# thinkpoint

## Alternative models for educator evaluation

*by Edward Roeber, Ph.D.*

### INTRODUCTION

In recent years, the use of student assessment scores in the evaluation of educational personnel, especially teachers, has gained considerable traction. A number of states, spurred on by the U.S. Department of Education, have instituted test-based educator evaluation models. In these models, change or growth in student performance is typically used, along with various types of supervisor observations, to evaluate teachers.

Except for the use of test scores (and the attendant statistical procedures associated with this test score use), these personnel evaluation procedures are relatively traditional in approach — the supervisor gathers data, observes the individual being evaluated, and pronounces the effectiveness of the individual. Various means of appeal of these judgments are provided, and those evaluated are sometimes afforded opportunities for improving their practice and thus their evaluations. Increasingly, however, these evaluations have significant stakes for those evaluated: they may receive (or be denied) pay increases or promotions, and in some cases, may be terminated and/or lose their teacher certification.

Thus, the nature of educator evaluation and how it impacts educators, as well as more broadly how it affects the education of students are important considerations. This paper delves into the traditional evaluation model often used, proposes an alternative to it that in the view of this author is far more effective, and then describes in more detail how this model could be put into place.

## Two Basic Educator Evaluation Models

Two basic models can be used to evaluate educators: The **Inspection Model** and the **Demonstration Model**. Each model is described below, followed by some advantages and challenges of using each. This is followed by a list of the parameters that describe and define how the two models differ. Finally, a proposed professional practices portfolio model is presented in outline format.

**Inspection Model** — This model was traditionally used most often 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 or their designee, so if a teacher is being evaluated, it may be the building principal, assistant principal, or a department chair. The key questions the evaluator seeks to answer include:

- [ ] What level of proficiency does this individual educator possess?
- [ ] Is this individual an effective 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 information, collect written evidence of educator effectiveness such as lesson plans, gather test scores, and query others (e.g., peers, parents, and/or students) about their perceptions.

Then, the supervisor rates the individual, making a determination of the level of effectiveness 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 Michigan starting in 2018-19, at least 40% of the annual year-end evaluation shall be based on student growth and assessment data. In grades and subjects in which state assessments are administered, 50% of student growth must be measured using state assessments, with the rest being determined locally.

While traditionally used primarily for probationary educators, this process expanded to all educators starting around 2011 as a way to better differentiate educator effectiveness and focus professional development and support.

## Advantages of the Inspection Model:

**1.** Each employee 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.

## Challenges in using 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 tenth-grade teacher might not have comparable achievement data available for use in their evaluations.

**2.** The criteria for conducting the "inspection" may not be well documented so that different supervisors 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 observing or rating the same educator might rate the teachers completely differently, perhaps even differing in their overall judgment of proficiency. This area of weakness can lead to successful legal challenges.

**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 multiple occasions throughout the school year. This may lead to more "drive-by" observations, where the supervisor drops in for a few minutes and is not able to see adequately what is taking place in the classroom.

**5.** Some 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 personal interactions, not objective evidence of performance collected during the evaluation process.

## Demonstration Model – This

model is one in which each individual educator to be evaluated is charged with the responsibility of demonstrating her or his own proficiency by collecting and organizing evidence. 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, the educator prepares, organizes, and presents evidence to support this assertion: "I am a proficient educator who can produce effective results. I am taking steps to improve my competence. Here is my proof of both assertions. . ." Thus, each educator needs to select appropriate evidence—whether supplied by themselves, by peers or supervisors, or located from other sources. A portfolio with no evidence in it would be rated "not effective."

Each educator to be evaluated should develop individual goals that are then reviewed and approved by his or her supervisor. The educator 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 supervisor may

need to provide support to assist the educator to accomplish his or her goals.

The educator being evaluated is also responsible for the collection of evidence about his or her proficiency. Multiple sources of information are best, so the educator will need to consider how they can demonstrate their skills, especially when it comes to showing evidence of student learning, which should be a key goal for every educator. In addition, the teacher will need to document how they provided the instruction that led to the gains in student achievement, and may also wish to consider how they impacted the learning of subsets of students (e.g., students who struggled initially, students who came to the class as high achievers). 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; however, 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 teaching skills. An outline of a potential professional practices portfolio is shown in Table 1.

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 the individuals involved in hiring decisions, such as building principals, already have some experience examining such collections and using them to make personnel 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.

## Advantages of the Demonstration Model:

1. This model works equally well for teachers and school leaders. Any educator (actually, anyone working in a paid or unpaid position in a school) 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 can and 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 be thinking constantly 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/principal?"

4. Evidence collection will need to be on-going, so that the educator is thinking about demonstrating effectiveness 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 effectiveness 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 teachers' classrooms two or more times each year. Plus, these observations will be less summa-

tive and more formative, serving to collect data for teachers to use in the teacher demonstration of effectiveness.

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. A school social worker

with no direct instructional time could also describe her effectiveness in supporting children and their families to improve student achievement. It would also work well for educators in non-state tested grades and content areas, as well as educational specialists and supervisors in schools.

## Challenges in using the Demonstration Model:

1. The public, parents and local and state policy makers (i.e., the local school board or state legis-

Table 1: Proposed Professional Practice Portfolio

### IN THE FALL, DEVELOP:

1. **Individual educator identifies his or her annual goals**
   A. Goals from the District/School Improvement Plan – the educator's role in achieving one or more of the goals
   B. Goals for the individual educator – the educator's own goal(s)
      ☐ Short-term – This school year
      ☐ Long-term – Next year and beyond

2. **Plans for growth and improvement**
   A. Plans to help accomplish district/school goals– How will the educator accomplish the district/school goals?
   B. Plans to accomplish individual goals – How will the individual accomplish his or her own goals?

3. **Measures of performance**
   A. State measures where available and applicable
      ☐ M-STEP/MME/MI-Access/WIDA/Other
   B. School measures
      ☐ School's comprehensive needs assessment
      ☐ Interim benchmark assessments
      ☐ Common assessments
   C. Educator-created measures
      ☐ Content organization measures
      ☐ Individually-collected data
         • Summative information
         • Interim benchmark assessments
         • Formative assessment information

### IN THE SPRING, ADD:

4. **Summary and evidence of activities used to accomplish the plans and goals**
   A. Individual educator achievement of district/school goals – What activities did the educator engage in to accomplish the goals of the district/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 District/School Improvement Plan was accomplished?
   B. Individual goals – What evidence is there that individual goals were accomplished?
      ☐ Educator-collected information
      ☐ Peer information
      ☐ 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 put forth by the individual educator?
   C. Supervisor(s) – Does the supervisor support the evidence of accomplishment as put forth by the individual educator?

lators) 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 will need to learn how to document their performance.

3. Each educator's collection of evidence is unique. 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 broad and disparate sets of evidence. Educators (and others) who review the collections of evidence will need 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 effectiveness or just inability to collect good evidence?

## Parameters of educator evaluation models

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 evalua-

tion 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 with, of course, variations on each:

■ *Identify/eliminate low performing educators* – The purpose here is to identify low performing educators who, if they are unable to show improvement, can be dismissed from their current employment. 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 future employment.

■ *Encourage improvement in all educators* – This purpose is a recognition that educators need to work to stay current in their field and learn new skills due to changes in the profession and our understandings of how students learn. This recognizes that education, like other professions, is always changing and that educators need focused professional learning opportunities 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 effectiveness. An educator evaluation system can encourage the use of the best professional development activities that lead to increased effectiveness 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 the success of efforts to encourage school teams 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 Improvement Plan. However, how will the team feel about weak educators in the group?

3. **Current status or improvement in status** — 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 "change" model.

■ *Current status* – In this sort of evaluation, a series of criteria and measures are defined, and each educator is judged against these criteria. Michigan, for example, has identified four models to use as a basis for judging the effectiveness of teachers and three models for judging the effectiveness of administrators. Achievement results are also used, and often– ironically – changes in student performance are the achievement judgments made in evaluating 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.

■ *Change 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 their change in performance that becomes the basis for the evaluation. In the Inspection Model, any deficiencies identified by the supervisor might be used as the basis for the demonstration of change. 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 successful in carrying out the improvement activities specified, will be able to show improved performance at later times.

**4. Locus of control** — Who is responsible to conduct the evaluation? In traditional evaluations, the supervisor provides the evaluation for each sub-ordinate. Yet, in many real-life situations (e.g., job inter-

views), 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, are described above.

**5. Evaluation methods** — What method(s) will be used to evaluate educators?

One report (Hinchey, 2010) reviews several tools that can be used in the evaluation of educators. 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 the effectiveness of an educator is to observe the ed-

ucator 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 perceptions of parents or guardians about the quality of education provided to their child, the interactions of the teacher with the parent/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 are based on actual test score gains, predicted versus actual gains, or other models.

**6. Types of assessment information** — What type(s) of student achievement information will be used?

In a previous paper (Roeber, 2010), a 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), 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-grade or end-of-course 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 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. 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.

## Summary

The type of educator evaluation model used can have a significant impact on educators. This paper has presented two fundamentally different approaches to this important activity. The advantages and challenges in using each have been presented, along with some of the important parameters of educator evaluation that should also be considered when implementing either model. The goal of this is to present an alternate to current practice that if carried out by thoughtful individuals could serve to enhance educator skills and job commitment, and thus, the learning and achievement of students, which is, after all, the primary reason why we want skilled teachers in all of our classrooms.

## References

Hinchey, P. (2010) *Getting Teacher Assessment Right: What Policymakers Can Learn From Research*. Boulder, CO: National Education Policy Center.

Roeber, E. (2010) *Using Tests to Evaluate Teachers*. East Lansing, MI: Michigan Education Association.

# Questions for reflection

1. Is the Demonstration Model one that you can use in your setting? What do you see as the advantages and challenges in using it in your setting? How might you address these challenges?

2. If the Inspection Model is used in your setting, could some aspects of the Demonstration Model (e.g., collections of evidence of student learning and what the educator did to assist students in their learning) be used along with the Inspection Model?

# Resources for further study

Baker, E.L., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J. & Shepard, L. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. Briefing Paper # 278. Washington, DC: Economic Policy Institute. Retrieved December 7, 2010 from http://www.epi.org/publications/entry/bp278

Goe, L., Bell, C., & Little, O. (2008). Approaches to evaluating teacher effectiveness: A research synthesis. Washington, DC: National Comprehensive Center for Teacher Quality

Hinchey, P. (2010). Getting Teacher Assessment Right: What Policymakers Can Learn From Research. Boulder, CO: National Education Policy Center.

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

---

## Author

ED ROEBER, PH.D., is the assessment director for Michigan Assessment Consortium (MAC) projects. He co-chairs the MAC's Assessment Resource Development Committee and serves as Board Secretary/Executive Committee.

Dr. Roeber has earned degrees in Psychology and Political Science (B.A.) and Educational Psychology/Measurement and Evaluation (A.M. and Ph.D.) from the University of Michigan.

---