Educator Evaluation – Models, Parameters, Issues and Implementation

Dr. Edward Roeber
Professor of Education
Department of Counseling, Educational Psychology and Special Education,
Michigan State University
East Lansing, MI 48824

February 2011

Table of Contents

**Educator Evaluation – Models, Parameters,
Issues and Implementation**
Edward Roeber
Michigan State University

Introduction

One of the major issues currently vexing Michigan educators is how to implement Public Act 205 of 2009, Sections 1249 and 1250, that requires the annual evaluation of teachers and school principals using student growth measures from national, state and local assessments as "a significant factor." These evaluations are to guide improvement efforts of individual educators, result in decisions about promotion, retention, or dismissal, as well as tenure decisions. In addition the districts are to maintain a method of compensation that is related to the job performance and accomplishment of the educator, based in part on the growth in student achievement as measured by state and local achievement measures, as well as other data about teacher and principal performance (Section 1250).

This new system of evaluating and compensating teachers and school administrators is to go into effect when school districts and employee groups next bargain new contracts. Thus, it is a timely issue for a number of Michigan school districts as they work to develop approve new contracts by the start of the next school year, or soon thereafter. The goal of this paper is to identify a few major approaches to the evaluation of educators, some key issues in the use of student assessment data in such evaluations, and to provide some thoughts on models for carrying out such evaluations in a constructive and feasible manner. The goal of this paper, then, is to add to the thoughtful consideration of the substantial measurement and policy issues inherent in complying with this new state law.

Overview of Educator Evaluation

In a previous paper (Roeber, 2010), a model that outlined how educator evaluation might occur was described. This model is an expansion of a model for educator evaluation first proposed by several Michigan professional educator organizations for use in the state's Race to the Top (RTTT) application (Michigan AFT, et. al., 2010). This model is the work of four Michigan professional associations (Michigan AFT, Michigan Education Association, Michigan Association of Secondary School Principals, and the Michigan Elementary and Middle School Principals Association). The paper also reviewed existing student achievement measures currently available or that could be developed for use in Michigan. The goal of the paper was to show how the ideas laid out in the Michigan AFT, et al paper prepared for Michigan's Race to the Top application could be operationalized.

The Roeber model is summarized in Figure 1. In this model, educator improvement goals are set by and for educators. In an ideal implementation of this model, the educator and the educator's supervisor would each select an improvement goal for the educator, both connected to the school's official School Improvement Plan. Then, the employee and the employee's supervisor would jointly develop activities and measures for these goals. These activities would be followed throughout the school year. Near the end of the school year, actual activities engaged in by the individual educator and the educator's accomplishments would be documented. Finally, the educator and the supervisor of the educator would write reflections on the activities and accomplishments. These reflections should focus on the extent to which the activities were engaged by the employee, as well as the extent to which these activities helped the educator to achieve the goals set for her or him and permitted the educator to improve instruction and student learning.

Figure 1
Proposed Professional Practice Portfolio

In the Fall, Develop:
1. Goals for the Individual Educator

    A. Goals from the School Improvement Plan (SIP)– the educator's role in achieving one or more of the SIP goals
    B. Goals for the Individual Educator – the educator's goal(s)
       o   Short-term – This school year
       o   Long-term – Next year and beyond

2. Plans for Growth and Improvement
    A. Plans to help accomplish school/team goals – How will the educator accomplish the school goals within the context of the school improvement team?
    B. Plans to accomplish individual goals – How will the individual teacher accomplish his or her goals?

3. Measures of Performance
    A. State measures where available and applicable
       o   MEAP/MME/MI-Access/ELPA
    B. School measures
       o   School's comprehensive needs assessment
       o   Interim benchmark assessments
       o   Common assessments
    C. Educator-Created Measures
       o   Content organization measures
       o   Individually-collected data
          •   Summative information
          •   Interim benchmark assessments
          •   Formative assessment information

In the Spring, Add:
4. Summary of Activities Used to Accomplish the Plans and Goals
    A. Individual educator achievement of team/school goals – What activities did the educator engage in to accomplish the goals of the school improvement team?
    B. Individual educator goals – What activities did the educator engage in to accomplish the goals he or she set for himself or herself?

5. Evidence of Accomplishment
    A. Team goals – What evidence is there that the selected goal(s) in the School Improvement Plan was accomplished?
    B. Individual goals – What evidence is there that the individual goals were accomplished?
       o   Educator-collected information
       o   Peer information
       o   Supervisor information

6. Reflective Feedback
    A. Individual educator – Looking back on the year, what would the educator have done differently? What does the educator plan for the coming year?
    B. Peers on the team/school – Do the peers of the educator support the evidence of accomplishment as put forth by the individual educator?
    C. Supervisor(s) – Does the supervisor support the evidence of accomplishment as put forth by the individual educator?

This model makes several assumptions:

- This model could be used by classroom teachers, supervisors, instructional support staff, and school leaders at the school, district, and intermediate school district levels, so as to encourage every educator to engage in improvement activities.
- *Every* educator (from newly-minted teachers or school leaders to experienced, veteran educators approaching retirement) should be striving to improve in areas identified both by the educator and the supervisor of the educator.
- This improvement process should be ongoing and assist improvement in all educators (from rookie to veteran teachers).
- Various achievement measures can and should be used, whether these already exist, are already available but yet to be selected, or are yet to be developed.
- The data to be used in Part 5, Evidence of Accomplishment, will vary among educators at each level and position, as there are many different types of assignments of educators employed in our schools.
- The evaluation is based on accomplishing the school's School Improvement Plan (SIP). District and intermediate district educators can base their goals on the support that they provide to schools within their district. This will encourage all educators to work together to achieve (or support the achievement of) the goals the school has set for itself.
- Peers and supervisors can reflect on what the candidate is able to achieve during the year, which could help to demonstrate the accuracy of the accomplishments that the educator documents.
- It presumes that teacher evaluation will be primarily formative in nature, especially in the early years, so that the emphasis is on helping **all** educators improve their practice. It could be used, after a period of time to provide summative judgments as well, so that while educators first have chances to improve, summary judgments could be made or implemented.
- These summative judgments could lead to positive or not so positive consequences for educators such as promotion, dismissal, compensation or bonuses. However, such judgments would be based on a substantial body of evidence collected over more than one school year, so it could provide a sound legal basis for personnel decisions that are made about the educator.
- The reflective piece (Part 6) might be the most important indication to an educator that improvement has not been adequate and that unless further improvement is forthcoming, some type of summative action (such as dismissal) might take place. On the other hand, substantial improvement might be recognized through compensation such as differential pay increases and/or promotions.

This model for evaluation would serve to accomplish the legislative purpose of using achievement information in educator evaluation, doing so in a manner that rather than focusing exclusively on identifying and eliminating a few poorly-performing educators, instead focuses on encouraging *all* educators to improve yet still providing the ability to flag those individuals unable or unwilling to improve.

Two Basic Evaluation Models

Before the evaluation model shown in Figure 1 can be implemented, a fundamental decision needs to be made about the locus of control of such a body of evidence of educator improvement activities. There are two basic models that could be used. Each of these is described below, followed by a more detailed description of the parameters that describe and define how the two models differ. The activities for implementing each model are also provided.

The two basic models for personnel evaluation are:

- Inspection Model – This is the traditional personnel evaluation model, most often used with new educators in their first two or three years of employment, to determine eligibility for tenure. In this model, a person or persons familiar with the work of the individual educator conducts the evaluation. This is usually the immediate supervisor of the educator or someone who is designated by the immediate supervisor, so if a teacher is being evaluated, it may be the building principal, assistant principal, or a departmental chair. The key questions to be answered include:

  o What level of proficiency does this individual educator possess?
  o Is this individual a competent educator, or at least on track to become one?

  Supervisors may observe the individual educator in action in the classroom or other settings and use various formal and informal rating forms to record this information, collect written evidence of educator competence such as lesson plans, gather test scores, and may even include surveys of others (e.g., peers, parents, and/or students) for their perceptions.

  Then, the supervisor rates the individual, making a determination of the level of proficiency of the educator, based on this collective body of information. In some evaluation systems, certain pieces of information (e.g., classroom observation or test scores) might be officially given a certain weight in the process (e.g., 50% of the overall evaluation), while in other cases this is left to the supervisor conducting the evaluation (and may differ among supervisors and/or educators being evaluated). In a number of states that received Race to the Top funding, this data collected locally is weighted at about 50% of the overall evaluation (Roeber, 2011).

  In the past, such evaluations did not usually include student achievement test scores or growth in student achievement in a formal manner. This is changing, however, due both to state Race to the Top application requirements and as a result, state legislation and/or state education policy changes. Almost all of the states that won Race to the Top funding set student achievement (variously measured and reported) as exactly or about 50% of the evaluation of teachers.

  Currently, such evaluations may be conducted annually only for probationary educators, while those with tenure or other form of permanency may be formally evaluated only once every two or three years, or even less often. One issue in expanding this "inspection model" to all educators on an annual basis (as called for in the Michigan legislation) is the capacity for supervisors such as secondary principals to adequately evaluate all of their instructional and specialist staff each year.

There are several advantages to the Inspection Model:

1. Each person such as a teacher is judged in a standard manner, which some would say makes this model "fair" to all such individuals. Thus, if legally challenged, a strong case can be made that all comparable educators went through the same process, using the same tools, implemented in the same manner.

2. The formal external evaluation model requires educators to provide evidence of their competence through the actual demonstration of their instructional prowess and/or school leadership. It is not just "talk the talk," but "walk the walk."

3. Supervisors can be taught to judge the competence of their subordinates. Such training would not only benefit the educator evaluation process, it would also assist the supervisor in the performance of their duties on an on-going basis.

There are several disadvantages to the Inspection Model:

1. By being "standard," the model may not be equally applicable to all persons in a particular job such as "teacher," because all such individuals are not the same. For example, a kindergarten teacher and a fifth grade teacher might not have comparable achievement data available for use in their evaluation.

2. The criteria for conducting the "inspection" may not be well documented, so that different supervisors might operationalize the criteria differently. For example, "effective class management" might be interpreted by one supervisor as 'a neat, orderly and quiet classroom,' while another might indicate that 'a classroom where students are actively engaged (and thus a bit noisy) in their own learning in small groups' is the best evidence of effective teaching.

3. Persons conducting the external evaluation may not be well trained to conduct the evaluation. If this is the case, then two supervisors might observe or rate the same educator completely differently, perhaps even differing in their overall judgment of proficiency. This is a real area of weakness in the current use of this model in the state.

4. Annually evaluating every educator in large schools may be very challenging to impossible, given that one or two individuals might be responsible for the annual evaluation of fifty, sixty or more classroom teachers, using methods that require observation of each teacher on two more occasions throughout the school year.

5. Some of the persons involved in the evaluation may not be neutral in their orientation to the process. For example, parents and students may "like" or "dislike" a teacher for reasons not directly related to the evaluation and thus bias their reviews of the educator. A principal might like a classroom teacher based on their personal interactions, not objective evidence of performance collected during the evaluation process.

- Demonstration Model – The second model is one in which each individual educator is charged with the responsibility of demonstrating her or his _own_ proficiency by collecting and organizing evidence of her or his own proficiency. This collection of evidence, along with reflections on the level of proficiency, is reviewed and agreed with/disagreed with by the supervisor. In this model, a body of evidence or portfolio with no evidence in it constitutes a rating of "not proficient." In this model, the educator prepares, organizes and presents evidence to support this statement: 'I am a proficient educator who can produce effective results. I am taking steps to improve my competence. Here is my proof of both assertion….' Thus, each educator needs to select appropriate evidence – whether supplied by themselves, by peers or supervisors, or located from other sources.

  Individual goals are developed by each educator (and her or his supervisor) who is then tasked with suggesting in advance the steps to be completed to accomplish the goals and, with supervisor approval, the types of evidence to be used to demonstrate learning and proficiency.

  The educator being evaluated is also responsible for the collection of evidence about their proficiency. Multiple sources of information are best, so the educator will need to consider just how they can demonstrate their skills, especially when it comes to student learning, which should be a key goal for every educator. Of course, the evidence of student learning will need to be collected within the context of each educator's job responsibilities and the mutually-set goals.

  A variety of tools can be provided to support the implementation of this model, so data used might include collection of some of the same types of evidence as in the Inspection Model, but the primary difference is locus of control – for example, the educator requests that an observation form or rating form be completed by someone who knows the educator and/or has seen the educator in an instructional setting. The evidence provided could be written or video, since videos showing the educator working with individual students, small groups of students, or the entire classroom could be effective means of showing her or his teaching skills.

  This model is akin to the manner in which pre-service educators compile evidence of their proficiency, showing coursework and grades, skill in teaching individual students and groups of students, and relevant assessment information. Newly-minted teachers often use their portfolios as part of the hiring process, so individuals involved in hiring decisions, such as building principals, already have some experience examining such collections and using them to make decisions. Thus, recently certified educators are already used to a system such as this. More experienced educators may need to learn how to best document their proficiency and use the tools to do so.

The advantages of the Demonstration Model are:

1. This model works equally well for teachers and school leaders. Any educator can be tasked with demonstrating their skills and proficiency within the context of their current job responsibilities.

2. This model motivates educators to demonstrate their own proficiency. They will internalize the criteria for what constitutes "good teaching" within their own job responsibilities.

3. Because this model requires each educator to be able to show their competence, the educators will need to constantly be thinking about what types of evidence they should be gathering, striving to answer the key question of "how can I show that I am an effective teacher or an effective principal?"

4. Evidence collection will need to be on-going, so that the educator is considering competency demonstration throughout the school year (and summer break), not just the one or two times a year when observed by a supervisor. This helps to keep educators focused on competency considerations during the entire school year (and beyond).

5. The building principal does not have to conduct so many observational visits to classrooms. Thus, a principal would be able to conduct annual evaluations of a larger number of staff than if the principal had to visit each teachers' classrooms two or more times each year. Plus, these observations will be less summative and more formative, serving to collect data for teachers to use in the teacher demonstration of proficiency.

6. This model can be customized to the unique job responsibilities of specific educators. This will permit an elementary teacher who serves on the school improvement team to indicate his or her goals for this work, while another teacher on a curriculum review committee can state goals for improvement in this area. It would also work well for educators in non-state tested grades and content areas, as well as educational specialists and supervisors in schools.

The disadvantages of the Demonstration Model are:

1. The public, parents and local and state policy makers (i.e., the local school board or state legislators) may not trust the types of evidence provided by educators. However, the review and concurrence of the supervisor can serve to assure accuracy and rigor of these evaluations.

2. Some educators that participate will need to learn how to document their performance.

3. Each educator's collection of evidence is different from each other. Will it be possible for these disparate sets of evidence to be judged using common rubrics and criteria?

4. Supervisors currently may not be prepared to evaluate such a broad and disparate sets of evidence. Educators (and others) who review the collections of evidence will need to be provided good examples of the different levels of proficiency and will need to be trained so as to demonstrate their proficiency in judging the collections.

5. If an educator submits a skimpy or thinly documented collection of evidence, is this truly a demonstration of lack of competence or just inability to collect good evidence?

Parameters for Educator Evaluation Systems

From an examination of the two models presented above, along with their advantages and disadvantages, several parameters along which such systems vary can be identified. These serve as key decision points in determining important aspects of educator evaluation systems no matter which model is selected. These choices can assist districts and their employee groups to better understand the range of options available to them, and hopefully, make more informed choices about the evaluation design and system to be used in each district.

The parameters for evaluation systems for educators include the following dimensions:

1. Purpose – What is the primary purpose of the evaluation system? There are two basic choices, of course, with variations on each:

   • Identify low performing educators – The purpose of identifying ineffective teachers and school administrators is to identify low performing educators who, if they are unable to show improvement, can be dismissed such from their current employment. Focusing on identifying and eliminating low performing educators is one of the key elements suggested by Public Act 205 of 2009. However, if the focus is entirely on this purpose, improvements needed by other educators (currently not rated as "ineffective") may not occur because of the perception by educators that identifying themselves as in need of improvement or needing to learn new skills may jeopardize their employment in the future.

   • Encourage improvement in all educators – The purpose of encouraging improvement in all educators is a recognition that educators need to work to stay current in their field, to learn new skills due to changes in the profession and our understanding of how students learn. This purpose recognizes that education, like other professions, is always changing and that educators need focused professional development in order to remain current with changes in the profession. By focusing evaluation efforts on identifying need for improvement and efforts engaged in to make the needed improvement, potentially every educator can be encouraged to seek to improve their proficiency. While such activities are needed by all educators, there is no certainty that just any professional development activity will help low performing educators improve enough to no longer be considered low performing. An educator evaluation system can encourage the use of best professional development activities that lead to increased proficiency of each educator.

2. Individual or Group Evaluation? – Will evaluation efforts be focused on individual teachers and school administrators or on school faculties as a whole?

   • Evaluate individual educators – The goal in evaluating individual educators is to determine how effective each educator is without reference to the effectiveness of other educators in the school or district. This would most clearly identify educators who do not produce the desired student results. It may also encourage, however, educators to "compete" with one another for high performance ratings, thus reducing school team efforts designed to improve instruction school-wide. Not every gain in student learning is the

result of individual teacher effort, so if the system sets one educator against others in a "zero-sum" game, collaborative efforts among educators may be harmed.

- Evaluate groups of educators (e.g., school faculties) – The purpose in evaluating teams of educators is a recognition that student achievement and performance is the result of student learning across multiple grades and courses, taught by multiple teachers, working together as a school faculty. Focusing on school teams also encourages educators to work together to improve student learning and bring about the changes needed to accomplish the goals in the school's School Improvement Plan.

3. Nature of the Evaluation – The evaluation of educators could focus on their current level of proficiency, or it could focus on how much the educator has improved his or her performance from one time to another. In essence, this turns the evaluation into either a "status" or a "progress" model.

   - Current Status – In this sort of evaluation, a series of criteria and measures are defined, and each educator is judged against these criteria. For example, a couple of states that won RTTT funding (Roeber, 2011) have chosen to use the work of Charlotte Danielson as the basis for judging the proficiency of teachers. Achievement results can also be used and often, these are – ironically – *changes in student performance* that are the achievement judgments made in judging an educator's Current Status. These criteria would be used to indicate how well each teacher is able to carry out each of these criteria.

   - Progress Made – In this type of evaluation, the proficiency of an educator is measured at two or more times, at least at the start and the end of the evaluation period. It is her or his *change in performance* that becomes the basis for the evaluation. In the Inspection Model, any deficiencies identified by the supervisor will be used as the basis for the demonstration of change. These judgments could also be against criteria such as those set forth by Danielson, but in this case, will serve to guide the pre- and post-improvements judgments that are the key to Progress Model.

     In the Demonstration Model, the focus will be on the performance in the goal areas identified by the individual educator and his or her supervisor. Presumably, in this model, educators will perform at a less than proficient level at the outset, and if success in carrying the improvement activities specified, will be able to show improved performance at a subsequent or later times.

4. Locus of Control – Who is responsible for the conduct of the evaluation? In traditional evaluations, the supervisor provides the evaluation for each sub-ordinate. Yet, in many real-life situations (e.g., job interviews), it is up to the employee to demonstrate his or her own competency and suitability for a job. Therefore, is the evaluation done by the employee or done to the employee?

   The two models, and their advantages and disadvantages, were described above.

5. Evaluation Methods – What method(s) will be used to evaluate educators?

A recent report (Hinchey, 2010) reviews the several tools that can be used in the evaluation of educators. Attachment A shows a summary description of each measure, a summary of the research on the use of that evaluation method, its strengths and its cautions. This chart, from the Hinchey report, is very useful in showing what methods can be used and ideas about how they can best be used.

The different evaluation methods reviewed by Hinchey are as follows (the page numbers shown in parentheses indicate the page number in the Hinckey report where the information cited can be found):

- Classroom Observation – One way to determine how effective an educator is to observe the educator in action. This method permits the educator to be "observed in the classroom, so that specific teacher practices, holistic aspects of instruction, and interaction between teachers and students" can be observed (p 27).

- Principal Evaluation – This is a structured or unstructured observation of the teacher by the building principal or his or her designee. This is generally done for "summative purposes, most commonly for tenure or dismissal decisions for beginning teachers" (p. 27).

- Instructional Artifacts – The artifacts used may include " lesson plans, teacher assignments, assessments, scoring rubrics, and student work" (p. 28). Typically, standardized procedures and forms are used to evaluate these artifacts. Limited evidence of the effectiveness of the use of the evaluation of the artifacts exists.

- Portfolio – These collections of evidence are used to "document a large range of teaching behaviors and responsibilities" and are "widely used …for assessing the performance of teacher candidates and beginning teachers" (p. 28).

- Teacher Self-Report Measure – In this method, "teachers report what they are doing in the classroom" and "may be assessed through surveys, instructional logs, and interviews" (p. 29).

- Student Survey – These surveys are "used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by students" (p. 29).

- Parent/Guardian Survey – These surveys are used to gather parental perceptions of parents or guardians about the quality of education provided to their child, the interactions of the teacher with the parent or guardian, and the level of satisfaction of the parent/guardian.

- Value-Added Measure – This is a statistical measure of the changes in achievement of students "used to determine teachers' contributions to students' test score gains" (p. 30). Some of these measures be based on actual test score gains, or predicted versus actual gains, or other models.

6. What type(s) of student achievement information will be used in the evaluation of educators?

In a previous paper (Roeber, 2010), the variety of assessment measures was identified. These include:

- Statewide Assessments – There are variety of summative assessments required of students in each state. These include the state's general education assessments, assessments for students with significant disabilities (the alternate assessment of alternate achievement standards – AA-AAS), measures of English language proficiency for English language learners, and other state-administered assessments.

- Locally-Developed/Selected Common Assessments – These assessments are comparable to state assessments in that they are summative in nature, but they are developed or selected locally. An example of this type of assessment is an end-of-course test or end-of-grade test.

- Interim/Benchmark Assessments – These are periodic, short-cycle summative assessments used during the school year, either at pre-determined intervals, such as quarterly or at the conclusion of instructional units. These assessments might be used to determine student achievement in order to grade the student (at the conclusion of each unit of instruction) or predict performance on the state or local summative assessments (so as to intervene to help students predicted to do poorly receive the assistance in order to improve their performance).

- Standardized Achievement Tests – These are commercially available achievement tests that typically cover a variety of subject areas from kindergarten through grade 12. One reason that they might be used is because scores on these tests are expressed on a common scale so that "value-added" measures can be based on such scores.

- Classroom-Based Assessments – These are measures developed and used by individual teachers. They may be formal tests comprised of one or more item types or may be less formal assessment methods such as observation or personal communications.

7. Standard versus Idiosyncratic Information – Are the same procedures, forms and processes used for all comparable educators (e.g., classroom teachers) or are different methods, materials and procedures?

There are a couple of choices that can be made:

- Provide standardized, comparable information on the performance of every educator, so that supervisors only need to learn one evaluation system and tools and use these with all sub-ordinates.

- Select from the variety of tools and procedures those that are most appropriate and applicable for each individual (or group of individuals) being evaluated. This will serve to customize the outcomes and the measures of the outcomes (student achievement, other student outcomes, and other data) for each educator or group of educators.

7. Use of evaluation results – How should the evaluation results be used? Are multiple uses of the same information appropriate?

- Formative Evaluation (Educator Improvement) – This purpose means that the results are used primarily to show areas where educator improvement is needed and then to indicate the success of these improvement efforts.

- Retention/Tenure decisions – This use means that the results are used to determine which educators are retained (or released), granted provisional certification, granted tenure, and/or promoted (such as promoted from classroom teacher to teacher leader).

- Compensation decisions – One goal of legislation on the use of student achievement in decisions about how educators are compensated. Thus, the evaluation results might be used to determine changes in compensation and/or bonus pay.

- Supervisor professional development – Summaries of the evaluations can serve to indicate areas in which supervisors participating in the evaluations need to learn new skills and improve their abilities to conduct the evaluations. The evaluation results can also be used eventually to gauge the success of these improvement efforts and might well be used in the evaluation of the supervisors such as building principals by his or her superiors.

Issues to Consider in Implementing Each Model

Regardless of the model used, there are several issues that need to be resolved in order to implement either one of the two basic evaluation models. Depending on where the control of the evaluation process resides (currently, P.A. 205 places the responsibility for determining the evaluation squarely on local school districts, although proposals to mandate a statewide system have been introduce in the Michigan legislature in the past. This is a model prevalent in a number of states.) Thus, currently, agreement on these basic issues will be necessary at local level, although this may change if the legislature decides to do so.

These major design issues, adapted from a list prepared by Saginaw ISD Committee on Teacher and Administrator Evaluation (2011), include the following:

- What is definition of "effective teaching?" What is "effective school leadership?" How can each be operationally defined? It is essential regardless of the evaluation model used (Inspection or Demonstration) for the definition of effective teaching and school leadership to be determined. There are many such definitions, yet without such a common definition or set of definitions, it will be impossible for supervisors to evaluate subordinates or for educators to demonstrate their proficiency in a manner fair to all employees. Factors to be considered include teaching skill, planning, delivery of instruction, motivation of students, assessment skill, care and concern about student learning, and so forth. Some of these are tangible attributes that can be observed readily, while others are less obvious. Stating what constitutes good teaching or good school leadership is the first step in determining how to measure these characteristics.

- What measures of teaching and leadership effectiveness already exist? Can these measures be used in the model to be used? If so, how? There are several systems of evaluation that are available for use. These include measures created by Marzano, Danielson, and others. What parts of these systems can be used? Of course, a determination will need to be made about whether such measures can be used selectively, or whether they must be adopted wholesale (with attendant costs). Or, will new measures, perhaps patterned after measures available commercially or in the research literature surrounding good teaching and leadership, be created for use. Using existing measures has the advantage of fast implementation but perhaps at higher cost. While creating new measures may save on implementation dollars, it will be more costly in the short range and presumes that high quality measures can be readily created.

- How can educators who are responsible for implementing the system (teachers, school leaders and their supervisors) be adequately prepared to carry out the duties required by the evaluation system? One essential feature of the system envisioned by P.A. 205 is that it will help every educator improve their performance. For this to occur, it will be necessary that the information collected in the evaluation system be of high quality. Since personnel evaluation is not a skill well known by school leaders or teachers, it will be essential regardless of the model adopted for the key participants to be fully prepared to implement the system. In the Inspection model, this is school leaders who evaluate teachers. In the Demonstration model, every educator needs to understand what high quality teaching and school

leadership looks like. This means being able, ultimately, to recognize good teaching and school leadership when it is presented to them (e.g., through observation, videotape or other means).

- Determine the manner in which student achievement can be measured, especially in non-state tested grades. If this personnel evaluation system is to be fair to all educators, it needs to be equal in its application to all educators, and the information used in the evaluation has to have comparable relevance to each educator. This means that careful consideration will need to be given to the non-MEAP and non-MME assessed grades and subject areas. What achievement measures will be used in each grade? In every content area? Will building results for a couple of grades and subjects be used for all educators in the building – is this fair? Is it fair not to do so? The determination of which measures will be used and how the data is collected and used will be a major issue to be resolved.

- Define how growth will be measured and reported; will a statistical model be used and if so, which one? The idea of "growth" is a seductively simple one – how much did students "grow" academically from one year to another? Is the extent of this change adequate or not? What makes these questions challenging is that the content of instruction varies substantially from one year to another, so that it is a bit of an "apples and oranges" comparison. The tests used, therefore, also differ from one year to another. Second, complete data is not available in many cases in schools currently. For example, there is no data available in any two adjacent grades in any content area at the high school level. How can change be reported when only one grade (grade 11) is assessed? While there is adjacent-grade data available in mathematics and English language arts at the middle school and late elementary levels, even these data do not cover all grades or subject areas. Is it fair to use such data for some but not all educators?

  In addition, what statistical procedures, if any, will be used to report changes in student performance from year to year? Currently, there are fairly simple procedures for reporting changes in student performance, such as value tables, there are more complex statistical procedures such as percentile change, and highly statistical procedures such as 'value-added' measures. If statistical procedures are adopted, will any local educator understand on what bases they were evaluated? If they don't, is it fair to use such information to evaluate them?

- What non-achievement measures will be used? Almost every evaluation system uses non-achievement measures for personnel evaluation. Such measures can be called for if the definitions of effective teaching and effective school leadership include such factors; most evaluation systems place great emphasis on such measures, in part because the use of student achievement in personnel evaluation is a relatively new phenomenon. As with the overall system discussed above, implementers of the evaluation systems locally will need to determine whether to use existing non-achievement measures (for example, they may be built into an evaluation system available commercially) or to create measures from scratch. As discussed above, there are advantages and challenges in each approach.

- How will educators receive feedback? How will continuous growth be measured? Can it be assured? The evaluation system suggested in P.A. 205 indicates that educators will be provided feedback so as to be encouraged and incented to seek to make improvements. This means that they must receive feedback, they need to receive this on some sort of on-going basis, and it has to be feedback that help them uncover areas where professional growth is needed and help them monitor whether the activities they engage in are helping make the needed changes. This suggests a system of feedback that is actionable – specific enough to indicate areas of need and comprehensive enough to cover the areas of good teaching and school leadership contained in the definitions of each.

- What improvement opportunities will be made available – locally and through out the state? Because the evaluation system called for by P.A. 205 needs to be designed to assist educators to improve, opportunities will need to be provided to educators to be able to take the actionable feedback and make needed changes. As currently written into P.A. 205, the system of evaluation will be determined locally (or, perhaps, regionally). As a result, it will be necessary for the professional development activities for teachers and school leaders to be available locally or regionally as well. This, of course, does not preclude the Michigan Department of Education or a statewide professional organization from offering professional development activities. The regional provision of professional development will help assure that needed assistance is available to each educator needing assistance without unnecessary and duplicative services being provided across the state (or, worse, for needed services to be unavailable). The availability of assistance to educators seeking to improve will be essential if the evaluation results will be used to terminate employees, since the employer will need to be able to demonstrate that reasonable efforts were made to assist educators to improve in areas of perceived weakness.

- How will the data be used to promote, certify, advance to tenure educators? This evaluation system is all encompassing, which means that the results will be used to make important decisions about teachers and school leaders. It will be essential that the manner in which the results will be used is determined in advance so that educators being evaluated may be given advance notice of this. What promotional opportunities will be provided to teachers? To school leaders? How will these decisions be made? Will a process of probationary teaching status lead to a certified status, and if so, by what processes, using what criteria? Will "tenure" still be possible, and if so, by what means? What will tenure mean if the educator continues to be reviewed (annually or perhaps less often, such as every second or third year) and could be dismissed?

- How will the system be used to dismiss ineffective tenured or untenured teachers or school leaders if they fail to show needed improvement? In this evaluation system, as required by P.A. 205, ineffective teachers (whether untenured or tenured) and school leaders can be dismissed. Attention will need to be paid to the quality of the information used to make the judgments of competence, as well as the processes engaged in. It will be essential to show that educators so judged were given information about their status, provided encouragement and opportunity for improvement through professional development, coaching, and/or other means, and still failed to

show sufficient improvement to warrant retention. One key issue to be addressed is the relative role of student achievement and non-student achievement measures in making these decisions, especially if these two types of measures do not agree. For example, if a teacher is judged to be proficient through observation, review of written lesson plans, contributes effectively to the school improvement activities, but has students who do not demonstrate change in academic achievement, is this a teacher who should be dismissed? It is thorny issues such as this that will need resolution.

• How will the compensation of educators (both teachers and school leaders) be tied to their evaluations? An important reason for the adoption of P.A. 205 and similar legislation is to move public education away from a system in which teachers are compensated by their level of education and years of teaching experience. However, in doing so, there may be a trade off of an objective (some would argue, however, irrelevant) system of compensation to one that is pegged to the level of performance of the educator. Such a system may merely provide bonuses to highly effective teachers and school leaders, or it may be a complete revision of the compensation system. Local school systems will need to reach agreement on how the evaluation results will be tied to the compensation of its employees. While P.A. 205 deals specifically with classroom teachers and school leaders, there are other school employees such as counselors, specialists, and aides not specifically mentioned by law that will also have to be covered in the design of the system of compensation.

Suggested Implementation Processes

All of these design decisions, issues and ideas, plus others that will arise as these discussions occur, will need to be made, and these decisions will need to be implemented within the model chosen. While some decisions will be the same, such as the achievement measure to be used by a fourth grade teacher to demonstrate the level of student achievement, the differences in the locus of control between the two models will affect how even this measure is used. For example, the MEAP performance may be an adequate measure of fourth grade teacher performance in the Inspection model, but may need to be supplemented by a variety of other measures in the Demonstration model. For example, the teacher might add interesting additional artifacts of student achievement such as several video clips of student-led conferences where they demonstrate how much they have learned in mathematics in fourth grade. Each of the models could be implemented in a local school system, although the processes of doing so will vary somewhat between the two models.

Some suggestions about how implementation could occur are provided in the following section. The initial steps for each model are quite similar (but not exactly the same). The manner in which the two models are implemented is a key

Inspection Model – This model might be implemented in the following manner:

1. A policy entity, such as the local school board or administration with the assistance of employees groups, needs to define the parameters of the model. Alternatively, the state board of education, state superintendent, state department of education and/or the state legislature might develop one or more models, which might be offered as models for local work. For example, if student achievement is to be included in the educator evaluation system, which actual measures will be used (from state assessment programs as well as locally-selected or developed assessments that are permitted)? Will student achievement be a pre-set percentage of the overall evaluation or is this a decision to be left to each school or supervisor?

   What other variables will be included in the evaluation system and how will these variables be measured or operationalized?

   What weight will be given to each of these variables in the model? Will standard weights be applied statewide or will this be left to the discretion of the local school systems?

2. A policy advisory committee, comprised of stakeholders in the evaluation process (e.g., teachers, school leaders, local and state policy makers and experts in educator evaluation, parents, business leaders) should assist the district in the development of resources and the implementation of the model.

3. A working group within the school system (or within the state education agency) should be tasked with developing the materials and procedures to be used in educator evaluation. This should include planning for the needed training in educator evaluation to assure adequate implementation of judgmental processes that will be used in this evaluation model. Examples include a) rating procedures and rating forms need to be developed (will existing ones or newly-created ones be used?), and b) the training that will need to be provided to school leaders to

be able to recognize different levels of teacher performance. This training may include actual samples of teaching, as well as training school leaders to classify different levels of performance correctly.

4. The district should provide training to local educators throughout the district (or, in other cases, this could be done by the state education agency regionally across the state). This might be live training presented in person, training provided through videoconferences (live or taped), or online training webinars. The goal is to assure that all persons who will judge the performance of teachers (whether school leaders or peer teachers) be trained to do so before they render any judgments of subordinates or peers in their buildings.

5. When the system is ready to start, it may be advisable to pilot it first with volunteer educators from the elementary, middle school and high school levels. This will permit any issues in implementation to be identified and addressed, before a district-wide application.

6. When the system begins, educators would be informed of the district's definition of effective teaching. The evaluation methods (e.g., observation) and materials (e.g., a principal rating form) would be provided to educators, explained as to their use, and then implemented.

7. Data would be collected on each educator by the supervisor, using in some cases information provided by the educator herself or himself (e.g., written lesson plans). Other data (e.g., test scores for the teacher) might come from external sources. The supervisor would be responsible for assembly of this information, and then using it to make formative and eventually, summative judgments about the subordinate.

8. The supervisor ratings and supporting documentation would be stored in some fashion to be referred to in future years as comparable data is collected so that changes in performance can also be reviewed and made part of the summative judgments of educator performance.

9. The district should monitor the application of its materials and procedures in order to determine where additional professional development is needed or if there are instances of where procedures or materials are not being implemented in a sound manner. The purpose of this quality control effort is to continue to make sure that educators are being properly evaluated and if not, that appropriate action is taken to assure that they are.

10. The district should use the monitoring information and other feedback mechanisms to gauge which parts of the system need to be worked on and improved in subsequent years. This needs to be viewed as an on-going effort that will take several years to adequately implement. Data collected from users (supervisors and subordinates) should be used to help to improve the system.

Demonstration Model – This model could be implemented in the following manner. As noted above, the manner in which the two models are developed begins in a very similar fashion, but then diverges.

1. As mentioned above, a policy entity, such as the local school board or administration with the assistance of employees groups, needs to define the

parameters of the model. Alternatively, the state board of education, state superintendent, state department of education and/or the state legislature might develop one or more models, which might be offered as models for local work. For example, if student achievement is to be included in the educator evaluation system, which actual measures will be used (from state assessment programs as well as locally-selected or developed assessments that are permitted)? Will student achievement be a pre-set percentage of the overall evaluation or is this a decision to be left to each school or supervisor? Will individual educators be able to use achievement measures relevant to the goals they select for improvement?

What other variables will be included in the evaluation system and how will these variables be measured or operationalized? Is this a standard set of measures or can individual educators "customize" these to fit with their improvement goals?

What weight will be given to each of these variables in the model? Will standard weights be applied statewide or will this be left to the discretion of the local school systems? Can different weights be applied to different educators within the same school or teaching the same grade or content areas?

2. A policy advisory committee, comprised of stakeholders in the evaluation process (e.g., teachers, school leaders, local and state policy makers and experts in educator evaluation, parents, business leaders) should assist the district in the development of resources and the implementation of the model.

3. A working group within the school system should be tasked with developing the materials and procedures to be used in educator evaluation. This should include planning for the needed training in educator evaluation to assure adequate implementation of judgmental processes that will be used in this evaluation model. Examples include a) rating procedures and rating forms need to be developed (will existing ones or newly-created ones be used?), b) training for participants in the evaluation system to show them how to document their performance as an educator, and c) the training that will need to be provided to school leaders to be able to recognize different levels of teacher performance.

    Part of this is the process of helping educators learn how to document their performance, what levels of performance are adequate and exemplary, and suggestions for how educators can demonstrate these levels (i.e., tools they could use, methods to use these tools, and so forth).

    This training may include actual samples of teaching, as well as training school leaders to classify different levels of performance correctly based on sample collections of evidence provided by subordinates.

4. The district should provide training to local educators throughout the district. Two types of training will be necessary in this model. The first might be designed for educators who will be evaluated. This professional development will introduce the standards for effective practice, the development of constructive improvement goals tied to school improvement plans, the manner in which these goals might be accomplished, as well as measures of the achievement of the improvement goals. Since this is a system where individual educators will have different goals from one another (yet tied to the school improvement plan for the school), it will be important to show educators the range of ways that they could assemble a body of evidence about their effectiveness. For recently certified teachers, this

may be relatively easy to understand but for more experience educators who never experienced this in pre-service preparation programs, more information and ideas will be necessary.

The second type of training will need directed at the individuals (i.e., school leaders or district leaders) who will be responsible for reviewing the collections of evidence from subordinates and helping to decide what level of proficiency these collections indicate. Because this is not an activity that is already taking place, it will be important to develop written descriptions of each level of performance (e.g., exemplary), rubrics for making this judgment based on the variety of types of evidence that might be submitted (i.e., ratings, video clips, surveys, samples of student work, lesson plans, etc.), training in how to use these materials, and using samples of educator documentation, learning to actually score collections of evidence of educator effectiveness. This should help supervisors to be able to make adequate judgments of what may well be very diverse evidence collections.

Each type of training might be presented live and in person, provided through videoconferences (live or taped), or presented through online training webinars or modules. Sample exemplars of collection of evidence at each level of proficiency should be posted and made available to educators and the public, both to improve the quality of the evidence submitted, but also to make the system as transparent as possible, The goal is to assure that all persons who will judge the performance of teachers (whether school leaders or peer teachers) be trained to do so before they render any judgments of subordinates or peers in their buildings.

5. When the system is ready to start, it may be advisable to pilot it first with volunteer educators from the elementary, middle school and high school levels. This will permit any issues in implementation to be identified and addressed, before a district-wide application.

6. When the system begins, educators would be informed of the district's definition of effective teaching. A description of the evaluation model, the rubrics, and evaluation methods and any resource materials that have been created for potential educator use should be provided to educators, with explanations as to their use.

   The educator to be evaluated should select one personal improvement goal; the educator's supervisor might select a second improvement goal. Ideally, these goals should be related to the school's School Improvement Plan. The two (the educator and his or her supervisor) should agree on improvement activities and measures of what would constitute success.

   The educator to be evaluated should identify a set of activities and the schedule for implementing these improvement activities that will be engaged in to achieve the two goals. These could include both short-term and longer-term activities. Some might help the educator achieve the needed improvement in a few months while others might require a full school year or more for success to occur. Hence, not only should the activities be laid out, but the schedule for implementing them should also be identified.

7. Then educators should begin to collect evidence of their effectiveness. The data to be collected for each educator would be determined in whole or in part by the

individual educator, although this could include observation and ratings by the supervisor of the educator or peers of that educator, using standard rating forms (as in the Inspection model). In many cases the collection of evidence will include information provided by the educator herself or himself (e.g., written lesson plans), plus other data (e.g., test scores for the teacher) that might come from external sources. The educator being evaluated (not her or his supervisor) would be responsible for assembly of this information.

It is up to the educator to collect evidence of participation in the planned activities and to obtain measureable outcome data, and then to prepare a summary of their success in accomplishing the two goals as well as what subsequent improvement activity might be desirable or necessary. This summary should include a reflection on the learning goals, activities, and accomplishments. All of this collection of evidence might be electronic (consisting of scanned written materials – from students or educators – as well as photographs, videos, and so forth). Using electronic means to store the evidence might permit a broader range of evidence to be collected and used.

8. Once assembled, the educator being evaluated should reflect on the contents of their collection of evidence. This reflection should include the goals selected for improvement, the anticipated and actual activities carried out to make improvement, the anticipated and actual data collected on effectiveness, what additional improvements are needed and how these improvements could be accomplished. Finally, the educator should state what he or she believes is her or his level of effectiveness and state the evidence in the collection of evidence that supports this summative judgment.

9. Then, the supervisor should be provided the collection of evidence from the educator, and the supervisor should review the evidence carefully. There are three key aspects to this review, namely whether the evidence provided is:

    a. *Accurate* (i.e., 'do I agree with the evidence provided?'),
    b. *Adequate* (i.e., is the evidence complete?) and
    c. Does the supervisor agree with the subordinate?

If the answer to any of the questions is "no," then the supervisor should provide the rationale for this judgment along with supporting evidence.

10. The subordinate and the supervisor should conference about the collection, their reflections on it, and their summative judgments about the educator's level of proficiency. Disagreements should be discussed and documented, since these could feed into improvement goals and activities for the following year.

11. The collection of evidence, along with educator and supervisor reflections and supporting documentation should be stored in some fashion to be referred to in future years as comparable data is collected by the educator so that changes in performance can also be reviewed and made part of the summative judgments of educator performance.

12. The district should monitor the application of its materials and procedures in order to determine where additional professional development is needed or if there are instances of where procedures or materials are not being implemented in a sound manner. The purpose of this quality control effort is to continue to

make sure that educators are properly evaluating themselves, that supervisors are accurately reviewing and rating these collections of evidence, and if not, that appropriate actions are taken to assure that they are.

13. The district should use the monitoring information and other feedback mechanisms to gauge which parts of the system need to be worked on and improved in subsequent years. This needs to be viewed as an on-going effort that will take several years to adequately implement. Data collected from users (supervisors and subordinates) should be used to help to improve the system.

The district should pull a small sample of educator collections of evidence and verify the scoring that was carried out in schools. If the level of agreement is below pre-set standards, the district will intervene and require all collections of evidence to be externally scored by a second set of educators from other schools (or even districts) in order to better gauge the actual performance of educators.

Summary

This paper laid out two basic models for educator evaluation: 1) an Inspection model, which is the traditional manner in which personnel evaluation occurs, and 2) a Demonstration model, which places the burden of demonstrating competency on the educator being evaluated, with review and concurrence by that individual's supervisor. A number of decision parameters were provided for school systems to consider as they develop the plans for an educator evaluation system, and implement the system in their district. For each model, suggestions were given on how the model could be implemented. It is the goal of this paper that systems carefully consider the decision parameters and the models as they determine how best to respond to the requirements of Public Act 205 of 2009.

References

AFT-Michigan, MEA, MASSP, and MEMSPA. "A framework for Michigan educator evaluations. Detroit, MI: Author (2010).

Danielson, Charlotte. "Assessing teacher effectiveness." Powerpoint presentation. (2010).

Heitlin, Liana. "Value added: Does it have merit?" Retrieved on from http://www.edweek.org/tm/articles/2010/12/06/valueadded.html on December 6, 2010.

Hinchey, Patricia H. "Getting Teacher Assessment Right – What Policymakers Can Learn from Research." Boulder, CO: National Education Policy Center (2010).

Pallas, Aaron M. "Measuring what matters." Phi Delta Kappan. December 2010/January 2011. Volume 92 Number 4, p. 68-71.

Roeber, Edward D. "Using Tests to Evaluate Teachers." East Lansing, MI: Michigan Education Association (2010).

Roeber, Edward D. "Attributes of Winning RTTP Applications." East Lansing, MI: Michigan State University.

Saginaw ISD Committee on Teacher and Administrator Evaluation. *Untitled document.* E-Mail from Kathy Dewsbury-White, Ingham ISD on February 6, 2011. Saginaw, MI: Author.

Sawchuk, Stephen. "Another scholarly take on teacher value-added. Retrieved on from http://blogs.edweek.org/edweek/teacherbeat/2011/11/another_scholar on November 21, 2010.

United States Department of Education. "Partners in reform." Address by the Secretary of Education to the National Education Association on July 2, 2009. Washington, DC: Author.

Attachment A – Brief Summaries of Teacher Evaluation Methods

The following table is excerpted from the Hinchey paper (2010) and shows teacher evaluation measures, descriptions of each, a summary of research on each method, its strengths and its cautions, as described by Hinchey.

Measure	Description	Research	Strengths	Cautions
Classroom Observation	Used to measure observable classroom processes, including specific teacher practices, holistic aspects of instruction, and interactions between teachers and students. Can measure broad, overarching aspects of teaching or subject-specific or context-specific aspects of practice.	Some highly researched protocols have been found to link to student achievement, though associations are sometimes modest. Research and validity findings are highly dependent on the instrument used, sampling procedures, and training of raters. There is a lack of research on observation protocols as used in the context for teacher evaluation.	• Provides rich information about classroom behaviors and activities. • Is generally considered a fair and direct measure by stakeholders. • Depending on the protocol, can be used in various subject, grades, and contexts. • Can provide information useful for both formative and summative evaluation.	• Careful attention must be paid to choosing or creating a valid and reliable protocol and training and calibrating raters. • Classroom observation is expensive due to cost of observers' time; intensive training and calibrating of observers adds to expense but is necessary for validity. • This method assesses observable classroom behaviors, but is not as useful for assessing beliefs, feelings, intentions, or out-of-classroom activities.
Principal Evaluation	Is generally based on classroom observation, maybe by structured or unstructured; uses and procedures vary widely by district.Is generally used for summative purposes, most commonly for tenure or dismissal decisions for beginning teachers.	Studies comparing subjective principal ratings to student achievement find mixed results. Little evidence exists on validity of evaluations as they occur in schools, but evidence exists that training for principals is limited and rare,	• Can represent a useful perspective based on principals' knowledge of school and context. • Is generally feasible and can be one useful component in a system used to make summative judgments and	• Evaluation instruments used without proper training or regard for their intended purpose will impair validity. • Principals may not be qualified to evaluate teachers on measures highly specialized for certain subjects or contexts.

		which would impair validity of evaluations	provide formative feedback.	
Instructional Artifact	Structured protocols used to analyze classroom artifacts in order to determine the quality of instruction in a classroom. May include lesson plans, teacher assignments, scoring rubrics, and student work.	Pilot research has linked artifact ratings to observed measures of practice, quality of student work, and student achievement gains. More work is needed to establish scoring reliability and determine the ideal amount of work to sample. Lack of research on use of structured artifact analysis in practice.	• Can be a useful measure of instructional quality if a validated protocol is used, if raters are well-trained for reliability, and if assignments show sufficient variation in quality. • Is practical and feasible because artifacts have already been created for the classroom.	• More validity and reliability research is needed. • Training knowledgeable scorers can be costly but is necessary to ensure validity. • This method may be a promising middle ground in terms of feasibility and validity between full observation and less direct measures such as self-report.
Portfolio	Used to document a large range of teaching behaviors and responsibilities. Has been used widely in teacher education programs and in states for assessing the performance of teacher candidates and beginning teachers.	Research on validity and reliability is ongoing, and concerns have been raised about consistency/ stability in scoring. There is a lack of research linking portfolios to student achievement. Some studies have linked NBPTS certification (which includes a portfolio) to student achievement, but other studies have found no relationship.	• Is comprehensive and can measure aspects of teaching that are not readily observable in the classroom. • Can be used with teachers of all fields. • Provides a high level of credibility among stakeholders. • Is a good tool for teacher reflection and improvement.	• This method is time-consuming on the part of teachers and scorers; scorers should have content knowledge of the portfolios. • The stability of scores may not be high enough to use for high-stakes assessment. • Portfolios are difficult to standardize (compare across teachers or schools) • Portfolios represent teachers' exemplary work but may not reflect everyday classroom activities.
Teacher Self Report	Teacher reports of what they are doing in classrooms. May be assessed	Studies on the validity of teachers self-report measures present mixed	• Can measure unobservable factors that may affect teaching, such	• Reliability and validity of self-report is not fully established and depends on

	through surveys, logs, and interviews. Can vary widely in focus and level of detail.	results. Highly detailed measures of practice may be better able to capture actual teaching practices but may be harder to establish reliability or may result in very narrowly focused measures.	as knowledge, intentions, expectation, and beliefs. • Provides the unique perspective of the teacher. • Is very feasible and cost-efficient; can collect large amounts of information at once.	instrument used. • Using or creating a well-developed and validated instrument will decrease cost-efficiency but will increase accuracy of findings. • This method should not be used as a sole or primary measure in teacher evaluation.
Student Survey	Used to gather student opinions or judgments about teaching practice as part of teacher evaluation and to provide information about teaching as it is perceived by students.	Several studies have shown that student ratings of teachers can be useful in providing information about teaching; may be as valid as judgments made by college students and other groups; and, in some cases, may correlate with measures of student achievement. Validity is dependent on the instrument used and its administration and is generally recommended for formative use only.	• Provides perspective of students who have the most experience with teachers. • Can provide formative information to help teachers improve practice in a way that will connect with students. • Makes use of students, who may be as capable as adult raters at providing accurate ratings.	• Student ratings have not been validated for use in summative assessment and should not be used as a sole or primary measure of teacher evaluation. • Students cannot provide information on aspects of teaching such as a teacher's content knowledge, curriculum fulfillment, and professional activities.
Value-Added Model	Used to determine teachers' contributions of students' test score gains. May also be used as a research tool (e.g., determining the distribution of "effective" teachers by student or	Little is known about the validity of value-added scores for identifying effective *teaching* though research using value added models does suggest that	• Provides a way to evaluate teachers' contributions to student learning, which most measures do not. • Requires no classroom	• Models are not able to sort out teacher effects from classroom effects. • Vertical test alignment is assumed (i.e., tests essentially measure the same thing from

	school characteristics.	teachers differ markedly in their contributions to students' test score gains. However, correlating value-added scores with teacher qualifications, characteristics, or practices has yielded mixed results and few significant findings. Thus, it is obvious that teachers vary in effectiveness, but the reasons for this are not known.	visits because linked student/ teacher data can be analyzed at a distance. • Entails little burden at the classroom or school level because most data is already collected for NCLB purposes. • May be useful for identifying upstanding teachers whose classrooms can serve as "learning labs" as well as struggling teachers in need of support.	grade to grade). • Value-added scores are not useful for formative purposes because teachers learn nothing about how their practices contributed to (or impeded) student learning. • Value-added measures are controversial because they measure only teachers' contributions to student achievement gains on standardized tests.