evaluation Review* 15, no. 2 (1991): 191-218.

¹⁹ William L. Sanders, S. Paul Wright, and Sandra P. Horn, "Teacher and classroom context effects on student achievement: Implications for teacher evaluation", *Journal of personnel evaluation in education* 11, no. 1 (1997): 57-67.

²⁰ Ludger Wöbmann, Elke Lüdemann, Gabriela Schütz, and Martin R. West, "School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003", OECD Education Working Papers (2007).

²¹ John H. Bishop, "The effect of national standards and curriculum-based exams on achievement", *The American Economic Review* 87, no. 2 (1997): 260-264.

²² Dan Goldhaber and Michael Hansen, "Using performance on the job to inform teacher tenure decisions", *The American Economic Review* 100, no. 2 (2010): 250-255.

-
- ²³ OECD, PISA 2009 results: What makes a school successful? Resources, policies and practices (Volume IV), 2010. Available at: <http://dx.doi.org/10.1787/9789264091559-en>
- ²⁴ Martin R. West and Paul E. Peterson, "The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments", *The Economic Journal* 116, no. 510 (2006).
- ²⁵ Andy Wiggins and Peter B. Tymms, "Dysfunctional effects of public performance indicator systems: a comparison between English and Scottish primary schools". In *Paper presented at the European Conference on Educational Research 20*, (2000), 23.
- ²⁶ OECD. *Teachers for the 21st Century: Using Evaluation to Improve Teaching*. OECD Publishing 2013. Available at: <http://www.oecd.org/site/eduistp13/TS2013%20Background%20Report.pdf>.
- ²⁷ Mark Paige, "Using VAM in high-stakes employment decisions", *Phi Delta Kappan* 94, no. 3 (2012): 29-32.
- ²⁸ Marcus A. Winters and Joshua M. Cowen, "Who would stay, who would be dismissed? An empirical consideration of value-added teacher retention policies", *Educational Researcher* 42, no. 6 (2013): 330-337.
- ²⁹ Pam Grossman, et. al., "Measure for measure...", op.cit.
- ³⁰ Dan D. Goldhaber, Pete Goldschmidt, and Fannie Tseng, "Teacher value-added at the high-school level: Different models, different answers?", *Educational Evaluation and Policy Analysis* 35, no. 2 (2013): 220-236.
- ³¹ Diane Ravitch, *The death and life of the great American school system: How testing and choice are undermining education* (Basic Books, 2016).
- ³² Eva L. Baker, Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, Richard Rothstein, Richard J. Shavelson, and Lorrie A. Shepard, "Problems with the Use of Student Test Scores to Evaluate Teachers", EPI Briefing Paper# 278, *Economic Policy Institute* (2010).