



REVIEW OF *TEACHER EVALUATION 2.0*

Reviewed By

H. Richard Milner, IV

Vanderbilt University

November 2010

Summary of Review

Teacher Evaluation 2.0, published by The New Teacher Project, offers a short critique of current teacher evaluation systems, structures, and practices. It describes the current system as broken and as falling short in terms of rigor and fairness. According to the report, evaluations are infrequent, unfocused, undifferentiated, unhelpful, and inconsequential. The report proposes six design standards for transforming teacher evaluation systems: an annual process; clear, rigorous expectations; multiple measures; multiple ratings; regular feedback; and significance. While the report outlines some logical recommendations, many of the ideas and recommendations are neither new nor innovative. However, this review identifies several assumptions, emphases, and approaches in the report that weaken its usefulness to policymakers.

Kevin Welner

Editor

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: 303-735-5290
Fax: 303-492-7090

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of 'Think Twice' Think Tank Reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

REVIEW OF *TEACHER EVALUATION 2.0*

H. Richard Milner IV, Vanderbilt University

I. Introduction

Attention to teacher quality and effectiveness¹ has been re-kindled by highly publicized actions, such as then-District of Columbia Chancellor Michelle Rhee’s firing of “ineffective” teachers last year,² the publication of test-based gain scores of individual teachers’ students in Los Angeles, and the movie *Waiting for Superman*.

Secretary of Education Arne Duncan’s emphasis on teacher evaluation and his claim that “Everyone agrees that teacher evaluations are broken,”³ underscore the prominence of the issue. Many observers, in fact, agree that teacher evaluation systems are generally perfunctory and in disrepair.⁴ However, the ongoing challenge has been to agree on actionable, working definitions specifying what it means to be a teacher of high quality.

Teacher Evaluation 2.0, a new report from The New Teacher Project (TNTTP),⁵ critiques current teacher evaluation systems and proposes six design “standards” that the report’s authors believe should be included in any teacher evaluation system. The report grounds the proposed design standards in a set of guiding principles and concludes with implementation considerations.

II. Findings and Conclusions of the Report

Teacher Evaluation 2.0 begins with the premise that “our schools have not done nearly enough to evaluate teachers accurately and use this information to improve educational quality” (p. 1). The guiding question of the report is: “*How can we address consistently ineffective teaching fairly but decisively?*” (p. 1). The report takes the stance that teacher evaluations should be used as tools to assist teachers in improving their practices but not as punitive tools that leave teachers intimidated. However, it also argues that teacher evaluations should be used in a high-stakes manner, and it includes the key proposal that teacher evaluations should be used to fire ineffective teachers and that promotion, tenure, and reappointments should rely on teacher evaluations.

The report has a logical organizational flow. It starts with an outline of five interrelated weaknesses with current teacher evaluation systems—namely, that they are:

- *Infrequent*: The report asserts that many teachers, especially experienced ones, are not evaluated annually.
- *Unfocused*: Evaluations tend to focus on “superficial” matters, such as teachers’ bulletin boards, and not on student academic progress.

- *Undifferentiated*: Teachers’ practices are either judged as “satisfactory/pass” or “unsatisfactory/fail,” and according to the report 99% of teachers in some districts receive the satisfactory rating.
- *Unhelpful*: The report claims that teachers say current evaluations are unhelpful in improving classroom practices and performances.
- *Inconsequential*: The report asserts that teacher evaluations are usually not used to make decisions related to teacher development, tenure, promotion, reappointment, or compensation.

Following this critique, the report sets forth guiding principles, which are quoted directly in full below:

- All children can master academically rigorous material, regardless of their socio-economic status.
- A teacher’s primary professional responsibility is to ensure that students learn.
- Teachers contribute to student learning in ways that can largely be observed and measured.
- Evaluation results should form the foundation of teacher development.
- Evaluations should play a major role in important employment decisions.
- No evaluation system can be perfect—in teaching or in any other profession.

The authors then propose six design standards that should be included in any teacher evaluation system:

Standard One: Annual Process

The basic premise of this standard is that all teachers should be evaluated by school leaders at least once a year. The idea is that teachers should be involved in ongoing evaluations that allow them to improve their practice.

Standard Two: Clear, Rigorous Expectations

This standard calls for overt links between teacher evaluations and student learning rather than evaluations on mundane teacher behaviors and routines. Moreover, this standard requires a high bar be set in the evaluation process.

Standard Three: Multiple Measures

A number of different data sources must be used to measure teacher effectiveness. The report argues that inclusion of measures derived from value-added models allows for a fuller picture of teacher effectiveness, and that relying on one or fewer data sources cannot capture what multiple measures can.

Standard Four: Multiple Ratings

This standard advances the idea that teacher rating scales should have four or five categories such as “highly effective,” “effective,” “needs improvement,” or “ineffective” (p. 7), rather than a two-point rating system (satisfactory/unsatisfactory). Multiple categories mean there are “at least two levels at or above expectations and two levels below expectations.”

Standard Five: Regular Feedback

This standard requires substantive, consistent, and frequent feedback that is used in teacher professional development. The report warns that regular feedback can be disadvantageous and ineffective if “teachers view development conversations as chores instead of opportunities to talk openly and constructively about instruction” (p. 8).

Standard Six: Significance

The sixth standard requires that the evaluation system should not be perfunctory; it should have real consequences for teachers. According to the report, results of teacher evaluations (both positive and negative) should have a bearing on decisions related to “teacher tenure, compensation, development, hiring, promotion and dismissal” (p. 8).

The final section of the report focuses on implementation of the six proposed standards. It reminds readers that the success of any evaluation system depends on the fidelity of implementation. Success depends on a school district and school’s ability to provide three things: (1) ongoing training and support for those involved, (2) more resources and personnel, and (3) “specific metrics to track” (p. 9) the functioning of the system. The report concludes with questions and considerations intended to help guide district implementation.

III. Rationale for Findings and Conclusions

The report seems to operate largely from conventional wisdom in its critique of current evaluation systems, its guiding principles, and its proposed design standards. There was no systemic review of current research and thinking on teacher evaluation, although references to popular and contemporary readings are dispersed throughout. Many of the ideas expressed are not new, and the limited citations are seldom complete. However, the District of Columbia’s IMPACT program and TNTP’s own 2009 report: *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, appear to be guiding documents of analyses for this report.

IV. Review of the Report’s Methods

While “guiding principles” are used to frame and shape the proposed standards, it is not stated how these principles were derived or how the proposed standards emerge from or connect to the principles. The report does offer some “real-life” examples to help illuminate and explicate the

six proposed standards. However, it is unclear how the real examples were selected, whether they are representative of ordinary practice in the U.S. or in particular school districts, and how they lead to the generalizations in this report. While the designs are logical and appear to be common sense, the same can be said of many theories that later failed to achieve their goals when put into practice. No evidence is presented in this report to suggest that these examples have proven effective or that they exemplify the desired traits.

V. Review of the Validity of the Findings

The main sections of the report (weaknesses of evaluation systems, guiding principles, standards, and implementation concerns) are broad and vaguely stated. They would be generally accepted as lofty truisms by many readers. However, in the specifics, a number of critical concerns come to light:

- Particularly in the opening sections, reforming teacher evaluation systems is implicitly and explicitly presented as a panacea for resolving far greater educational problems. While teacher evaluations can be a critical element in teacher development, learning, and improvement, they cannot single-handedly resolve the multifaceted, deeply nuanced, and complex troubles faced by school districts (in particular, urban districts facing great disparities between resources and student needs).
- The report recommends a heavy emphasis on standardized tests and measures. Looking back on more than a decade of this approach, such an emphasis has not proven to be a valid and generalizable practice. There are countless outside factors that have an impact on student test scores which are not given proper consideration in the report.⁶ The report recommends value-added measures, but this approach has been repeatedly shown to be insufficient to overcome validity concerns.⁷
- While student test score approaches are combined in the report's approach with administrator classroom observations, there is minimal consideration of quality indicators (beyond tests and observations) that capture key information about a teacher's ability. Portfolio evaluations and other artifacts that provide evidence that teachers are effective or improving and that students are learning receive very little attention in the report.
- Although teachers can and do make an important difference in student progress, the report fails to put this teacher effect in a meaningful context. If readers do not understand the role of additional factors such as parental involvement, student motivation, administrative support, and resources, they may easily accept results of teacher evaluations as the independent reason for student success or failure.⁸
- An unaddressed contradiction is embedded in two conflicting goals of the proposed evaluation system: (1) the use of evaluations for teacher personal/professional development, and (2) the use of teacher evaluation for administrative personnel decisions. Although a sound system including both elements may be workable, a district will face real tensions and challenges – obstacles that are never mentioned, acknowledged, or explored in this report.

- In urban and highly diverse settings, key attributes of effective teaching may not be clearly demonstrated through tests or may be obscured by other circumstances connected to structural and systemic inequities.
- Administrators and other evaluators often spend disproportionate amounts of time with new, inexperienced, poorly trained, and sometimes ineffective novice teachers. In tight fiscal circumstances, having sufficient staff to implement the ideal evaluation structure may not be realistic.⁹ While all teachers should be evaluated yearly and perhaps multiple times each year, the report does not consider the realities of schools (particularly under-resourced¹⁰ and understaffed schools) that may prevent annual reviews of teachers.
- The report’s suggestion that teacher evaluations should be completed by “school leaders” (p. 4) assumes that principals possess the time, knowledge, skills, attitudes, dispositions, and expertise necessary for teacher development and improvement. In some cases, principals do not have the subject-matter knowledge necessary to make curricular or instructional evaluations or recommendations to teachers. While principal evaluations may be the best option in some contexts, more flexible recommendations may be called for given the great diversity of needs, demands, and contexts throughout the U.S. Even when principals are excellent “leaders,” they sometimes are far better at administrative and school leadership than curriculum and instructional practices. Without proper provision of qualified evaluators, no evaluation system will prove successful.

VI. Usefulness of the Report for Guidance of Policy and Practice

Teacher Evaluation 2.0 provides some commonsensical and conventional suggestions and recommendations for changing teacher evaluation systems. Having systems that are logical, clear, context-responsive and universally implemented represent sound practices. The report’s ideas are therefore neither irrelevant nor trivial. But many of them are delivered at a high level of abstraction and do not provide fresh, cutting-edge, or innovative solutions or recommendations for recognized problems in teacher evaluation. And the report errs in its excessive attribution of benefits that may result from transforming the shortcomings of teacher evaluation systems. While important, teacher evaluation is only one piece of the puzzle. More troubling is the uncritical endorsement of value-added modeling for teacher evaluations – an approach which if used as a major part of a teacher evaluation system would simply outpace the usefulness of its information, possibly harming both teachers and students.

In short, this report generally points policy in the right direction, although it does not tell us anything that is not already suggested or asserted in the teacher evaluation discourse. The recommendations rest on appeals to common sense rather than on actual empirical evidence. Before implementing the recommendations, policymakers will want to test the conjectures dispersed throughout the report. With these caveats, the report has some recommendations that may prove useful for policymakers and could be implemented, but only with deliberate caution.

Notes and References

- 1 Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher*, 31(9), 13-25.
- 2 The bottom 5% of teachers (241 in all) were found to be ineffective and were consequently fired in the D.C. public schools.
- 3 Duncan's comment is quoted, without citation, at The New Teacher Project (2010, October). *Issue analysis: Building a Thriving Teacher Workforce*. (See box in lower right-hand corner.) Retrieved November 5, 2010, from <http://www.tntp.org/index.php/publications/issue-analysis/view/teacher-evaluation-2.0/>
- 4 For example, see two recent articles in the practitioner journal *Educational Leadership*: Donaldson, M.L. (2010). No more valentines. *Educational Leadership*, 67(8), 54-58. Toch, T. (2008). Fixing teacher evaluation. *Educational Leadership*, 66(2), 32-37.
- 5 The New Teacher Project (2010). *Teacher Evaluation 2.0*. New York: Author. Retrieved November 5, 2010, from <http://www.tntp.org/files/Teacher-Evaluation-Oct10F.pdf>
- 6 Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3-12.
- 7 See e.g., Baker, E. L. et. al. (2010, August 29) *Problems with the use of student test scores to evaluate teachers*. EPI briefing paper #278. Washington, DC: Economic Policy Institute.
- 8 Milner, H. R. (2010). *Start where you are, but don't stay there: Understanding diversity, opportunity gaps, and teaching in today's classrooms*. Cambridge, MA: Harvard Education Press.
- 9 Milner, H. R. (2010). *Start where you are, but don't stay there: Understanding diversity, opportunity gaps, and teaching in today's classrooms*. Cambridge, MA: Harvard Education Press.
- 10 Anyon, J. (1980). Social class and the hidden curriculum of work. *Journal of Education*, 162(1), 366-391.

DOCUMENT REVIEWED:	Teacher Evaluation 2.0
AUTHOR:	The New Teacher Project
PUBLISHER/THINK TANK:	The New Teacher Project
DOCUMENT RELEASE DATE:	October 2010
REVIEW DATE:	November 16, 2010
REVIEWER(S):	H. Richard Milner IV, Vanderbilt University
E-MAIL ADDRESS:	rich.milner@vanderbilt.edu
PHONE NUMBER:	(615) 480-7474 or (615) 322-4503
SUGGESTED CITATION:	

Milner, H.R. IV (2010). *Review of "Teacher Evaluation 2.0."* Boulder : National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-teach-eval-TNTP>