Getting Teacher Evaluation Right:
A Brief for Policymakers

American Education Research Association and
National Academy of Education

There is a widespread consensus among practitioners, researchers, and policy makers that current teacher evaluation systems in most school districts do little to help teachers improve or to support personnel decision making. For this reason, new approaches to teacher evaluation are being developed and tested.

There is also a growing consensus that evidence of teachers' contributions to student learning should be a component of teacher evaluation systems, along with evidence about the quality of teachers' practice. "Value Added Models" (VAMs) for looking at gains in student test scores from one year to the next are promoted as tools to accomplish this goal. Policy makers can benefit from research about what these models can and cannot do, as well as from research about the effects of other approaches to teacher evaluation. This brief addresses both of these important concerns.

Research on Value-Added Models of Teacher "Effectiveness"

Researchers have developed value-added methods for looking at gains in student achievement by using statistical methods that allow them to measure changes in student scores over time, while taking into account student characteristics and other factors often found to influence achievement. In large-scale studies, these methods have proved valuable for looking at a range of factors affecting achievement and measuring the effects of programs or interventions.

When applied to individual teacher evaluation, the use of VAM assumes that measured student achievement gain, linked to a specific teacher, reflect that teacher's "effectiveness." Drawing this conclusion, however, assumes that student learning is measured well by a given test, is influenced by the teacher alone, and is independent from the growth of classmates and other aspects of the classroom context.

However, research reveals that a student's achievement and measured gains are influenced by much more than any individual teacher. Others factors include:

- School factors such as class sizes, curriculum materials, instructional time, availability of specialists and tutors, and resources for learning (books, computers, science labs, and more)
- Home and community supports or challenges
- Individual student needs and abilities, health, and attendance
- Peer culture and achievement
- Prior teachers and schooling, as well as other current teachers
- Differential summer learning loss, which especially affects low-income children
- The specific tests used, which emphasize some kinds of learning and not others, and which rarely measure achievement that is well above or below grade level.

Most of these factors are not actually measured in value-added models, and the teacher's effort and skill, while important, constitute a relatively small part of this complex equation. As a consequence, researchers have documented a number of problems with VAM as accurate measures of teachers' effectiveness.

1. Value-added models of teacher effectiveness are highly unstable.

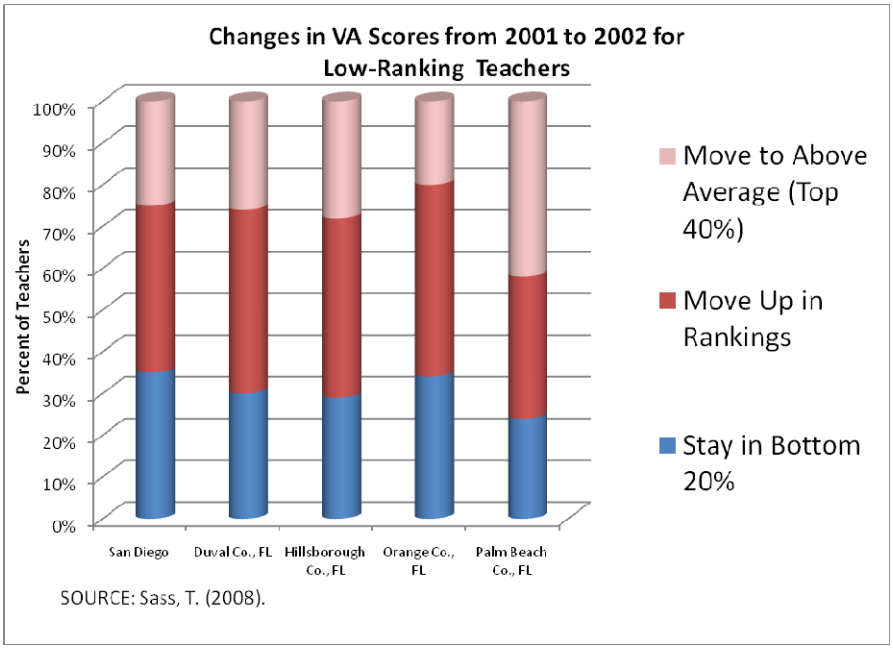
Researchers have found that teachers' effectiveness ratings differ substantially from class to class and from year to year, as well as from one statistical model to the next, as Table 1 shows.[1]

Table 1: Percent of Teachers Whose Effectiveness Rankings Change

	By 1 or more Deciles	By 2 or more Deciles	By 3 or more Deciles
Across models [a]	56-80%	12-33%	0-14%
Across courses [b]	85-100%	54-92%	39-54%
Across years [b]	74-93%	45-63%	19-41%

Note: [a] Depending on pair of models compared. [b] Depending on the model used.
Source: Newton, Darling-Hammond, Haertel, and Thomas (2010).

A study examining data from five separate school districts found, for example, that of teachers who scored in the bottom 20% of rankings in one year, only 20-30% had similar ratings the next year, while 25-45% of these teachers moved to the top part of the distribution, scoring well above average. (See Figure 1.) The same was true for those who scored at the top of the distribution in one year: A small minority stayed in the same rating band the following year, while most scores moved to other parts of the distribution.



Teachers' measured effectiveness varies significantly when different statistical methods are used.[2] For example, when researchers used a different model to recalculate the value-added scores for teachers that were published in the *Los Angeles Times* in 2011, they found that 40%-55% of them would get noticeably different scores using a VAM that accounted for student assignments in a different way.[3]

Teachers' value-added scores also differ significantly when different tests are used, even when these are within the same content area.[4] For example:

- In a study using two tests measuring basic skills and higher order skills, 20%-30% of teachers who ranked in the top quartile in terms of their impacts on state tests ranked in the bottom half of impacts on more conceptually demanding tests (and vice versa).[5]

- Teachers' estimated effectiveness is very different for "Procedures" and "Problem Solving" subscales of the same math test.[6]

- Teacher effects on high-stakes tests are not highly related to their effects on low stakes tests, and dissipate more quickly.[7]

This raises concerns both about measurement error and, when teacher evaluation results are tied to student test scores, about the effects of emphasizing "teaching to the test" at the expense of other kinds of learning, especially given the narrowness of most tests currently used in the United States.

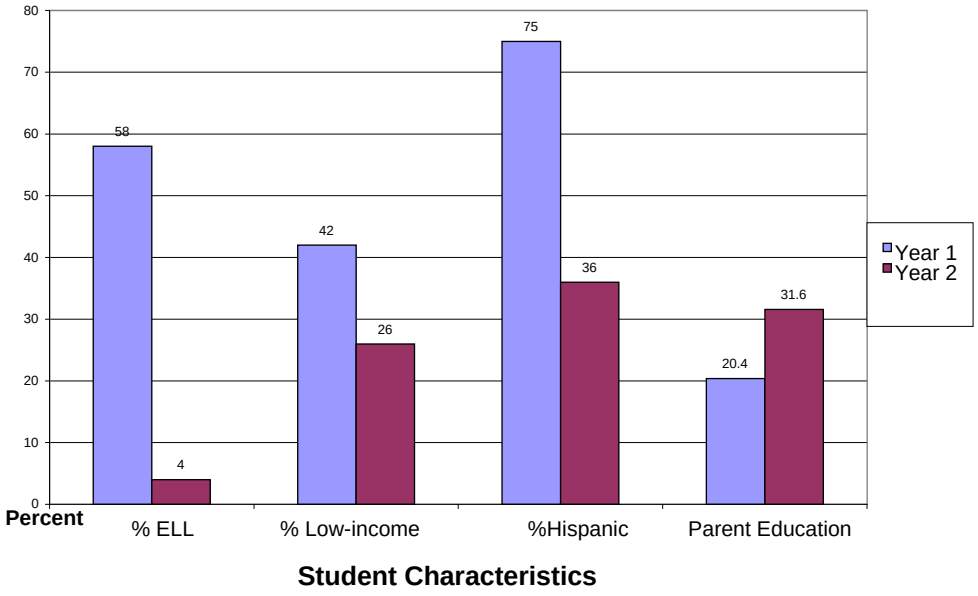
2. Teachers' value-added ratings are significantly affected by differences in the students who are assigned to them

VAMs require that students be assigned to teachers randomly. However, students are not randomly assigned to teachers – and statistical models cannot fully adjust for the fact that some teachers will have a disproportionate number of students who have greater challenges (students with poor attendance, who are homeless, who have severe problems at home, etc.) and those whose scores on traditional tests may not accurately reflect their learning (e.g., those who have special education needs or who are new English language learners). These factors can create both misestimates of teachers' effectiveness and disincentives for teachers to want to teach the students who have the greatest needs.

Even when the model includes controls for prior achievement and student demographic variables, teachers are advantaged or disadvantaged based on the students they teach. Several studies have shown this by conducting tests which look at a teacher's "effects" on their students in grade levels *before* or *after* the grade level in which he or she teaches them. Logically, for example, 5th grade teachers can't influence their teachers' 3rd grade test scores. So a VAM that identifies teachers' true effects should show *no* effect of 5th grade teachers on their students' 3rd grade test scores two years earlier. But studies that have looked at this have shown large "effects" – which suggest that students have at least as much bearing on the value-added measure as the teachers who actually teach them in a given year.[8]

One study that found considerable instability in teachers' value-added scores from class to class and year to year examined changes in student characteristics associated with the changes in teacher ratings. After controlling for prior test scores of students *and* student characteristics, the study still found significant correlations between teachers' ratings and their students' race/ethnicity, income, language background, and parent education. Figure 2 illustrates this finding for an experienced English teacher in the study whose rating went from the very lowest category in one year to the very highest category the next year (a jump from the 1st to the 10th decile). In the second year, this teacher had many fewer English learners, Hispanic students, and low-income students, and more students with well-educated parents than in the first year.

Figure 2 - Student Characteristics in Years 1 and 2 for a Teacher Whose Ranking Changed from the 1st to the 10th Decile



This variability raises concerns that use of such ratings for evaluating teachers could create disincentives for teachers to serve high-need students. This could inadvertently reinforce current inequalities, as teachers with options would be well-advised to avoid classrooms or schools serving such students, or to seek to prevent such students from being placed in their classes.

3. Value-added ratings cannot disentangle the many influences on student progress

It is impossible to fully separate out the influences of students' other teachers, as well as school conditions, on their reported learning. No single teacher accounts for all of a student's learning. Prior teachers have lasting effects, for good or ill, on students' later learning, and current teachers also interact to produce students' knowledge and skills. For example, the essay writing a student learns through his history teacher may be credited to his English teacher, even

if she assigns no writing; the math he learns in his physics class may be credited to his math teacher. Specific skills and topics taught in one year may not be tested until later, if at all. Some students receive tutoring, as well as help from well-educated parents. A teacher who works in a well-resourced school with specialist supports may appear to be more effective than one whose students don't receive these supports. As Henry Braun, then at ETS, noted,

> It is always possible to produce estimates of what the model designates as teacher effects. These estimates, however, capture the contributions of a number of factors, those due to teachers being only one of them. So treating estimated teacher effects as accurate indicators of teacher effectiveness is problematic.[9]

Initial research on the use of value-added methods to dismiss some teachers and award bonuses to others shows that value-added ratings often do not agree with the ratings teachers receive from skilled observers, and are influenced by all of the factors described above.

For example, among several teachers dismissed in Houston as a result of their value-added test scores, one teacher, a ten-year veteran, had been "Teacher of the Year" and was rated each year as "exceeding expectations" by her supervisor. She showed positive VA scores on 8 of 16 tests over four years (50% of the total observations), depending in part on the grade level she was asked to teach. Another teacher, also consistently rated as "exceeding expectations" or "proficient" by her supervisor, and also receiving positive VA scores about 50% of the time, had a noticeable drop in her value-added ratings when she was assigned to teach a large number of English Language Learners who were transitioned into her classroom.[10] Overall, the study found that, in this system,

- Teachers cannot identify a relationship between their instructional practices and their ratings on value-added, which appear unpredictable.
- Ratings change considerably when teachers change grade levels, often from "ineffective" to "effective" and vice versa.
- Teachers teaching in grades in which English Language Learners (ELLs) are transitioned into mainstreamed classrooms are the least likely to show "added value."
- Teachers teaching larger numbers of special education students in mainstreamed classrooms are also found to have lower "value-added" scores, on average.
- Teachers teaching gifted students add little value because their students are already near the top of the test score distribution.
- Teachers report seeking to boost their scores by avoiding certain subjects and types of students, and by seeking assignments to teach particular subjects / grades.
- Teachers also report being confused and demoralized by the system.

Professional Consensus about the Use of Value-Added Methods in Teacher Evaluation

For all of these reasons, most researchers have concluded that VAM is not appropriate as a primary measure for evaluating individual teachers. A major report by the RAND Corporation concluded that:

> The research base is currently insufficient to support the use of VAM for high-stakes decisions about individual teachers or schools.[11]

Similarly, Henry Braun of the Educational Testing Service concluded in his review of research:

> VAM results should not serve as the sole or principal basis for making consequential decisions about teachers. There are many pitfalls to making causal attributions of teacher effectiveness on the basis of the kinds of data available from typical school districts. We still lack sufficient understanding of how seriously the different technical problems threaten the validity of such interpretations.[12]

Finally, the National Research Council's Board on Testing and Assessment concluded that:

> VAM estimates of teacher effectiveness that are based on data for a single class of students should not used to make operational decisions because such estimates are far too unstable to be considered fair or reliable.[13]

Other Approaches to Teacher Evaluation

While VAMs based on student test scores are problematic for making evaluation decisions for individual teachers, they are useful for looking at groups of teachers for research purposes – for example, to examine how specific teaching practices or measures of teaching influence the learning of large numbers of students. The larger scale of these studies reduces error, and their frequent use of a wider range of outcome measures allows more understanding of the range of effects of particular strategies or interventions.

These kinds of analyses provide other insights for teacher evaluation, since there is a large body of evidence over many decades concerning how specific teaching practices influence student learning gains. For example, there is considerable evidence that effective teachers:

- Understand subject matter deeply and flexibly;
- Connect what is to be learned to students' prior knowledge and experience;
- Create effective scaffolds and supports for learning;
- Use instructional strategies that help students draw connections, apply what they are learning, practice new skills, and monitor their own learning;
- Assess student learning continuously and adapt teaching to student needs;
- Provide clear standards, constant feedback, and opportunities for revising work; and
- Develop and effectively manage a collaborative classroom in which all students have membership.[14]

These aspects of effective teaching, supported by research, have been incorporated into professional standards for teaching that offer some useful approaches to teacher evaluation.

Using Professional Standards for Teacher Evaluation

Professional standards defining accomplished teaching were first developed by the National Board for Professional Teaching Standards (NBPTS) to guide assessments for veteran teachers. Subsequently, a group of states working together under the auspices of the Council for Chief State School Officers created the Interstate New Teacher Assessment and Support Consortium (INTASC), which translated these into standards for beginning teachers, adopted by over 40 states for initial teacher licensing. A recent revision of the INTASC teaching standards has resulted in their alignment with the Common Core Standards in order to reflect the kind of teacher knowledge, skills, and understandings needed to enact the standards.

These standards have become the basis for assessments of teaching that produce ratings which are much more stable than value-added measures. At the same time, they incorporate classroom evidence of student learning and they have recently been shown in larger-scale studies to predict teachers' value-added effectiveness, so they help ground evaluation in student learning in more stable ways. Typically the performance assessments ask teachers to document their plans and teaching for a unit of instruction linked to the state standards, adapt them for special education students and English language learners, videotape and critique lessons, and collect and evaluate evidence of student learning.

A number of studies have found that the National Board Certification assessment process identifies teachers who are more effective in raising student achievement than other teachers.[15] Equally important, studies have found that teachers' participation in the National Board process stimulates improvements in their practice.[16] Similar performance assessments, used with beginning teachers in Connecticut and California, have been found to predict their students' achievement gains on state tests.[17] The Performance Assessment for California Teachers (PACT) has also been found to improve beginning teachers' competence and to stimulate improvements in the teacher education programs that use it as a measure.[18]

Professional standards have also been translated into teacher evaluation instruments in use at the local level. In a study of three districts using standards-based evaluation systems, researchers found significant relationships between teachers' ratings and their students' gain scores on standardized tests, and evidence that teachers' practice improved as they were given frequent feedback in relation to the standards.[19] In the schools and districts studied, assessments of teachers were based on well-articulated standards of practice evaluated through evidence including observations of teaching along with teacher pre- and post-observation interviews and, sometimes, artifacts such as lesson plans, assignments, and samples of student work.

Finding Additional Measures Related to Teacher Effectiveness

The Gates Foundation has launched a major initiative to find additional tools that are validated against student achievement gains and that can be used in teacher evaluation at the local level. The *Measure of Effective Teaching (MET) Project* has developed a number of tools, some of them based on the standards-based assessments described above, and others taking a new tack. Among these are observations or videotapes of teachers, supplemented with other artifacts of practice (lesson plans, assignments, etc.), that can be scored according to a set of

standards which reflect practices associated with effective teaching. Also included are tools like student surveys about teaching practice, which have been found, in an initial study, to be significantly related to student achievement gains.[20]

Countries like Singapore include a major emphasis on teacher collaboration in their evaluation systems. This kind of measure is supported by studies which have found that stronger value-added gains for students are supported by teachers who work together as teams[21] and by higher levels of teacher collaboration for school improvement. [22]

Some systems ask teachers to assemble evidence of student learning as part of the overall judgment of effectiveness. Such evidence is drawn from classroom and school-level assessments and documentation, including pre- and post-test measures of student learning in specific courses or curriculum areas, and evidence of student accomplishments in relation to teaching activities. A study of Arizona's career ladder program, which requires the use of various methods of student assessment to complement evaluations of teachers' practice, found that, over time, participating teachers demonstrated an increased ability to create locally-developed assessment tools to assess student learning gains in their classrooms; to develop and evaluate pre- and post-tests; to define measurable outcomes in hard-to-quantify areas like art, music, and physical education; and to monitor student learning growth. They also showed a greater awareness of the importance of sound curriculum development, more alignment of curriculum with district objectives, and increased focus on higher quality content, skills, and instructional strategies.[23] Thus, the development and use of student learning evidence, in combination with examination of teaching performance, can stimulate improvements in practice.

Building Systems for Teacher Evaluation that Support Improvement and Decision Making

Systems that help teachers improve and that support timely and efficient personnel decisions have more than good instruments. Successful systems use multiple classroom observations across the year by expert evaluators looking at multiple sources of data that reflect a teacher's instructional practice, and they provide timely and meaningful feedback to the teacher.

For example, the Teacher Advancement Program, which is based on the standards of the NBPTS and INTASC, as well as the standards-based assessment rubrics developed in Connecticut,[24] ensures that teachers are evaluated four to six times a year by master / mentor teachers or principals who have been trained and certified in a rigorous four-day training. The indicators of good teaching are practices that have been found to be associated with desired student outcomes. Teachers also study the rubric and its implications for teaching and learning, look at and evaluate videotaped teaching episodes using the rubric, and engage in practice evaluations. After each observation, the evaluator and teacher meet to discuss the findings and to make a plan for ongoing growth. Ongoing professional development, mentoring, and classroom support are provided to help teachers meet these standards. Teachers in TAP schools report that this system, along with the intensive professional development offered, is substantially responsible for improvements in their practice and the gains in student achievement that have occurred in many TAP schools.[25]

In districts that use Peer Assistance and Review (PAR) programs, highly expert mentor teachers conduct some aspects of the evaluation and provide assistance to teachers who need it. Key features of these systems include not only the instruments used for evaluation but also the expertise of the consulting teachers or mentors – skilled teachers in the same subject areas and school levels who have released time to serve as mentors to support their fellow teachers – and the system of due process and review that involve a panel of both teachers and administrators in making recommendations about personnel decisions based on the evidence presented to them from the evaluations. Many systems using this approach have been found not only to improve teaching, but also to successfully identify teachers for continuation and tenure as well as intensive assistance and personnel action.[26]

Summary and Conclusions

New approaches to teacher evaluation should take advantage of research on teacher effectiveness. While there are considerable challenges in the use of value-added test scores to evaluate individual teachers directly, the use of value-added methods can help to validate measures that are productive for teacher evaluation.

With respect to value-added measures of student achievement tied to individual teachers, current research suggests that high-stakes, individual-level decisions, or comparisons across highly dissimilar schools or student populations, should be avoided. Valid interpretations require aggregate-level data and should ensure that background factors – including overall classroom composition – are as similar as possible across groups being compared. In general, such measures should be used only in a low-stakes fashion when they are part of an integrated analysis of what the teacher is doing and who is being taught.

Other teacher evaluation tools that have been found to be both predictive of student learning gains and productive for teacher learning include *standards-based evaluation processes.* These include systems like National Board Certification and performance assessments for beginning teacher licensing as well as district and school-level instruments based on professional teaching standards. Effective systems have developed an integrated set of measures that show what teachers do and what happens as a result. These measures may include evidence of student work and learning, as well as evidence of teacher practices derived from observations, videotapes, artifacts, and even student surveys.

These tools are most effective when embedded in systems that support evaluation expertise & well-grounded decisions, by ensuring that evaluators are trained, evaluation and feedback are frequent, mentoring and professional development are available, and processes are in place to support due process and timely decision making by an appropriate body.

With these features in place, evaluation can become a more useful part of a productive human capital system, supporting accurate information about teachers, helpful feedback, and well-grounded personnel decisions.

Endnotes

[1] Newton, X., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010) Value-Added Modeling of Teacher Effectiveness: An exploration of stability across models and contexts. *Educational Policy Analysis Archives, 18* (23). http://epaa.asu.edu/ojs/article/view/810; Sass

[2] Newton et al. (2010); Rothstein, J. (2007). Do Value-Added Models Add Value? Tracking, Fixed Effects, and Causal Inference. National Bureau for Economic Research.

[3] Briggs, D. & Domingue, B. (2011). Due Diligence and the Evaluation of Teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the Los Angeles Times. Boulder, CO: National Education Policy Center.

[4] Lockwood, J. R., McCaffrey, D. F., Hamilton, L.S., Stetcher, B., Le, V. N., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44* (1), 47 – 67.

[5] Bill & Melinda Gates Foundation (2010). Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project. Seattle: Author. Rothstein, Jesse (2011). Review of 'Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project. Boulder, CO: National Education Policy Center.

[6] Lockwood et al. (2007).

[7] Corcoran, Sean P., Jennifer L. Jennings, and Andrew A. Beveridge (2011). Teacher effectiveness on high- and low-stakes tests. Working paper. NY: New York University.

[8] Briggs & Domingue (2011). Rothstein, Jesse (2010). "Teacher Quality in Educational Production: Tracking, Decay, and Student Achievement," *Quarterly Journal of Economics* 125(1).

[9] Henry Braun (2005). Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models (Princeton, NJ: Educational Testing Service.

[10] Audrey Amrein-Beardsley & C. Collins (in press). The SAS® Education Value-Added Assessment System (EVAAS®): Its Intended and Unintended Effects in a Major Urban School System.

[11] Daniel F. McCaffrey, Daniel Koretz, J. R. Lockwood, Laura S. Hamilton (2005). Evaluating Value-Added Models for Teacher Accountability. Santa Monica: RAND Corporation.

[12] Henry Braun, Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models (Princeton, NJ: ETS, 2005), p. 17.

[13] National Research Council, Board on Testing and Assessment (2009). Letter Report to the U.S. Department of Education.

[14] For a summary of studies, see L. Darling-Hammond & J. Bransford (2005). *Preparing Teachers for a Changing World: What Teachers should Learn and Be Able to Do.* San Francisco: Jossey-Bass.

[15] See for example, L. Bond, T. Smith, W. Baker, & J. Hattie (2000). The certification system of the National Board for Professional Teaching Standards: A construct and consequential validity study (Greensboro, NC: Center for Educational Research and Evaluation); Cavaluzzo, L. (2004). Is National Board Certification an effective signal of teacher quality? (National Science Foundation No. REC-0107014). Alexandria, VA: The CNA Corporation; Goldhaber, D., & Anthony, E. (2005). Can teacher quality be effectively assessed? Seattle, WA: University of Washington and the Urban Institute; Smith, T., Gordon, B., Colby, S., & Wang, J. (2005). An examination of the relationship of the depth of student learning and National Board certification status (Office for Research on Teaching, Appalachian State University). Vandevoort, L. G., Amrein-Beardsley, A., & Berliner, D. C. (2004). National Board certified teachers and their students' achievement. *Education Policy Analysis Archives, 12*(46), 117.

[16] Steven Athanases (1994). Teachers' reports of the effects of preparing portfolios of literacy instruction. *Elementary School Journal, 94*(4), 421-43; Mistilina Sato, Ruth Chung Wei, & Linda Darling-Hammond (2008). Improving Teachers' Assessment Practices through Professional Development: The Case of National Board Certification, American Educational Research Journal, 45: pp. 669-700; Tracz, S.M., Sienty, S. & Mata, S. (1994, February). The self-reflection of teachers compiling portfolios for National Certification: Work in progress. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education. Chicago, IL; Tracz, S.M., Sienty, S. Todorov, K., Snyder, J., Takashima, B., Pensabene, R., Olsen, B., Pauls, L., & Sork, J. (1995, April). Improvement in teaching skills: Perspectives from National Board for Professional Teaching Standards field test network candidates. Paper presented at the annual meeting of the American Educational Research Association. San Francisco, CA.

[17] Mark Wilson & P.J. Hallum (2006). Using Student Achievement Test Scores as Evidence of External Validity for Indicators of Teacher Quality: Connecticut's *Beginning Educator Support and Training* Program. Berkeley, CA: University of California at Berkeley; Newton, S.P. (2011). Predictive Validity of the Performance Assessment for California Teachers. Stanford, CA: Stanford Center for Opportunity Policy in Education, 2010, available at http://scale.stanford.edu

[18] Ruth R. Chung (2008). Beyond Assessment: Performance Assessments in Teacher Education, Teacher Education Quarterly, 35 (1): 7-28; Ruth Chung Wei and Raymond Pecheone (2010). Teaching Performance Assessments as Summative Events and Educative Tools. In Mary Kennedy (ed.), Teacher Assessment and Teacher Quality: A Handbook (New York: Jossey-Bass).

[19] Anthony Milanowski, S.M. Kimball, B. White (2004). The relationship between standards-based teacher evaluation scores and student achievement. University of Wisconsin-Madison: Consortium for Policy Research in Education; Anthony Milanowski (2004). The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati, Peabody Journal of Education 79 (4): 33-53. Jonah Rockoff & Cecilia Speroni (2010). Subjective and Objective Evaluations of Teacher Effectiveness (New York: Columbia University, 2010).

[20] Bill & Melinda Gates Foundation (2010). *Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project*. Seattle: Author.

[21] C. K. Jackson & E. Bruegmann (2009, August). *Teaching students and teaching each other: The importance of peer learning for teachers.* Washington, DC: National Bureau of Economic Research.

[22] Y. Goddard & R. D. Goddard (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

[23] Richard Packard & Mary Dereshiwsky (1991). *Final quantitative assessment of the Arizona career ladder pilot-test project.* Flagstaff: Northern Arizona University.

[24] The teacher responsibility rubrics were designed based on several teacher accountability systems currently in use, including the Rochester (New York) Career in Teaching Program, Douglas County (Colorado) *Teacher's Performance Pay Plan*, Vaughn Next Century Charter School (Los Angeles, CA) Performance Pay Plan, and Rolla (Missouri) School District Professional Based Teacher Evaluation

[25] Lewis Solomon, J. Todd White, Donna Cohen & Deborah Woo (2007). *The effectiveness of the Teacher Advancement Program.* National Institute for Excellence in Teaching, 2007.

[25] National Commission on Teaching and America's Future (1996). What Matters Most: Teaching for America's Future. NY: NCTAF; Piet Van Lier (2008). Learning from Ohio's best teachers: A homegrown model to improve our schools. Policy Matters Ohio.