

# Defensible Teacher Evaluation

*Student Growth Through  
Classroom Assessment*

*Rick Stiggins*

# Contents

<b>List of Tables</b>	<b>v</b>
<b>Acknowledgments</b>	<b>vii</b>
<b>About the Author</b>	<b>ix</b>
<b>Introduction</b>	<b>1</b>
<b>Chapter 1. Keys to Productive Teacher Evaluation in the New Era</b>	<b>7</b>
Ah, but There Is a Problem . . .	11
What Does It Mean to Evaluate Teachers Well?	12
<b>Chapter 2. Annual Accountability Test Scores? Unacceptable in This Case</b>	<b>19</b>
Test Scores: Inappropriate Evidence in Teacher Evaluation	21
Value-Added Analysis: If It Works in Economics, Will It Work in Education?	33
Summary and a Comment	37
<b>Chapter 3. Classroom Assessment to the Rescue</b>	<b>39</b>
Step 1: Establish and Prioritize the Achievement Standards	43
Step 2: Develop Assessment Plans	45
Step 3: Carry Out the Plans	48
Step 4: Compile and Submit a Final Portfolio and Report	50
Summary	54
The Key Is Universal Classroom Assessment Literacy	55

<b>Chapter 4. Similar to Student Learning Objectives, But Better</b>	<b>59</b>
The SLO Process	61
Summary	66
<b>Chapter 5. Tapping the Power of Assessment Literacy</b>	<b>69</b>
The Keys to Assessment Quality	72
Learning Targets as Assessment Foundations	75
Strong Classroom Assessments Begin with Strong Target Foundations	79
Matching Methods with Targets	82
Test Administration and Scoring	84
Effective Communication of Results	85
Factoring Student Affect into the Teacher Evaluation Equation	88
The Benefits of Classroom Assessment Literacy	93
Imagine What Would Become Possible	95
<b>Chapter 6. Required: Local School District Leadership and Assessment Action</b>	<b>99</b>
Action 1: Balanced Local Assessment Systems	101
Action 2: Continue to Refine Achievement Expectations	104
Action 3: Assure Assessment Quality	105
Action 4: Communicate Results Effectively	106
Assessment Literacy: The Foundation of Action	107
Conclusion	108
<b>Appendices</b>	
A. Classroom Assessment Quality Rating Scales for the Teacher Evaluation Context	111
B. Sample Affective Questionnaires	117
C. Rating Scales for Evaluating Completion of Five Essential Local Assessment Actions	127
<b>References</b>	<b>135</b>
<b>Index</b>	<b>139</b>

Chapter 4 addresses a currently popular approximation of a classroom assessment-based definition of student growth I am advocating called Student Learning Objectives. I will analyze the similarities and then detail the problems with this version that are remedied with the model defined in Chapter 3.

Chapter 5 describes the levels of assessment literacy needed in every classroom and school in order for teachers and supervisors to take full advantage of my approach. The challenge we face is that pathways for attaining those levels of competence have almost never been open to teachers or school leaders. The time has come to put a solid foundation of assessment literacy in place. As I will explain, sufficiently high levels of competence in classroom assessment can be achieved very economically.

Faculty-wide assessment literacy is not sufficient in and of itself, however. It represents only one part of a multi-layered assessment system infrastructure that local district and school leaders must put in place to achieve truly effective teacher evaluation. In Chapter 6, I identify a series of essential assessment actions local leaders must take to build the local assessment systems that can support effective teacher evaluation and meet the wide-ranging informational needs of practitioners in their organizations. The benefits associated with the completion of these actions are legion and, as I will demonstrate in this concluding chapter, can include immense learning benefits for students.

### **A MESSAGE FOR ALL OF THE KEY AUDIENCES**

For *local, state, and federal policy makers*, the message implicit in my presentation is that much of what has been unfolding in recent years by way of political action surrounding the consideration of student growth in teacher evaluation is technically unsupportable and, therefore, wrong. I will defend this proposition. Currently required practices very often are doomed to ineffectiveness for a variety of practical reasons. I will detail those. As policy makers, you must

# 2

## Annual Accountability Test Scores? Unacceptable in This Case

**T**hree federal school improvement initiatives are driving the national movement to weave student achievement growth into the teacher evaluation process. The Obama administration's 2009 American Recovery and Reinvestment Act launched the Race to the Top competition, encouraging states to strengthen their school improvement efforts. To receive federal funds, the program required increased rigor in teacher evaluation systems, including the integration of evidence of student growth into the evaluation process. Then, in 2011, the administration introduced their ESEA (Elementary and Secondary Education Act) Flexibility Program, in which states could receive a waiver from No Child Left Behind (NCLB) achievement growth requirements if they agreed to develop rigorous teacher evaluation systems based on multiple

**Table 2** Ineffective Evaluation Resulting from Reliance on Standardized Test Scores

<i>Criterion for effective evaluation</i>	<i>Difficulty with standardized test scores</i>
Specific academic achievement standards are identified and agreed to in advance.	The full array of academic achievement standards identified and agreed to in advance probably will not be well reflected in the test and those standards that are represented will not be sampled in sufficient depth to support inferences about student achievement.
Standards align with and sample the teacher's normal instructional responsibilities.	Even if some targets align with the teacher's assignment, most of those expectations probably will not be well reflected in the test and its resulting score(s).
Each standard is accompanied by an assessment plan and quality assessment(s).	Descriptions of the domains tested typically are available, but they may not overlap extensively with any particular teacher's assignment. Inferences about student mastery of specific achievement standards are therefore unsupported.
Pretest/post-test administration permits causal links between the instruction and learning success.	A pre/post span of a year opens the door to a wide variety of influences that are far beyond a teacher's control.
The evaluation process affords teachers the opportunity to describe factors, positive or negative, that impacted results.	When change in standardized test scores is the required criterion, scores alone count for a preset percentage of a teacher's final rating regardless of extraneous factors.

Considered in light of accepted standards for personnel action in the world of work in general, these issues create a major problem for policy makers wishing to make standardized test scores part of teacher evaluation. When personnel action is pending, it is expected that those judged not to have

literate supervisor and/or an assessment specialist, if one is available. Reviewers can recommend and oversee any needed revisions before the assessment is used.

As previously mentioned, if teachers from different schools or classrooms share responsibility for achieving student mastery of the same standard, then it would be in everyone's best interest to collaborate in the design and development of the assessment. This not only saves time and effort but also assures comparability of results across instructional settings.

Pre/post-test results should be reviewed with a teacher's supervisor as soon as those results become available. This review will either affirm the desired impact or suggest modifications in preparation for further instruction.

Finally, as instruction and assessment unfold, any unanticipated influences on student achievement that are beyond the teacher's control (e.g., integration of new students into the population, shifts in policy, required use of new texts or other resources) should be noted and reviewed with the supervisor. These will need to be included in the final report, but early review makes it far easier for the supervisor to make a fair and valid evaluation of the teacher's overall performance.

*Step 3 Prime Directive:* Each teacher must bring a solid foundation of assessment literacy to the process in order to manage it efficiently and effectively. Only if supervisors bring a firm understanding of the basic principles of sound assessment practice to this process can they help staff members collect and present dependable and credible evidence of their teaching proficiency.

#### **STEP 4: COMPILE AND SUBMIT A FINAL PORTFOLIO AND REPORT**

As the evaluation period nears an end, teachers must prepare to present evidence of their own proficiency to a supervisor. This presentation must do two things: (1) defend the efficacy of a teacher's instruction by reporting in detail

really important standards, inferences about a teacher's overall effectiveness based upon so few learning targets simply cannot be defended. The sample is far too small. The Chapter 3 classroom assessment approach asks teachers to select at least one learning target from each subject and level they teach. Further, if a teacher has fewer levels or subjects to teach, more targets per level and/or subject would be expected. In any event, a far larger sample is needed to yield stronger inferences about the impact of a teacher's instruction.

Typically, state guidelines provide an opportunity for mid-course evaluation of student and teacher progress in the SLO context, and it makes sense that the likelihood of positive teacher impact on student growth increases with the frequency of such reviews. My recommendations afford the teacher/supervisor team an ongoing set of opportunities for course corrections. For instance, throughout the year, as soon as a teacher completes a pretest/instruction/post-test sequence for a priority achievement standard, there should be an interim review of results and action planning based on those results. Given the larger sample of targets called for, the result will be a far more productive collaboration between teacher and supervisor.

There are several problems with the management of assessment relating to many state and local Student Learning Objectives guidelines. Most state recommendations intentionally focus the spotlight on commercially published assessments versus locally developed district and classroom assessments. Several states publish approved lists of acceptable commercial tests and require that any deviation from those lists receive state approval. Given the broad range and depth of learning targets spanning all school subjects and grade levels that would be candidates for SLO identification, it seems unlikely that a brief list of commercial tests could cover anything more than a fraction of them. To demand the use of tests which have a minimal chance of matching the objectives they would be purchased to assess seems pointless and wasteful—in short, damaging to assessing student learning or teacher impact.



*Avoiding Potential Sources of Bias and Distortion.*

- Do you notice anything in the assessment or way the assessment is carried out that might prevent a student from adequately demonstrating what he or she knows or can do? Potential sources of bias include unclear targets or purposes, an inappropriate assessment method, tasks and rubrics that are unclear or that give an unfair edge to particular groups, inadequate sampling, and problems with context—e.g., insufficient time or other testing conditions that inhibit maximum performance.

**Key 4: Good Communication.** Has the assessment developer planned for adequately managing information from the assessment and reporting it in ways that will meet the needs of all intended users?

- Has communication been planned from the outset as part of the assessment?
- Do teachers record assessment information accurately over time and appropriately synthesize it for reporting?
- Will the recipients of the results understand them and find them useful?

## LEARNING TARGETS AS ASSESSMENT FOUNDATIONS

A few details will help readers understand how these keys link to quality classroom assessment. For instance, our ATI classroom assessment task analysis reveals that we can subdivide achievement targets comfortably into four kinds, thus simplifying assessment planning (Stiggins and Chappuis, 2012). Here are the four:

- *Knowledge*—mastery of content—both knowing and understanding (consider, for example, the difference between knowing how to solve a linear equation and understanding when or why a person might use that knowledge).
- *Reasoning*—the ability to use that knowledge and understanding to figure things out and to solve problems, as in using knowledge to compare or classify things.
- *Performance Skills*—proficiency that can only be demonstrated through actual behavior, and where it is the behavior itself (not a resulting product) that is assessed

must replace it, a rationale for this change, and the likely impact of evolving beliefs on the prospects of weaving student growth into teacher evaluation.

The presentation that follows is derived from the ATI assessment action planning guide for local school district leadership teams designed to permit them to conduct their own local self-analysis of the quality of their local assessment system (Chappuis, Commodore and Stiggins, 2010). The discussion of essential assessment actions below is accompanied by a series of straightforward rating scales for each action in Appendix C. This combination reveals how completion of these actions can support using evidence of student growth in teacher evaluation, among other important instructional purposes.

### **ACTION 1: BALANCED LOCAL ASSESSMENT SYSTEMS**

The time has come to abandon the belief that annual standardized test scores provide sufficient information to support the development of effective schools. They have not done so for 60 years and cannot ever do so because they fail to meet the informational needs of crucial instructional decision makers—such as those who seek to evaluate teacher performance.

Let's replace this old thinking with the new belief that we must balance standardized test data with high-quality interim/benchmark and day-to-day classroom assessments, each serving their own unique set of users and uses. Further, a balanced assessment system must formalize and institutionalize use of assessment to both support student learning (formative applications) and certify that learning (summative).

The rationale for balancing assessments in these terms is that crucial data-based instructional decisions are made at many levels and for formative and summative purposes. Table 10 on the next page depicts the active ingredients in a truly balanced local assessment system. Note that each cell of the table makes a distinct contribution to instructional decision making and student well-being that the other cells cannot make.