

**HOW TO EVALUATE AND RETAIN EFFECTIVE TEACHERS**  
**Study by the League of Women Voters of South Carolina, 2011-2013**

**STUDY REPORT**

**About the Study**

The genesis of this study was a forum: Getting the Best Teachers in our Classrooms: A Conversation on Evaluating Teacher Effectiveness held by the LWV of the Charleston Area on March 22, 2011. The forum explored the current movement in South Carolina and across the nation toward “performance based” evaluation of teachers. LWVCA members determined that the topic was timely and worthy of further investigation. Since any change of policy on teacher evaluation in South Carolina would be at the behest of the SC Superintendent of Education or the General Assembly, LWVCA brought a study proposal to the LWVSC Convention in the spring of 2011. Delegates at that convention adopted a two-year study, “How to Evaluate and Retain Effective Teachers. “

The purpose of the study is to examine the growing movement toward “results based” evaluation nationally and in South Carolina. Goals of the study are to propose one or more positions for consensus and action; raise public interest and knowledge about the issues surrounding teacher effectiveness and evaluation; and stimulate a dialog across the state among the public and members of the education community, i.e., teachers, school administrators, school board, SC Department of Education.

The study has been led by two members of the LWV Charleston Area: Jon Butzon, LWVSC Education chair, and Peggy Huchet, LWV Charleston Area.

In the summer of 2011, LWVSC President Barbara Zia sent several letters to local LWV presidents announcing the formation of a study committee and inviting them to appoint at least one member. The committee held its first meeting in Columbia in September, 2011. During the following months, committee members independently carried out research and wrote reports. The committee chairs compiled and distributed the reports to the committee members. A second committee meeting was held in March 2012 to discuss consensus questions.

Study Committee members include: Pat Borenstein, LWV Greenville County; Jon Butzon, LWV Charleston Area; Paula Egelson, LWV Charleston Area; Pamela Hollinger, LWV Hilton Head Island/Bluffton; Peggy Huchet, LWV Charleston Area.

**Teachers Matter**

Hundreds of studies have examined the factors that affect student academic achievement. Virtually all of them conclude that the most important school-based factor is an effective teacher. A sample of the findings by respected investigators includes:

Eric Hanushek, leader in the development of the economic analysis of educational issues, writes: “The general finding about the importance of teachers comes from the fact that the average gains in learning across classrooms, even classrooms within the same school, are very different. Some teachers year after year produce bigger gains in student learning than other teachers. The magnitude of the differences is truly large, with some teachers producing 1.5 years of gain in achievement in an academic year while others with equivalent students produce only 1/2 year of gain. In other words, two students starting at the same level of achievement

Report, p. 1

can know vastly different amounts at the end of a single academic year due solely to the teacher to which they are assigned. If a bad year is compounded by other bad years, it may not be possible for the student to recover.”<sup>1</sup> Hanushek goes on to say that the importance of a highly-qualified teacher is overwhelming. The quality of teaching is “far and away the most important factor in student achievement.”

“It is fairly well documented that the best school predictor of student outcomes is high-quality, effective teaching as defined by performance in the classroom (Goldhaber, 2002; Goldhaber & Brewer, 2000; Hanushek, Kain, O'Brien, & Rivkin, 2005; Wright, Horn, & Sanders, 1997). More recently Aaronson, Barrow, and Sander (2007) examined data from the Chicago Public Schools and found that a one-standard-deviation, one-semester improvement in math teacher quality raised student math scores by 0.13 grade equivalents or, over one year, roughly one-fifth of average yearly gains.

“William Sanders, who pioneered the Tennessee Value-Added Assessment System, summarizing his own studies, stated that especially in math, the cumulative and residual effects of teachers are still measurable at least four years after students leave a classroom (Sanders, 2000, p. 335). A study by Nye, Konstantopoulos, and Hedges (2004) was unusual because it randomly assigned students to classes, estimated teacher effects on student achievement over four years. Their estimates of teacher effects on achievement gains were similar in magnitude to those of previous studies done by economists, but they found larger effects on mathematics than on reading achievement.”<sup>2</sup>

In 2012, researchers at Harvard and Columbia published a large longitudinal study that confirmed the long-term impact of high quality teachers. They found that when a highly effective teacher joins a school, test scores immediately rise. They also were able to associate long term gains with students assigned to effective teachers, e.g. they were more likely to attend college, earn higher salaries, live in better neighborhoods and save more for retirement. “Overall, our study shows that great teachers create great value and that test score impacts are helpful in identifying such teachers.” Although there are issues about best way to use tests, “it is clear that improving the quality of teaching – whether using value-added or other tools – is likely to have large economic and social returns.”<sup>3</sup>

Linda Darling-Hammond wrote: “For more than two decades, policymakers have undertaken many and varied reforms to improve schools, ranging from new standards and tests to redesigned schools, new curricula and new governance models. One important lesson from these efforts is the repeated finding that teachers are the fulcrum determining whether any school initiative tips toward success or failure. Every aspect of school reform depends on highly skilled teachers for its success. This is especially true as educational standards rise and the diversity of the student body increases. Teachers need even more sophisticated abilities to teach more complex curriculum to the growing number of public school students who have fewer educational resources at home, those who are new English language learners, and those who have distinctive learning needs.”<sup>4</sup>

### **Traditional Evaluation Practices**

Toch and Rothman's 2008 report *Rush to Judgment* provided a thought-provoking perspective on teacher evaluation. They asserted that current supervisory and evaluative practices were "superficial, capricious, and often don't even directly address the quality of instruction, much less measure students' learning." Specifically, they described teaching as a profession that

Report, p. 2

focuses on formal credentials rather than on instructional effectiveness and student achievement. Regarding typical evaluation procedures, Michigan State professor Mary Kennedy was quoted as saying, "in most instances, it's nothing more than marking satisfactory or unsatisfactory."<sup>5</sup>

*The Widget Effect*, published in 2001, was a blunt critique of traditional teacher evaluation policy "[P]olicymaking around improving teacher quality to date has focused almost exclusively on a teacher's qualifications – teacher credentials, majors, degrees and licensing. Those criteria would be all well and good if they were associated with positive gains in student learning. Unfortunately, by and large, they are not." (p. 2) "Put simply, they fail to distinguish great teaching from good, good from fair, and fair from poor. A teacher's effectiveness—the most important factor for schools in improving student achievement—is not measured, recorded, or used to inform decision-making in any meaningful way."

The authors of the *Widget Effect* studied teacher evaluation practices in 12 school districts and found:

- There was a striking disconnect between student achievement and teacher ratings. Over 90% of teachers were rated as satisfactory even as large percentages of students were failing.
- Most districts rated teachers in two categories, satisfactory and unsatisfactory. Virtually all tenured teachers consistently were rated as satisfactory. Even when rating systems had more choices, most teachers rated in the highest one or two categories.
- Typically, very few teachers were dismissed, ratings were inflated and feedback was minimal. Because most teachers received high ratings, teachers resisted an evaluation that indicates that they need improvement. "The result is a dysfunctional school community in which performance problems cannot be openly identified or addressed"
- "Excellence goes unrecognized, development is neglected and poor performance goes unaddressed." Teachers are treated as if they are "interchangeable parts." Hence, the widget effect.<sup>6</sup>

Here are some **South Carolina** statistics.

During the 2012-2011 school years, 98% of all South Carolina teachers were rated as "satisfactory."

Recent student achievement outcomes were somewhat different. On the 2012 Palmetto Assessment of State Standards, depending on grade level (3<sup>rd</sup> – 8<sup>th</sup>), students who did not meet the minimum state standard in each area tested: Reading and Research – between 19.7% and 30.3%; Mathematics – between 21.6% and 31.4%; Science – between 24.6% and 39.3%; and Social Studies – between 19.1% and 31.3%.

South Carolina's on time graduation rate in 2011 was 73.6%. On the SAT in 2012, South Carolina (1431) ranked 48<sup>th</sup> out of 50 states and the District of Columbia.

## **Performance Based Evaluation**

Report, p. 3

These and similar findings have prompted a shift from relying on teacher qualifications to evaluating teacher performance. States and school districts, searching for “effective” teachers, are examining results, i.e., the academic achievement of their students.

Goldhaber and Hansen write: “Since the turn of the 21st century, emphasis has shifted from supervision of teachers to the evaluation of their performance, as well as from teacher behavior to student achievement. In their 2005 book *Linking Teacher Evaluation and Student Learning*, Tucker and Stronge argued for evaluation systems that determine teacher effectiveness using evidence from student gains in learning as well as observations of classroom instruction. They examined the supervisory systems in four different school districts that used data on instructional practices and learning gains and drew a series of recommendations supporting the use of both types of data. However, their recommendations regarding the use of student achievement data were the most forcefully stated: ‘Given the clear and undeniable link that exists between teacher effectiveness and student learning, we support the use of student achievement information in teacher assessment. Student achievement can, and indeed should be, an important source of feedback on the effectiveness of schools, administrators, and teachers’.”<sup>7</sup>

### **What is Being Measured: Value-Added Growth Model**

It is important to understand that performance-based teacher evaluation systems measure student growth over a given time period, typically a school year. They do not measure whether students are on “grade level” or have acquired all of the skills spelled out for a specific grade. Growth models are designed to assess the progress that individual students make over time. In the value-added growth model, student academic growth over a school year is measured against anticipated yearly student achievement goals. Students are tested at the end of each school year. The difference between the anticipated score and the actual test score is the growth score. Teachers whose students repeatedly perform better than predicted can be considered highly effective.

Many factors that affect a child’s academic performance are not within the control of the school or teacher, e.g. poverty, race, ethnicity. The value-added growth model may consider the impact of these factors among others in order to determine how much of a student’s progress can be attributed to the teacher.

As explained by the National Association of State Boards of Education (NASBE): “To address these issues, many education policymakers, researchers, and practitioners have turned to value-added assessments (or more accurately, ‘value-added models’) for analyzing assessment data. Value added assessment models are statistical approaches that use multiyear student test score data, and student background characteristics in some models, to attribute student growth to schools, teachers, or both. In other words, value-added models attempt to determine how far a student has progressed compared to where the student started and to what degree that growth can be attributed to educational factors (as opposed to ‘external’ factors such as socioeconomic status, race, parents’ educational levels, or innate ability).”<sup>8</sup>

As explained by Laura Goe: “In its most simple form, the value-added measure as it is used for evaluating teachers is calculated as follows: Students’ previous test scores are used to create *predicted* test scores for a given year. The difference between the predicted and actual test scores are growth scores. Teachers’ contribution to students’ learning is determined by looking at the average of all of their students’ growth scores. The teachers are then ranked against

other teachers within a district (or other unit of interest) according to how much they contributed to students' growth, and this ranking is their value-added "score." In some value-added models, only students' prior achievement scores are used in the calculation; other models include students' gender, race, and socioeconomic background; still others include information about teachers' experience. With a value-added measure, teachers whose students performed about as well as predicted are considered "average" teachers, those whose students performed much better than predicted are considered "above average" or "highly effective," and those whose students performed worse than expected are considered "below average."<sup>9</sup>

The value-added model has great potential to provide information to strengthen school performance, identify effective practices and teaching techniques, predict a teacher's future performance and stimulate improvement in the quality of teacher preparation programs. There is not consensus among researchers about what are the "right" factors to incorporate into a value-added measure and many different formulas are proposed. Most authorities caution against evaluating teacher performance based solely on a value-added measure.

NASBE comments: "perhaps the most important feature of value-added assessment is that it serves to keep everyone's focus on student growth and learning momentum which is the essence of the schooling experience. After all, is it not the main role of educators to take children from where they find them and then "add value?"<sup>10</sup>

### **What Should a Teacher Evaluation System Accomplish?**

Teacher evaluation should help school districts attain the goal of putting an effective teacher in each classroom. It should be a self-perpetuating system for attracting, rewarding and retaining effective teachers. To do so, it should:

- Provide the means to identify, reward and retain top notch teachers
- Distinguish between outstanding, average and ineffective teachers
- Advance the effective teachers and eliminate the non performers.
- Improve skills of all teachers

### **Country-Wide Trend**

Statistics show the dramatic trend of states moving from reliance on teacher qualifications as determined by educational background and years of service toward policies that require systematic, performance-based teacher evaluation. Investigation by the National Council on Teacher Quality in 2011 reveals:

- In the past 3 years, 32 states and the District of Columbia have changed the state evaluation policy.
- Twenty-four states and DC now require an annual evaluation for all teacher compared to 15 two years ago.
- In 2009, 35 states did not require evidence of student learning in teacher evaluations. Today, 22 states require evidence of student growth.
- Seventeen states specify that student achievement will be a preponderant or significant factor.
- In 18 states plus DC, poor evaluation ratings can prompt dismissal.<sup>11</sup>

### **Federal Incentives: No Child Left Behind, Race to the Top**

Report, p. 5

The No Child Left Behind Act of 2001 (NCLB) for the first time sought to raise teacher quality, particularly in schools serving low-income students. The goal was to ensure that the children of poor families had the same access to good teachers as other students.

NCLB has produced some positive changes. Chief among those changes has been the requirement for schools, school districts, and states to disaggregate their student achievement data. No longer can the bad news about certain groups of children be hidden or diluted by being folded into the “average.”

Under NCLB, teachers are required to be “highly qualified.” i.e., have at least a bachelor’s degree, have full state certification, and demonstrate knowledge of the subject matter they teach. What we now realize is that “highly qualified” and “highly effective” are very different criteria. The Center for Public Education wrote in 2009, “The overwhelming majority of teachers have met the law’s definition of ‘highly qualified,’ yet there is little evidence that teacher quality has improved markedly. While these characteristics defined base-level requirements, they did not ensure that qualified teachers were effective in the classroom.”<sup>12</sup>

In 2009, the federal Department of Education announced Race to the Top (RTTT), “a competitive grant program designed to encourage and reward States that are creating the conditions for education innovation and reform; achieving significant improvement in student outcomes, including making substantial gains in student achievement, closing achievement gaps, improving high school graduation rates, and ensuring student preparation for success in college and careers.”<sup>13</sup>

RTTT defines an “effective teacher” as “a teacher whose students achieve acceptable rates (e.g., at least one grade level in an academic year) of student growth and a “highly effective teacher” as “a teacher whose students achieve high rates (e.g., one and one-half grade levels in an academic year) of student growth. Teacher evaluation must include multiple measures, “provided that teacher effectiveness is evaluated, in significant part, by student growth.”<sup>14</sup>

To be competitive, an applicant must provide evidence of substantial progress toward: “a) adopting internationally benchmarked standards and assessments that prepare students for success in college and the workplace; b) recruiting, developing, retaining, and rewarding effective teachers and principals, especially where they are most needed; c) building data systems that measure student success and inform teachers and principals about how they can improve instruction; and d) turning around the lowest achieving schools.”<sup>15</sup>

The opportunity to capture a share of the \$4 billion offered by RTTT has been the incentive for many state legislatures to approve teacher evaluation policies that incorporate measures of student growth. Even states that did not apply have begun to follow suit.

### **Issues Associated with Performance-based Evaluation**

“The task of building better evaluation systems is as difficult as it is important. Many hurdles stand in the way of rating teachers fairly on the basis of their students’ achievement, the solution favored by many education experts today. And it’s increasingly clear that it’s not enough merely to create more-defensible systems for rewarding or removing teachers. Teacher evaluations pay much larger dividends when they also play a role in improving teaching.”<sup>16</sup>

Log on to any chat room concerning education and you will find vigorous arguments for and against performance-based evaluation policies. On the negative side, many educators believe that teachers are being held accountable for outside factors beyond their control, e.g., poverty, broken families, disabilities. They point to obstacles presented by inadequate resources, lack of support from administration and breakdown of school discipline. They point to the dearth of reliable, unbiased measures of student growth and performance. The new rating systems, they fear, will encourage competition, destroy teacher morale and diminish the teaching profession. To them, the reform movement represents just the latest version of “teacher-bashing.”

Advocates of performance-based evaluation counter that it is possible to isolate and account for the issues that teachers can and cannot control. Given the evidence of the strong impact that teachers have on student achievement, they assert that it is incumbent on school districts to institute a system that reveals the tremendous variations in teacher skills and advances the highest performers. And student academic performance is an integral part of the process. They consider that systematic evaluations and acknowledgement of excellence will add credibility to the teaching profession.

### **Cautions: Use of Test Scores**

Much of the controversy surrounding performance-based evaluation centers on the use of standardized test scores as the measure of academic growth. Often, these tests are the only available instruments to measure student achievement on a large scale. Among the concerns expressed about using standardized test scores to evaluate teacher performance.

*Reliability:* One criticism is that achievement tests scores are not statistically accurate or reliable enough to support high stakes personnel and compensation decisions. Is it valid to evaluate a teacher on the basis of one yearly test? And what accounts for the variation in scores of a given teacher’s students from year to year?

*Limited applicability:* Students are tested in a few subjects, typically math, reading, science and social studies and only in grades 3 through 8. It is of concern that there are no comparable measures for teachers who teach in other subject areas and for personnel such as librarians.

*Accountability:* Beyond the primary grades, students study with several teachers. The question raised is how is it possible to isolate a given teacher’s contribution to student progress? The same holds true for team teaching or similar arrangements.

Students are transient, especially in unstable, low income neighborhoods. So, another issue is how you can ensure the impact of a teacher when many of her students are only in her classroom for a portion of the school year.

Some people have asserted that test scores do not necessarily reflect the impact of the current teacher but the cumulative effect of previous teaching.

*Equity:* Another expressed concern deals with equity. Given the negative impact of societal factors beyond the control of the school, teachers of disadvantaged students have a more difficult job than those who teach children from well-to-do families. How can it be fair to compare them?

*Unintended consequences:* Some people worry that teachers might narrow their curriculum and “teach to the test.” Often the stakes are so high that teachers will resort to cheating if necessary.

*Narrow scope:* And it is pointed out that tests do not measure teacher contributions to valued student outcomes, such as moral character, civic engagement, enthusiasm for learning, open-mindedness.

### **Multiple Measures**

The bulk of researchers and practitioners readily acknowledge the limitations of existing tests and caution against relying on test scores alone to identify effective teachers. There is widespread agreement that what is needed is a comprehensive evaluation system that uses multiple measures. In fact, we could find no example of anyone who is advocating for the use of test scores alone in evaluating teacher performance. Most authorities would keep student test scores as part of the assessment and encourage development of instruments that better measure student growth.

Goe observes: “Even though value-added models are useful in tracking student progress over time, there are limitations and complexities involving these methods and the resultant data, particularly when used as the sole measure of teacher effectiveness. Educators and policymakers should give careful consideration to these concerns before committing to using value-added methods in ways that may not be warranted, given the current state of our understanding about the methodology and theory. The bottom line is that value-added data are limited in what they can tell about teacher quality and effectiveness; these data are more useful when supplemented by other measures and sources of evidence.”<sup>17</sup>

Toch and Rothman state: “[S]tandardized tests, with their many limitations, provide only a partial picture of what students know and are able to do. They aren’t great measures of student achievement. As a result, it’s important to evaluate teachers’ actual instruction—the way they work with their students in their classrooms, from their teaching techniques to the types of homework they assign.

“To get a fuller and fairer sense of teachers’ performance, evaluations should focus on teachers’ instruction—the way they plan, teach, test, manage, and motivate. They need to move far beyond principal drive-bys to multiple measures, multiple evaluations, and multiple evaluators. And they should contribute to helping teachers improve their performance to a far greater degree than they do in most public schools today—both to promote a climate that attracts the best and brightest into teaching and to spend public education’s vast “professional development” monies far more efficiently than most school systems do today. Where possible, in the most defensible ways possible, student test scores should have a role in teacher evaluations. School systems should evaluate both the work that teachers do in their classrooms and the results of that work. As Joan Baratz-Snowden, a former director of educational issues at the American Federation of Teachers, says: “Anyone who dismisses student learning in evaluations is naïve. Anyone who defines student learning as tests scores is also naïve.”<sup>18</sup>

One measure that is used universally for teacher evaluation is the structured classroom observation. Essential to the process is a well trained observer, who provides immediate feedback, engages the teacher in analyzing her performance and provides support and

encouragement to develop new skills. Classroom observation has also been found to be an indicator of future teacher performance.

Kane et.al. reported on an ongoing study in the Cincinnati public schools: "We find that evaluations based on well-executed classroom observations do identify effective teachers and teaching practices. Teachers' scores on the classroom observation components of Cincinnati's evaluation system reliably predict the achievement gains made by their students in both math and reading. These findings support the idea that teacher evaluation systems need not be based on test scores alone in order to provide useful information about which teachers are most effective in raising student achievement."<sup>19</sup>

Darling-Hammond has developed a structured Teacher Performance Assessment (TPA) model to measure classroom performance. ".These assessments require teachers to document their plans and teaching for a unit of instruction, videotape and analyze their teaching, and collect and evaluate evidence of student learning. All of these pieces of evidence are then assembled and evaluated by highly trained raters who score them in a consistent manner against specific criteria that reflect standards of best practice."<sup>20</sup>

Structured performance assessments can be used to evaluate both new and veteran teachers. They can be applied consistently and scored reliably. Darling-Hammond and others emphasize that the structured classroom observation is an opportunity to give novice teachers some much-needed support and mentoring. "Developing teacher effectiveness is as important as measuring it. Many studies have concluded that teachers' participation in standards-based performance assessments can help teachers improve their practice."<sup>21</sup> Engaging teachers at all levels to analyze and improve their techniques will enable school districts to identify, motivate and retain effective teachers.

Relevant to but outside the scope of this study is the often mentioned need to improve the quality and thoroughness of teacher and principal preparation programs. Structured assessments of new teachers will reveal the weaknesses and gaps in their training. States can use this information to prompt teacher-training programs to fill in these gaps and can raise their standards for accreditation credentials,

School districts may use other measures of teacher performance to supplement classroom observations and testing, such as portfolios of student's work; quality of lesson plans; surveys of parents and students; survey of peers; professionalism and contribution to the school.

### **Elements of a Performance-Based Evaluation Policy**

There is much written about how to construct a comprehensive, performance based evaluation system. Among the many models that are proposed, there is a core of essential components that are repeatedly emphasized. Here are the recommendations from the authors of *The Widget Effect*:

- Institute a "comprehensive performance evaluation and development system" that differentiates among teachers re effectiveness in promoting achievement of students. "This demands clear performance standards, multiple rating options, regular monitoring of administrator judgments, and frequent feedback to teachers."<sup>22</sup>
- Provide individualized professional development and intensive support when needed.
- Incorporate a value added model as part of a comprehensive evaluation system<sup>23</sup>

Report, p. 9

- Train evaluators thoroughly and hold them accountable for using providing constructive feedback and support.
- Establish an evaluation system that is perceived as credible, i.e., fair and accurate, and use it to make decisions about tenure, compensation, advancement, retention and dismissal.
- A system that has clear performance standards, frequent evaluations and follow up support will put unsuccessful teachers on notice about their performance. This will encourage unsatisfactory performers to voluntarily exit the system or expedite their dismissal because evaluation record is adequate documentation.<sup>24</sup>

The authors of *The State of the States* examined 17 states in which new evaluation policies were being implemented. Drawing on their research and their professional experience, the authors put forward early lessons regarding pitfalls and strategies that seem to be worth emulating. Many of their recommendations echo those above. In addition, some comments that provide additional guidance:

- Measures of teacher effectiveness, although imperfect, are a marked improvement over evaluation systems that ignore teacher impact. However, value added measures are “still emerging.” (p. 28) and it is necessary to use existing measures with caution. The system will fail if expectations are unrealistic.
- Not all teachers can or need to be assessed exactly the same way. For the subject areas where no state-wide tests are available, states should develop comparable measures where it is possible. However, for subject areas where comparable tests do not exist or are not appropriate, states must seek other measures of effectiveness. “Comparability of all measures isn’t the ultimate goal; *fair, rigorous and appropriate measures* of teacher performance are the bottom line.” (p. 29)
- Classroom observations are important and must be conducted by skilled evaluators using standardized protocol. This will require states to make a significant investment in training the evaluators.
- States policy should assure “an equitable distribution” of the most effective teachers to high-need schools. (p.37)<sup>25</sup>

*The State of the State* authors conclude: “What this policy review and early lessons suggest is that performance-based teacher evaluation must be approached in a measured, realistic and transparent way. Performance measures are not perfect and good teachers are not the product of formulas. Conducting teacher performance evaluations that focus on the results and the behaviors that matter most will move us toward a system that recognizes and encourages effective instruction and prepares and values highly-effective teachers.”<sup>26</sup>

Careful implementation of a new evaluation system is deemed to be essential. Reform will succeed only if the policy is perceived to be fair and transparent. Involving teachers in its development and introduction will smooth the way to acceptance by all of the stakeholders. “[A]ny new system is only as strong as the team that builds it. In other words, collaboration is essential. As states and districts begin to reform their teacher evaluation systems, involving as many stakeholders as possible and encouraging feedback throughout the process is crucial. Gaining perspective through the lens of teachers, parents, district leaders, unions, administrators, and community members ensures that the process produces valuable

information from all stakeholders, is fair and inclusive, and finally, that the evaluation system will be successful and sustainable.<sup>27</sup>

Based on the research of the Bill and Melinda Gates Foundation, Bill Gates said: “The **first and most important feature** of a strong evaluation and development system is heavy teacher involvement throughout – from the conceptual stage, to the roll out, to revising the program once it's underway. If someone wants to rush an evaluation system into place – and they think they can speed it through by doing it without the teachers – that is a grave mistake. The system will be low-quality, and will never get buy-in from the teachers.”

Gates continued: “A new teacher evaluation system is not automatically a good thing. If states and school districts feel pressured to rush out new systems, those systems could evaluate teachers unfairly and fail to help teachers improve. That would be a disaster. A flawed execution of a good idea could convince people it is a bad idea – and that could kill this push for reform.”<sup>28</sup>

### **Costs**

There are costs associated with successful implementation of a performance-based evaluation policy. Among the expenses:

- States and school districts need to establish a data base that will enable them to track the achievement of students from one year to the next and compare the progress made by similar students assigned to different teachers.
- There will be personnel costs. It will be necessary to train the evaluators. School principals also will need training and probably will have to commit significant time to conducting evaluations. It may be necessary to employ additional teachers and support staff. And often districts contract with consultants to provide guidance, conduct training and oversee implementation.
- There is a need to improve and expand the measures of student progress. This means developing measures for teachers at grade levels and in disciplines where no standardized tests are available as well as improving existing instruments.

The state or the school district or both will have to make a substantial commitment of funds to initiate and implement a credible system. It's possible that some existing funds can be reallocated but new investment will be required.

Gates estimates that “it will cost between 1.5 and 2 percent of the overall budget for teacher compensation and benefits to implement an evaluation system based on multiple measures of teaching performance.” He has cautioned against trying to “skimp on paying teachers to do classroom observations, to cut corners on training the evaluators, to get stingy on providing the feedback that will help teachers improve. But saving money on those measures would be like saving money on a car by leaving out the engine. One-point-five to two percent is a small investment compared to what is paid now for teacher development that shows little results. It's possible that the costs could be met by reallocating existing dollars. And the returns in student achievement will be many times the investment”<sup>29</sup>.

### **What Others States are Doing**

There are many models of evaluation systems developed by teachers, teacher educators, researchers and advocacy groups. Among them: Charlotte Danielson's Framework for

Teaching; TAP: The Teacher and Student Achievement Program; The University of Virginia's Teaching Performance Record; The Bill & Melinda Gates Foundation's Measures of Effective Teaching; The New Teacher Center's Formative Assessment System; Educational Leadership Policy Standards from the Interstate School Leaders Licensure Consortium. Some states have adopted a specific evaluation model and have contracted with the author to develop the policy, train staff and oversee implementation.

There are common components in the performance based evaluation policies being adopted across the country:

All teachers are evaluated throughout their careers. Teachers are evaluated on a regular schedule. More experienced teachers may be evaluated on a less frequent schedule.

Ratings go beyond satisfactory/ unsatisfactory. There are three to five categories, e.g. superior, above average, average, below average, ineffective.

Evaluation consists of multiple measures, one of which is student performance. Student performance typically accounts for 25 to 50 percent of a teacher's rating score. Other evaluation factors: observations; student portfolio, survey of student, parents, peers, professionalism and collegiality.

To insure consistency in implementation, the policy specifies that raters are to be qualified and trained.

Following each evaluation, teachers receive immediate feedback and support. A significant objective of the evaluation is to identify areas needing improvement and target training to address these areas.

Ratings have consequences. Ratings are the basis for decisions regarding retention/dismissal, compensation, assignments. Retention of even long-term employees is no longer guaranteed.

### **ADEPT: South Carolina's Evaluation System**

The South Carolina ADEPT system (Assisting, Developing, and Evaluating Professional Teaching) was initiated in 1998. It was revised in 2005 to assure that the performance standards were "consistent with current research and best practice."

According to an overview prepared by the S.C. Department of Education, the ADEPT system provides different amounts of assistance, development and evaluation according to the needs of teachers as they progress through stages of their careers. Novices receive formal evaluation and assistance, whereas veteran teachers are presumed capable of self-evaluation and improvement. In addition, ADEPT sets standards for certification and requires that teacher education programs in the State offer courses that focus on the necessary skills to meet these standards.

Phase one. Beginning teachers are employed under "induction contracts." Evaluation provides them with feedback, guidance and support so that they can transition into professional practice. In 2010-2011, 92% of the 2027 teachers in this category were judged to meet ADEPT standards.

Report, p. 12

Phase two: Teachers who have completed an induction year are eligible for employment under an annual contract. An evaluation, called the annual-formal 1 evaluation, is required. In 2010-2011, 87% of the 2450 teachers employed under annual contracts met ADEPT standards.

Phase three: Annual contract teachers undergo the annual-formal 2 evaluation the following year. "It is at this point that high-stakes, consequential decisions are made on the basis of teaching performance. Given appropriate preparation and support during the previous two stages, most teachers are successful in achieving the high level of performance defined by the ADEPT standards. Nonetheless, the formal evaluation process is an important part of quality assurance." Seventy-five % of 117 teachers in this category met ADEPT standards in 2012-2011.

A small number of teachers remain at the annual contract level for another year and receive an additional year of annual diagnostic assistance (ADA) In 2010-2011, 85% of the 197 teachers in this category met ADEPT standards. Teachers who fail this formal process twice are automatically suspended from teaching in any public school in S.C. for a minimum of two years and must complete a state-approved program of remediation to have their teaching certificate reinstated.

Phase four: Most teachers move into "continuing contract status", which guarantees due process rights related to employment and dismissal. From that time forward, the responsibility for professional growth now falls on the teacher" in what is called "goals-based evaluation" (GBE). Evaluations may be formal or informal at the discretion of the district. "It is through this self-directed goals-based process that experienced, successful teachers are able to engage in lifelong learning and, further, to give back to the profession."

The great majority of teachers in South Carolina are in continuing contract status. In 2010-2011, there were 43,548 continuing GBE teachers; 99% of them were judged as meeting ADEPT standards.

Altogether, of the 50,630 teachers in the state in 2010-2011, 98% met ADEPT standards, 1% did not meet ADEPT standards. Data was not available for the remaining few (just over 1%). These statistics have been consistent over the past 10 years and probably before.<sup>30</sup>

#### **Issues to Consider About ADEPT:**

- The pass/fail system does not make it possible to differentiate among teachers.
- After a teacher is under a continuing contract, retention is virtually assured. There is no systematic, universally applied process to dismiss ineffective teachers.
- For "continuing contract" teachers, the frequency and scope of teacher evaluation depends on the practices of individual school principals. Consequently, experienced teachers do not receive the feedback, guidance and support that are the benefits of systematic formal evaluations.
- ADEPT evaluation criteria fall into four domains: Planning, Instruction, Classroom Environment, Professionalism. These criteria might best be described as "inputs" in the teaching process and are not necessarily indicative of the results of teaching in terms of student achievement.
- There is no meaningful way to incorporate student performance as a basis for teacher evaluation. .

---

<sup>1</sup> Hanushek, Eric, *The Economic Value of Higher Teacher Quality*, Economics of Education Review 30(2011) 466-479, p. 467

<sup>2</sup> Gargani, John, Ozge Hacifazlioglu and Michael Strong, Do we know a successful teacher when we see one? Experiments in the Identification of effective teachers, *Journal of Teacher Education*. 62.4 (September-October 2011): p. 367

<sup>3</sup> Chetty, Raz, John Friendam and Jonah Rockoff, *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood*," December 2011

<sup>4</sup> Darling Hammond, Linda, *Evaluating Teacher Effectiveness, How Teacher Performance Assessments Can Measure and Improve Teaching*, Center for American Progress, October 2010, p. 1

<sup>5</sup> Toch, Thomas and Robert Rothman, *Rush to Judgment: Teacher Evaluation in Public Education*, Education Sector Reports, 2008, pp. 1-2

<sup>6</sup> Weisberg, Daniel, Susan Sexton, Jennifer Mulhern, David Keeling, *The Widget Effect,, Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*, 2<sup>nd</sup> edition, The New Teacher Project, 2001

<sup>7</sup> Goldhaber, Dan and Michael Hansen, *Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions*, Center of Reinventing Public Education working paper, 2009

<sup>8</sup> *Evaluating Value-Added: Findings and Recommendations From the NASBE Study Group of Value-Added Assessments* (National Association of State Boards of Education, 2005).

<sup>9</sup> Goe, Laura, *Key Issue: Using Value-Added Models to Identify and Support Highly Effective Teachers*, National Comprehensive Center for Teacher Quality, May, 2008, p. 2

<sup>10</sup> NASBE, op. cit., p.

<sup>11</sup> *State of the States, Trends and Early Lessons on Teacher Evaluation and Effectiveness Policies*, National Council on Teacher Quality, October, 2011

<sup>12</sup> "Does Highly Qualified Mean Highly Effective?", Center for Public Education, 2009, p. 1

<sup>13</sup> *Race to the Top Executive Summary*, U.S. Department of Education, Washington, DC 20202, November 2009, p.2

<sup>14</sup> U.S. Department of Education, op. cit., p.11

<sup>15</sup> *Reauthorizing ESEA, Making Research Relevant, Educator Effectiveness*, A Pocket Guide, American Institutes for Research, 2011, p. 4

Report, p. 14

---

<sup>16</sup> Toch and Rothman, op. cit, p. 1

<sup>17</sup> Goe, Laura, op. cit p.

<sup>18</sup> Toch and Rothman, op. cit. pp. 11-18

<sup>19</sup> Kane, Thomas J., Amy L. Woolen, John H. Tyler and Eric S. Taylor, *Evaluating Teacher Effectiveness, Can Classroom Observations Identify Practices that Raise Achievement?*. Education Next, volume 11, number 3, Summer 2011.

<sup>20</sup> Darling-Hammond, op.cit, pp. 7-9

<sup>21</sup> Ibid, p. 12

<sup>22</sup> Weisberg, Daniel, et. al., op Cit, p. 7

<sup>23</sup> Ibid, p. 27

<sup>24</sup> Weisberg, Daniel, et. al., op. cit.

<sup>25</sup> National Council on Teacher Quality, op. cit, pp 28-37

<sup>26</sup> Ibid, p. iv

<sup>27</sup> YuenTing Liu, Vivian and Laura Steen Mullinger, *Making it Meaningful: Building a Fair Evaluation System*, TE<sup>2</sup> Community Brief, September 2011

<sup>28</sup> Gates, Bill, Speech to the Education Commission of the States Annual Conference, July 11, 2012, Bill and Melinda Gates Foundation

<sup>29</sup> Ibid

<sup>30</sup> *An overview of the ADEPT System*, SC Department of Education, June, 2006 and *Final ADEPT Results 2010-11*, Division of School Effectiveness, SC State Department of Education.”