



School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
Telephone: 802-383-0058

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

RESEARCH-BASED OPTIONS FOR EDUCATION POLICYMAKING

Teacher Evaluation

*William Mathis, University of Colorado Boulder
September 2012*

Teachers are important, and policies mandating high-stakes evaluations of teachers are at the forefront of popular school reforms. Today's dominant approach labels teachers as effective or ineffective based in large part on a statistical analysis of students' test-score performance. Teachers judged effective are rewarded, and those found ineffective are sanctioned.

While such *summative* evaluations can be useful, lawmakers should be wary of approaches based in large part on test scores: the error in the measurements is large—which results in many teachers being incorrectly labeled as effective or ineffective;¹ relevant test scores are not available for the students taught by most teachers, given that only certain grade levels and subject areas are tested; and the incentives created by high-stakes use of test scores drive undesirable teaching practices such as curriculum narrowing and teaching to the test.²

Summative initiatives should also be balanced with *formative* approaches, which identify strengths and weaknesses of teachers and directly focus on developing and improving their

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

teaching. Measures that de-emphasize test scores are more labor intensive but have far greater potential to enrich instruction and improve education.

Teacher quality is among the most important *within-school* factors affecting student achievement. However, research also suggests that teacher differences account for no more than about 15% of differences in students' test score outcomes.³ Other school factors such as class size reduction⁴ and adequate, focused funding⁵ are also research-based ways to improve education. Further, *non-school factors*, which are generally associated with parental education and wealth, are far more important determinants of students' test scores.⁶

Care must be taken in selecting or designing a balanced evaluation system. Given the extensive range of activities, skills, and knowledge involved in teachers' daily work, the system's goals must be clear, explicit and reflect practitioner involvement.⁷ Effective teacher evaluation also requires an investment in sufficient numbers of qualified evaluators. Otherwise, the system will likely be irregular, uneven and ineffective.⁸

Many established evaluation systems are available, and some have a strong research base. Among the more widely known approaches are Charlotte Danielson's Framework for Teaching⁹ and the Peer Assistance and Review (PAR)¹⁰ approach. Connecticut's Beginning Educator's Support and Training (BEST) system along with the National Board for Professional Teaching Standard's system for advanced teachers are also recognized as promising systems for promoting both student learning and professional improvement.¹¹ Properly preparing teachers is also receiving renewed attention, and Stanford's *edTPA* consortium of 24 states is developing comprehensive assessments of prospective teachers.¹²

Any single measure of teaching or teachers will emphasize one important element at the expense of others.¹³ Accordingly, all teacher evaluation systems should employ a diverse set of measures to capture the complex nature of the art and science of teaching.¹⁴ In fact, the wisest choice may be to have two or more *separate* measurement systems within a district, allowing for the possibility of different results—which in turn would provide a check and a caution against relying on only one measurement system.

Key Research Points and Advice for Policymakers

- If the objective is improving educational practice, formative evaluations that guide a teacher's improvement provide greater benefits than summative evaluations.¹⁵
- If the objective is to improve educational performance, outside-school factors must also be addressed. Teacher evaluation cannot replace or compensate for these much stronger determinants of student learning.¹⁶ The importance of these outside-school factors should also caution against policies that simplistically attribute student test scores to teachers.
- The results produced by value-added (test-score growth) models alone are highly unstable. They vary from year to year, from classroom to classroom, and from one

test to another.¹⁷ Substantial reliance on these models can lead to practical, ethical and legal problems.

- High-stakes evaluations based in substantial part on students' test scores narrow the curriculum by diminishing or pushing out non-tested subjects, knowledge, and skills.¹⁸
- Teacher evaluation systems necessarily involve trade-offs, and specific design choices are controversial, so it is important to involve all key stakeholders in system design or selection.¹⁹
- To be successful, schools must invest in their teacher evaluation systems. An adequate number of highly trained evaluators must be available.²⁰
- Given the wide variety of teacher roles and the many factors that influence learning that are outside the control of the teacher, a wide variety of measures of teacher effectiveness is also indicated.²¹ By diversifying, the weakness of any single measure is offset by the strengths of another.²²
- High-quality research on existing evaluative programs and tools should inform the design of teacher evaluation systems.²³ States and districts should investigate balanced models such as PAR and the Danielson Framework, closely examine the evidence concerning strengths and weaknesses of each model, and never attach high-stakes consequences to teachers which the evidence cannot validly support.

Notes and References

1 Briggs, D. & Domingue, B. (2011). *Due Diligence and the Evaluation of Teachers*. Boulder, CO: National Education Policy Center. Retrieved September 16, 2012, from <http://nepc.colorado.edu/publication/due-diligence>.

Durso, C. S. (2012). *An Analysis of the Use and Validity of Test-Based Teacher Evaluations Reported by the Los Angeles Times: 2011*. Boulder, CO: National Education Policy Center. Retrieved September 16, 2012, from <http://nepc.colorado.edu/publication/analysis-la-times-2011>.

2 Rothstein, R., Ladd, H. F., Ravitch, D., Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Linn, R. L., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the Use of Student Test Scores to Evaluate Teachers*. Economic Policy Institute. Retrieved September 16, 2012, from <http://www.epi.org/publication/bp278/>.

3 Hanushek, E. A., Kai, J. F., & Rivken, S. J. (August 1998). *Teachers Schools and Academic Achievement*. NBER working paper 6691. Cambridge MA. Retrieved September 16, 2012, from http://www.cgp.upenn.edu/pdf/Hanushek_NBER.PDF.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (Fall 2004). How Large are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26 (3), 237-257. Retrieved September 16, 2012, from

<http://steinhardt.nyu.edu/scmsAdmin/uploads/002/834/127%20-%20Nye%20B%20%20Hedges%20L%20%20V%20%20Konstant>.

Rowan, B., Correnti, R., & Miller, R. J. (November 2002). *What Large-scale, Survey Research Tells Us about Teacher effects on Student Achievement: Insights from the Prospects Study of Elementary Schools*. CRPE Research Report RR-051. Retrieved September 16, 2012, from <http://cw.marianuniversity.edu/mreardon/755/document%20repository/Teacher%20Effects%20on%20Student%20Achievement.pdf>

4 Class size: Project STAR. Retrieved September 16, 2012, from <http://www.aypf.org/publications/rmaa/pdfs/ClassSizeSTAR.pdf>.

Class size: Project SAGE. Retrieved September 16, 2012, from <http://www.aypf.org/publications/rmaa/pdfs/ClassSizeSAGE.pdf>.

These stand-alone documents are extracted from

James, D. W., Jurich, S., & Estes, S. (2001). *Raising minority academic achievement: A compendium of education programs and practices*. Washington, DC: American Youth Policy Forum.

5 Baker, B. (2012). *Revisiting the Age-Old Question: Does Money Matter in Education?* Washington: The Albert Shanker Institute. Retrieved September 16, 2012, from http://www.shankerinstitute.org/images/doesmoneymatter_final.pdf

6 Rothstein *et al.*, 2010 (see note 2).

7 Hinchey, P. H., (2010). *Getting Teacher Assessment Right: What Policymakers Can Learn from Research*. Boulder, CO: National Education Policy Center. Retrieved September 16, 2012, from http://nepc.colorado.edu/files/PB-TEval-Hinchey_o.pdf.

8 Hinchey, 2010 (see note 7).

Tucker, P. (1997). Lake Wobegon: Where all teachers are competent (Or have we come to terms with the problem of incompetent teachers?). *Journal of Personnel Evaluation in Education*. 11(2), 103-26.

Ward, M. E. (1995). Teacher dismissal: The impact of tenure, administrator competence, and other factors. *School Administrator*, 52, 16-19.

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The Widget Effect: Our National Failure to Acknowledge and Act on Teacher Differences*. Brooklyn, NY: The New Teacher Project. Retrieved September 16, 2012 from <http://widgeteffect.org/>.

9 See Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

10 Papay, J. P., & Johnson, S. M. (2012). Is PAR a Good Investment? Understanding the Costs and Benefits of Teacher Peer Assistance and Review Programs. *Educational Policy*, 26(5) 696-729.

11 Hinchey, 2010 (see note 7).

12 Darling-Hammond, L. (2012, August 15). A New Way to Evaluate Teachers – by Teachers. The Answer Sheet. *Washingtonpost.com* Retrieved September 16, 2012, from http://www.washingtonpost.com/blogs/answer-sheet/post/a-new-way-to-evaluate-teachers--by-teachers/2012/08/14/5844dc86-e677-11e1-936a-b801f1abab19_blog.html#pagebreak.

13 Guarini, C. & Stacy, B. (March 2012). *Review of Gathering Feedback for Teachers*. Boulder, CO: National Education Policy Center. Retrieved September 16, 2012, from <http://nepc.colorado.edu/thinktank/review-gathering-feedback>

Hinchey, 2010 (see note 7).

14 Steele, J. L., Hamilton, L. S., & Stecher, B. M. (2010). *Incorporating Student Performance Measures into Teacher Evaluation Systems*. Palo Alto, CA: RAND. Retrieved September 16, 2012, from http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR917.pdf

15 Hinchey, 2010 (see note 7).

16 Rothstein *et al.*, 2010 (see note 2).

17 Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating Teacher Education. *Phi Delta Kappan*, 93(6), pp. 8-15. Retrieved September 16, 2012, from <http://www.kappanmagazine.org/content/93/6/8.full.pdf>

18 Burris, C. & Welner, K. (2012, June 29). *5 reasons parents should oppose evaluating teachers on test scores*. *Washington Post, Answer Sheet*. Retrieved September 16, 2012, from http://www.washingtonpost.com/blogs/answer-sheet/post/5-reasons-parents-should-oppose-evaluating-teachers-on-test-scores/2011/06/05/AGTppaJH_blog.html#pagebreak.

19 Hinchey, 2010 (see note 7).

20 Hinchey, 2010 (see note 7).

21 Briggs & Domingue, 2011 (see note 1).

Durso, 2012 (see note 1).

Rothstein *et al.*, 2010 (see note 2).

22 Hinchey, 2010 (see note 7).

23 Hinchey, 2010 (see note 7).

*This is a section of **Research-Based Options for Education Policymaking**, a multipart brief that takes up a number of important policy issues and identifies policies supported by research. Each section focuses on a different issue, and its recommendations to policymakers are based on the latest scholarship. **Research-Based Options for Education Policymaking** is published by The National Education Policy Center, housed at the University Of Colorado Boulder, and is made possible in part by funding from the Great Lakes Center for Education Research and Practice.*

*The mission of the **National Education Policy Center** is to produce and disseminate high-quality, peer-reviewed research to inform education policy discussions. We are guided by the belief that the democratic governance of public education is strengthened when policies are based on sound evidence. For more information on NEPC, please visit <http://nepc.colorado.edu/>.*