

# *Study Questions For*

## *Evaluating America's Teachers: Mission Possible?*

Presented below are several study questions linked to each of the nine chapters in *Evaluating America's Teachers: Mission Possible?* Although the questions, written by the book's author (W. James Popham), were initially generated for use by book-study groups, the questions need not be employed only in that fashion. For instance, people who read the book all by themselves—without being part of a study group—should not be precluded from giving additional thought to a chapter's contents. All of the questions are intended to elicit a reader's personal reactions rather than a “correct” or a “yes/no” response. Accordingly, readers will typically discover that, if they draw on a chapter's content as they arrive at their responses to a given chapter's study questions, their understanding of the chapter will usually be deepened. Because the questions were intended to be thought provoking, any thoughts they provoke will be welcomed.

### **CHAPTER 1**

#### *What Underlies the Tightening of Today's Teacher-Evaluation Programs?*

1. The book's opening chapter attributes today's dramatically increased attention to teacher evaluation to the federal government because of (1) the financial incentives linked to the Race to the Top Program and (2) the ESEA Flexibility Program's potential for states

to avoid NCLB sanctions. In both of these programs, a state's officials have a better chance of being successful if their state establishes a teacher-evaluation program consistent with federal guidelines embedded in those two programs. Given our nation's longstanding assignment of responsibility for public education to the states rather than to the federal government, what is your opinion regarding the appropriateness of the two-pronged "persuasion strategy" associated with this federal push toward more rigorous teacher evaluation?

2. Critics of the many teacher-evaluation programs emerging throughout the nation argue that, to the extent these new programs are flawed, inaccurate evaluations will be seen for many more teachers than necessary. The chapter concludes that when teacher evaluations are inaccurate, those most harmed will be students—not teachers. Why do you think this conclusion was reached, and do you agree with it?

3. Four implementation mistakes were identified in Chapter 1—mistakes that could render a state's teacher-education program ineffectual. The most serious of these, it was argued, is that the wrong kind of evidence will be selected to contribute to a "multiple-measures" approach to teacher evaluation. As you review the four potential implementation errors described in the chapter, do you agree with the contention that the most significant shortcoming of a teacher-evaluation strategy would be for teacher evaluators to rely on inappropriate evidence? If you agree, why so? If not, what other implementation mistake might be more serious?

4. Six recommended elements are identified in the chapter as components of the current federal vision of teacher evaluation. For example, one of those six elements was a call for such evaluation systems to be used for continual improvement of instruction. What is your opinion of these six requirements—both individually and as a group?

## CHAPTER 2

### *Human Judgment: Needed or Not?*

1. Do you believe that human judgment should play a prominent role in the evaluation of our nation's teachers? What reason(s) would you invoke to support your stance on this issue?

2. Some states' current teacher-evaluation frameworks allow greater implementation flexibility to local school districts than is seen in other states. If occasions for the exercise of human judgment are numerous (or significant) in the implementation of a state-approved teacher-evaluation process, how can an acceptable level of judgmental consistency be maintained among a state's diverse school districts?

3. If you were called on by a local school board to supply a brief, yet compelling argument in support of the role of human judgment when evaluating teachers, what sort of presentation would you put together?

4. What is your opinion about combining the formative and summative functions of teacher evaluation? Why?

## CHAPTER 3

### *Defensible Teacher Evaluation*

1. A five-step strategy for evaluating teachers was presented in this chapter. In Figure 3.1, this "weighted-evidence judgment" approach was depicted graphically. As you consider the five steps in weighted-evidence judgment, which *one* of those steps is, in your opinion, the most difficult to accomplish? Why?

2. Many phases of today's education process are dependent on financial resources that seem to be continually shrinking. Given that a judgment-based approach to teacher evaluation surely depends on the quality of those doing the judging, how do you believe a teacher-evaluation program can ensure that its judges are sufficient in number and sufficient in training to render first-rate judgments? In other words, what practical configuration of human judges seems to provide the most sensible, fiscally affordable way of carrying out weighted-evidence judgments? Must teacher evaluations be carried out only by solo school-site administrators?

3. How can *teacher-specific* adjustments in the weight of certain evaluative evidence be made without destroying the perceived legitimacy of this judgmental approach to evaluation? Or do you think such adjustments, because they would not be made identically for all teachers, should simply not be allowed? Why?

4. How can teacher evaluators' judgments be made in a manner that elicits external approval rather than criticism?

## CHAPTER 4

### *Evidence From Standardized Tests*

1. Early on in this chapter, in Figure 4.1, a medley of commonly employed sources of evidence was set forth, for example, classroom observations. If you were called on to identify what you regard as the most compelling types of evidence by which to evaluate teachers, which of the evidence sources presented in that table would you choose? Try to rank, in order, the top three evidence sources from Figure 4.1's illustrative evidence to determine a teacher's instructional ability. Please assume that each of the evidence sources listed was appropriately obtained, that is, collected as well as it could be collected.

2. Because the chapter's focus was on the use of students' performances on standardized tests, it was emphasized that greater evaluative weight should be given to students' scores on standardized tests whose technical (psychometric) adequacy had been verified. Do you agree that, if a group of educators do not possess the technical skills to ascertain the quality of the reliability, validity, and absence-of-bias data associated with a standardized test, this task can be assigned to external consultants? If not, what alternatives do you suggest should be employed to make sure that a standardized test passes psychometric muster?

3. How important is it to verify the degree to which, if a standardized test is used to help evaluate teachers, its instructional sensitivity must be demonstrated? Regardless of your personal preferences on this issue, if you were called on to persuade the designers of a new teacher-evaluation system that a test's instructional sensitivity was crucial, what kinds of argument do you think might be most persuasive?

4. If you were trying to explain to a colleague where standardized testing's traditional emphasis on supplying comparative score interpretations came from, what brief explanation could you put together to do that explanatory job well?

## CHAPTER 5

### *Evidence From Classroom Assessments*

1. Under what circumstances, if any, would evidence of student growth collected by classroom assessments be more indicative of a

teacher's skill than would students' performances on standardized tests?

2. At the close of this chapter, four factors are cited that should increase the evidential weight given to classroom assessment evidence when evaluating teachers. It was recommended that assessment results be given greater evaluative weight if those assessments (1) measure genuinely challenging curricular aims; (2) are accompanied by evidence supporting reliability, validity, and absence of bias; (3) were built, improved, and scored in accord with accepted practices; and (4) are accompanied by information attesting to the accuracy and honesty of the results. Now, looking back at those four factors, how would you rank their importance? That is, if you were obliged to list the four factors as most important, second most important, and so on, how would you rank these four considerations?

3. If you were trying to guide a particular school's teachers about ways of increasing the "believability" of any classroom-assessment evidence that's employed to evaluate teachers, what practical procedures would you suggest to those teachers so that "the world" would view the evaluative implications of students' performances on classroom assessments as credible?

4. As is true with standardized tests, the instructional sensitivity of classroom tests is a serious concern when using students' performances on those tests to reflect a teacher's instructional effectiveness. What tangible procedures might you recommend to teachers who want their teacher-made tests to be instructionally sensitive rather than instructionally insensitive?

## **CHAPTER 6**

### *Evidence From Classroom Observations*

1. If you had been asked by a district's school board to explain why a teacher who, based on observations of that teacher in action, might fail to display "instructionally sanctioned" behavior in the classroom, yet still promote high levels of students' achievement, what would your explanation be?

2. The two classroom-observation systems most widely used in the United States are those created by Charlotte Danielson and, more recently, the approach generated by Robert Marzano. As you contrast

these two systems, briefly described in Chapter 6, do you conclude that—for purposes of teacher evaluation—one is preferable to the other? And, if so, why do you think so?

3. Do you believe it is an advantage or a disadvantage for a classroom-observation system that was initially generated to support a formative-evaluation function to also be used for the summative evaluation of teachers? Why do you hold this view?

4. It was argued in the chapter that classroom observations can rather readily identify exceptionally strong and exceptionally weak teachers, but for most teachers—the ones who are not atypically weak or strong—classroom observations are not as useful. Do you agree or disagree with this contention? And, of course, why do you take your position on this issue?

## CHAPTER 7

### *Evidence From Ratings*

1. Suppose you were a teacher who is taking part in your school's "Back to School Night for Parents," and you have been asked to participate in a panel explaining to parents about the general nature of teacher evaluation and, in particular, the difference between (1) classroom observations of a teacher and (2) ratings of a teacher's quality. What would be the main points you would want to make during your part of the panel presentation?

2. Two sources of rating information were described in this chapter, namely, ratings supplied by administrators and ratings supplied by students. If you were a teacher evaluator who was assigning evaluative weights to these two types of ratings—assuming that the ratings had been collected as well as is possible—which of the two kinds of ratings, that is, ratings collected from administrators or students, do you believe should be given more weight? Why?

3. Do you think it is possible to reassure students about the anonymity of their teacher-evaluation ratings sufficiently well so that the bulk of students will end up rating their teachers honestly? How do you support your position on this issue?

4. If a state's teacher-evaluation framework is intended to satisfy recently articulated federal preferences regarding the evaluation of that state's teachers (the preferences described in the book's early

chapters), do you think that administrators' ratings of teachers will be meaningfully different than the kinds of administrative ratings we have seen in this nation during the past 100 years? What are the most prominent similarities or differences between today's administrative ratings of teachers and yesteryear's administrative ratings of teachers?

## CHAPTER 8

### *Evidence From Sundry Sources*

1. Suppose you were taking part in a school district's "customizing" of a state-decreed framework for teacher evaluation—a framework in which the state's school districts are permitted to tailor, modestly, their district's use of the state system. Please assume that you can choose two additional sources of information for the evaluation of your district's teachers. Now, as you review the chapter's 11 sorts of evaluative evidence, which *two* of them do you believe should be added to the more commonly employed state-determined sources of evaluative evidence (that is, the evidence sources described in Chapters 4–7)? Why?

2. One source of evaluative evidence treated positively in Chapter 8 was focused on shifts in students' affect. If you were a teacher who was attempting to persuade the other teachers in your school to adopt this kind of evaluative evidence, what arguments might you employ? On the other hand, if you were opposed to relying on evidence related to students' affective changes, what would those arguments be?

3. If you were called on to consider the 11 sundry sources of teacher-evaluation evidence treated in this chapter—with a view to isolating the three *least* appropriate kinds—which three would you select? And, of course, why?

## CHAPTER 9

### *Mission Possible?*

1. This wrap-up chapter opens with a brief reprise of the weighted-evidence judgmental approach to teacher evaluation advocated in the book. If you had been invited by a committee of state legislators

to describe the chief features of this judgment-based approach to teacher evaluation, how would you go about doing so? Assuming for the moment that you regard weighted-evidence judgment as a viable way to evaluate teachers, and you wanted to persuade the legislators not only to understand this approach, but also to support its use, what kind of rationale would you employ in favor of weighted-evidence judgment?

2. Five suggestions were offered in Chapter 9 regarding ways for individuals to either construct or improve a teacher-evaluation program. Which of these five suggestions do you regard as the easiest to follow? Why? Which of the five suggestions do you think is the most difficult to follow? Why?

3. In the book's final paragraph, a position is taken that the evaluation of America's teachers is an endeavor which, if carried out properly, can not only evaluate teachers accurately but can also, as a consequence of such accurate evaluations, end up benefitting our students educationally. What is your position on this issue? Do you agree or disagree with the author's conclusion? And, for the final time, why?

*W. J. P.*

*1/2/13*