EFF Research Principle: An Approach to Assessment Based on Cognitive Science

By Regie Stites

What Do We Mean by Assessment Based on Cognitive Science?

The fourth key research principle underlying the Equipped for the Future system reform initiative is an approach to assessment in adult basic education based on recent advances in cognitive science and educational measurement. Cognitive scientists have explored the development and nature of expertise in ways that help us better understand how adults learn and improve their ability to accomplish important tasks. Experts in educational measurement have used this research to develop new approaches to assessment design that move beyond traditional testing theory and practice. This new approach to test design will inform the next generation of educational assessments, including new and more valid tests for use in the adult education and literacy system.

The EFF approach to assessment builds on the body of cognitive science research described in *Research to Practice Notes 1-3*. Cognitive science research findings on the nature of expertise are the foundation for the **EFF Dimensions of Performance**. This four-dimensional model of adult performance (see discussion in *Research to Practice Note 2*) suggests a design for assessments that can answer the following questions about the use of the skill process defined by an EFF Standard:

1) What knowledge, skills, and abilities are evident in applying the EFF Standard to accomplish a purposeful activity?
2) How fluently can this integrated skill process be used to accomplish a task?
3) How independently can it be used?
4) How broadly (range of tasks and task environments) can it be used?

This *Research to Practice Note* provides an overview of the cognitive science and measurement theory and research findings that support the EFF approach to assessment. Among the key findings addressed are the following:

- Good assessment is grounded in a clear and explicit model of the cognitive processes that underlie complex and purposeful activity.
- Good assessment is a process of reasoning from evidence that permits valid and reliable inferences about ability to be drawn from a limited opportunity to observe and interpret performance.
- Good assessment provides information that learners, teachers, and others can easily understand and use to reflect upon and improve learning and teaching.

The EFF publication *Results That Matter: An EFF Approach to Quality* presents five key principles that reflect the theoretical foundations of EFF. Program practices that support these principles provide guideposts by which programs, teachers, students, and their communities can assess their implementation of the EFF Framework. They help practitioners to better answer the questions "What does it mean to practice EFF?" and "What does EFF implementation look like in action?" These *Research to Practice Notes* will help you to:
- identify the research basis for the principles;
- learn key concepts and terms associated with the principles;
- see examples of how other programs have implemented the program practices;
- reflect on how you and your program can implement the program practices.

How Cognitive Science Informs EFF

Researchers have recently made great progress in understanding how people think and learn. In the last few years, there has been an extraordinary growth in scientific work on the mind and brain. We now have ways to study not only the *products* of thinking and learning but also the *processes* by which people acquire new information, such as the neural processes that occur during thought and learning and the process through which people develop competence and exper-tise. The multidisciplinary group of researchers who have conducted this work have coined a term for the study of thinking and learning: *cognitive science*. Cognitive science research represents one of the key conceptual underpin-nings of the EFF Framework for teaching, learning, and assess-ment and provides the research basis for the constructivist theory of learning. For teachers who would like to read more about cognitive science research, the National Academy of Sciences has produced three publications that synthesize this work and its implications for education: *How People Learn: Brain, Mind, Experi-ence, and School*, by Bransford, Brown, and Cocking (1999); *How People Learn: Bridging Research and Practice*, by Donovan, Brans-ford, and Pellegrino (1999); and *Knowing What Students Know: The Science and Design of Educational Assessment*, by Pellegrino, Chudowsky, and Glaser (2001).

What Research Says About Good Assessment

Good assessment is grounded in a clear and explicit model of the cognitive processes that underlie complex and purposeful activity.
Quality in any assessment is always first and foremost a matter of being clear about what it is you want to assess—what we call the construct. Starting with a clear construct is an essential first step in designing good assessments. Accord-ing to Samuel Messick (1992, p. 17), "A construct-centered approach would begin by asking what complex of knowledge, skills, or other attributes should be assessed, presumably because they are tied to explicit or implicit objectives of instruction or are otherwise valued by society. Next, what behaviors or per-formance should reveal those constructs, and what tasks or situations should elicit those behaviors? Thus, the nature of the construct guides the selection or construction of relevant tasks as well as the rational development of construct-based scoring criteria and rubrics." Educational content standards are one way to answer the question of "what complex of knowledge, skills, or other attrib-utes should be assessed." The **EFF Assessment Framework** has been developed to enable teachers and assessment developers to align assessments with EFF content standards and with teaching focused on the skill domains defined by the standards. This alignment is the key to a standards-based reform process that includes both educational accountability and educational improvement (see NRC, 1999a; Stites, 2002, EFF Assessment Consortium, 2004).

Aligning assessment with important content. Ideally, the constructs that assessments are designed to measure should reflect broad-based agreement on what adults should know and be able to do. When assessment results influence high-stakes decisions (for example, educational program funding) it is espe-cially important that test results match the learning goals and objectives that are important to learners, teachers, and society. Unfortunately, the limitations of traditional testing approaches often result in a construct that is too narrow-ly defined. In this case, test results will only capture a small part of the impor-tant learning that has taken place. According to a number of sources cited in a recent report by the National Research Council's Committee on the Founda-tions of Assessment (NRC, 2001), traditional testing approaches and designs do not adequately capture the complex knowledge and skills that are called for in contemporary educational standards and that are needed for success in 21st century work and life (Resnick and Resnick, 1992; Glaser, Linn, and Bohrnst-edt, 1997; NRC, 1999b; Rothman, Slattery, Vranek, and Resnick, 2002). New approaches to assessment, such as model-based performance assessment (Baker, 1998), evidence-centered design (Mislevy, Steinberg & Almond, 1999; Mislevy et al., 2001), and related approaches (see, NRC, 2001) have laid out principles for design of assessments to capture and evaluate evidence of com-plex problem-solving and higher-order think-ing skills under conditions that simulate (and better predict transfer to) real-world environments

Upon reflection…
- *What kind of assessment best reveals understanding and strategic thinking?*
- *What kind of assessment task best reveals a student's ability to use the Components of Performance of an EFF Standard?*

Good assessment is a process of reasoning from evidence that permits valid and reliable inferences about ability to be drawn from a limited opportunity to observe and interpret performance.

According to the authors of *Knowing What Students Know: The Science and Design of Educational Assessment*, "an assessment result is an estimate, based on samples of knowledge and performance from the much larger universe of everything that a person knows and can do" (NRC, 2001, p. 37, emphasis in the original). In essence, educational tests are tools that permit us to draw reasonable inferences about what a person knows and can do and thus assessment is a process of "reasoning from evidence" (NRC, 2001, p. 38, citing Mislevy, 1994, 1996). In order to ensure that the inferences we draw from test results about performance ability are sound, we need to collect good evidence and interpret it carefully.

The assessment triangle. In order to design assessments that support sound reasoning from evidence we need to make sure that three critical elements of good assessment design are defined clearly and aligned coherently:

- a cognition model (student model) that defines the aspects of achievement to be assessed;
- an observation model (task model) that defines tasks that will be used to collect evidence about performance ability; and
- an interpretation model (scoring model) that defines methods used to analyze the evidence resulting from the performance on the tasks.

The relationships among these three elements are represented in the assessment triangle below.

Observation   Interpretation
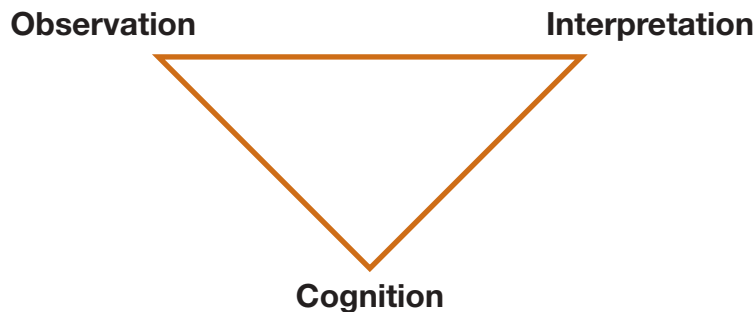
Cognition

Starting with cognition. In applying the principles of the assessment triangle to the design of assessments for the EFF Standards, we start with the cognition corner of the triangle, or with an EFF Content Standard and the Performance Continuum for that standard. Each EFF Content Standard and Performance Continuum defines an integrated skill process and includes:

- the four dimensions of the EFF Performance Continuum (increasing depth and structure of the knowledge base, and increasing fluency, independence, and ability to perform in a range of conditions).
- the performance level descriptors, which identify the specific behavioral indicators on each of the four dimensions.

According to Robert Mislevy: "*Good teachers have always relied on a wider array of means to learn about how the students in their classes are doing and to help plan further learning. Alongside the tests and quizzes they design and score … they also use evidence from projects, work in class, conversations with students… Teachers call these 'informal' assessments, in contrast with the 'formal' assessments typified by large-scale standardized tests. Good teachers implicitly exploit the principles of cognitive psychology, broadening the universe of discourse to encompass local information and address the local problem at hand. Yet precisely because informal assessments are thus individuated, neither their rationale or their results are easily communicated beyond the classroom. Standardized tests do communicate efficiently across time and place – but by so constraining the universe of discourse that the messages often have little utility in the classroom.*

"The challenge … is to devise assessments that, in various ways, incorporate and balance the strengths of formal and informal assessments by capitalizing on an array of methodological, technological, and conceptual developments."

—Robert J. Mislevy (1997), Postmodern test theory, p. 189.

We need to look at the Performance level descriptions for an EFF Standard when we develop assessments because they identify benchmarks for knowledge, skill, and strategic ability that can be used to guide development of tasks and scoring criteria (see examples in *Improving Performance, Reporting Results: The Guide to the Read With Understanding Assessment Prototype*, EFF Assessment Consortium, 2004).

From cognition to observation. The observation corner of the triangle reflects a set of beliefs about the kinds of tasks or situations that will best afford opportunities to observe and collect evidence on the important knowledge, skills, and other aspects of performance identified in the cognition corner (what students need to know and be able to do). As noted in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999, p. 137), "[e]very test, regardless of its format, measures test-taker performance in a specified domain. Performance assessments, however, attempt to emulate the context or conditions in which the intended knowledge or skills are actually applied." We believe that complex, performance-based assessment is the most appropriate method for capturing evidence of proficient performance on the complex, integrated skill process defined by an EFF Standard and its performance levels. Other testing formats (for example, selected response formats such as multiple choice items) can also be developed as measures of the EFF Standards, but any single test of this type may capture evidence of only a portion of the construct (cognition model) defined by an EFF Standard.

From observation to interpretation. Moving from the observation to the interpretation corner of the assessment triangle takes us to methods and tools for reasoning from the evidence produced in a performance task. Our goal in the interpretation corner is to make valid inferences about a student's ability to perform on the skill process that constitutes the targeted content for the assessment (the cognition corner or student model). One of the tools for reasoning from evidence included in EFF assessment design specifications is a template for the scoring rubrics that will be used to evaluate the performance evidence produced by tasks. The criteria in the these rubrics are based on the descriptions of knowledge and strategies to be assessed at each performance level and also linked directly to the nature of the evidence of performance produced by the assessment task. Because all observations are limited, multiple measures are always preferable. Design specifications, model assessment tasks, and scoring rubrics for the EFF Standard Read With Understanding can be found in *Improving Performance, Reporting Results* (EFF Assessment Consortium, 2004).

Good assessment provides information that learners, teachers, and others can easily understand and use to reflect upon and improve learning and teaching.

The passage from Lorrie Shepherd in the text box on this page highlights the potential for assessment to have both positive and negative effects on teaching and learning. If tests measure what is important, then teachers are more likely to teach—and learners are more likely to learn—important knowledge, skills,

Assessment reform is part of a larger effort to raise standards and improve the quality of education. Standards-based reform envisions a more challenging curriculum for all students focused on higher order thinking skills and depth of understanding. It involves a thoroughgoing reconceptualization of what it means to know in each of the disciplines, as well as fundamental changes in teaching and learning consistent with constructivist theory. Transforming assessment is seen as an essential part of curriculum reform because of widespread beliefs and evidence documenting the distorting effects of high-stakes basic skills tests on teaching and learning (Madaus et al., 1992; Resnick & Resnick, 1992; Romberg, Zarinnia, & Williams, 1989). The belief, that the content of assessments had to be changed to effect other changes, was captured in the slogan "WYTIWYG," "What You Test Is What You Get."

—from Lorrie A. Shepherd (2000). The Role of Classroom Assessment in Teaching and Learning, p. 19.

EFF Research Principle: An Approach to Assessment Based on Cognitive Science

and abilities. Assessments need to clearly communicate information (or be "transparent") and alternative (or performance-based) assessments do that well. As many experts have pointed out (see Wilson & Adams, 1996; Wiggins, 1998; NRC, 1999a; NRC, 2001), this "transparency" is a key advantage of alternative assessments over traditional assessment approaches. Such assessments are better able to communicate to all stakeholders—adult learners, teachers, policymakers, and the public—what tests are measuring, what the results of tests mean, and how test results can be used as indications of program quality and as guides for instructional improvement.

The National Research Council convened a panel of assessment and adult education experts in December 2001 to discuss issues in alternative assessment for accountability in adult basic education. The panel concluded that a serious obstacle to improving assessment and accountability practices in adult basic education is that there is, as yet, no national consensus on the content to be assessed (NRC, 2002). EFF was developed to address the need for such a common framework.

Putting Good Assessment Practices to Work in Your Program

Results That Matter: An Approach to Program Quality Using Equipped for the Future (Bingman & Stein, 2001) provides a vision of what system reform at the program level looks like using EFF Standards, referred to as the **EFF Quality Model.** Within this model, the principles of EFF described in Research to Practice Notes 1-4 are linked to key program practices and to predicted short- and long-term outcomes. These Program Practices provide guideposts by which administrators, teachers, students and their communities can assess their implementation of the EFF Framework. How might your program implement these Program Practices? Below are two examples of how your program can implement good assessment practices and questions for further discussion and reflection. For more information, see the EFF Assessment Resource Collection (http://eff.cls.utk.edu/assessment/).

EXAMPLE 1:

Teachers and students use the EFF Standards and Performance Levels to determine appropriate starting points for instruction and to plan learning activities.

The following scenario takes place in a workplace-based ABE class. The class is working on reading and understanding materials related to health and health insurance. As this is a multi-level class, the students have a range of reading abilities. The teacher uses the Read with Understanding performance level descriptions of key knowledge, skills, and strategies (see *A Guide to Using the Read With Understanding Performance Levels* in the online EFF Assessment Resource Collection at http://eff.cls. utk.edu/assessment/guides/read/read1.html) to evaluate each student's reading performance ability and to decide what level of performance on Read with Understanding they should be working toward. Based on a variety of sources of evidence, including in-class assessments and observations of each student's performance in reading activities in the classroom, the teacher has determined that some students in the class are performing at EFF Read With Understanding performance level 2, others typically show performance ability in reading consistent with the descriptions in EFF performance level 3, and still others show performance at EFF level 4. The teacher uses this appraisal of her students' performance levels to help plan their learning activities and to monitor their learning progress and achievement.

Upon reflection…

- *How does the EFF approach to planning instruction differ from a text-based planning process?*

- *How might you use the EFF Standards and performance levels to help plan learning activities?*

The EFF performance level descriptions give the teacher ideas for planning instruction. In planning learning activities, the teacher looks to the EFF performance level just above the level that best describes the students' current reading performance. The students who generally show reading performance ability matching the performance level descriptions at EFF level 2 will work on the knowledge, skills, and strategies described in EFF performance level 3; those showing EFF level 3 ability will work on the skills described in level 4; and so on. For the students who will be working on the level of reading performance described at EFF level 3, the teacher has prepared a set of activities that focus on identifying portions of the text that answer the pre-reading questions using simple strategies such as scanning for key words. For those students working on the reading performance described at EFF level 5, on the other hand, the teacher has prepared learning activities that include reading to complete a comparison chart of several health insurance plans and identifying what information is missing from a plan's description.

EXAMPLE 2:

Teachers use the EFF Standards and Performance Levels to select or develop assessment tasks, rubrics and other evaluation tools to monitor learning progress and achievement.

This scenario in this example takes place in a community college ESOL program. To create her own instructional assessments and scoring rubrics aligned to the EFF Standard Listen Actively, the teacher plans her approach by addressing two key issues in assessment design: gathering sufficient performance evidence and then scoring it fairly and appropriately.

In developing assessment tasks, she first considers the issue, problem or topic that is the context for teaching and learning on the standard. (For example, students in the class have recently been working on issues related to parenting and helping children with schoolwork.) Next, she thinks about a meaningful, real-life task related to this issue that learners will care about accomplishing, that is doable in the instructional setting, and that will allow learners to produce evidence that they have mastered (or have not yet mastered) the targeted standard at the appropriate level. Since the students in the class are all concerned about upcoming parent-teacher conferences, the teacher decides to design a series of instructionally-embedded assessment tasks and an end-of-unit mastery test using simulated parent-teacher conferences as the task context in which to observe and collect evidence of progress in improving proficiency in application of the EFF Standard Listen Actively.

For reflection…

- *How could your program use EFF Standards and Performance Levels to develop instructionally-embedded assessments?*

- *What are the advantages of aligning the content of assessments used for instructional purposes and the content of assessments used for achievement and program accountability reporting?*

The teacher uses the Listen Actively performance level descriptions to design learning activities that incorporate performance assessment tasks and scoring rubrics. These embedded assessments (part learning and part assessment) are designed as opportunities for students to try out their listening skills and to provide both students and teacher with a window on what students can do well and what they still need to work on. As the time of the actual parent/teacher conferences draws near, the teacher develops a more formal performance assessment task to document the progress that students have made and give them confidence in their improved listening ability going into the conferences. The assessment task again takes the form of a simulated parent/teacher conference, but unlike the embedded assessment activity, the conditions for performance are more authentic. A native English-speaking assistant takes the role of "teacher" in the conference and students must go through the role-play from start to finish without pauses, restarts, or requests for outside assistance. The teacher observes the interaction and afterwards debriefs each student on their performance using a scoring rubric as a guide to point out areas of strength and weakness

in the listening performance. The teacher and student also discuss strategies for making best use of the listening abilities the student has demonstrated mastery of in the upcoming parent/teacher conference.

Glossary

Continuum of Performance: A multidimensional, developmental description of performance on an EFF Standard ranging from the novice level to the expert level. The continuum is built around the four EFF Dimensions of Performance, and performance levels are defined by identifying key features of performance at various points along the continuum. (Stein, 2000, pp.58-59.).

Dimensions of Performance: The theoretical foundation, based in cognitive science, on which the EFF Continuum of Performance for each skill is built. The Dimensions of Performance identify developmental differences in performance on the EFF Standards related to four areas
- *Structure of the knowledge base:* Description of vocabulary, content knowledge, and strategies for organizing and applying content knowledge,
- *Fluency of performance:* Description of the ease, fluidity, and automaticity evidenced in performance,
- *Independence of performance:* Description of the degree of initiative and self-reliance evidenced in performance, and
- *Range of conditions for performance:* Degree to which a task and the contexts of a task are familiar or unfamiliar to the learner, the extent to which tasks are structured (scaffolded) or unstructured, and the complexity of the tasks. (Stein, 2000, pp.55-60; Bransford, Brown, & Cocking, 1999.)

EFF Assessment Framework: The EFF Assessment Framework defines levels of performance and measures of performance for a variety of assessment purposes. The Framework describes adult performance along four dimensions (structure of the knowledge base, fluency of performance, independence of performance, and range of conditions for performance). When completed, the Framework will include tools that enable programs using EFF as a framework for instruction to report student progress from level to level on the National Reporting System.

EFF Quality Model: A vision of what system reform at the program level looks like using EFF Standards. The EFF tools, foundational theory and research, expected program practices, and predicted short- and long-term outcomes are presented and explained in the publication *Results That Matter: An Approach to Program Quality Using Equipped for the Future* (Bingman & Stein, 2001). Ordering and downloading information can be found at http://www.nifl.gov/lincs/collections/eff/eff_publications.html.

Performance-based assessment: A tool for measuring student learning that requires the student to construct or produce a response to an assessment item or task. Performance assessments attempt to emulate the context or conditions in which the intended knowledge or skills are actually applied. Examples might include on-demand writing tasks, projects resulting in a product, performance, or event, and portfolios involving a collection of student work related to multiple standards or themes. Performance-based assessment plays an important role in the EFF standards-based education and improvement model. (Ananda, 2000.)

Performance Task: A meaningful, real world learning activity that includes all of the components of performance for a standard and provides opportunities for development of all four dimensions of performance. A performance task informs all steps of the instructional planning, teaching, and assessment cycle. (Ananda, 2000; McGuire, 2000.)

A complete Equipped for the Future Glossary is available online at:
http://www.nifl.gov/lincs/collections/eff/eff_glossary. html

References

Ananda, S. (2000). *Equipped for the Future assessment report: How instructors can support adult learners through performance-based assessment.* Washington, DC: National Institute for Literacy.

Baker, E.L. (1998). *Model-based performance assessment. CSE Technical Report 465.* Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Bingman, B. & Stein, S. (2001). *Results that matter: An approach to program quality using EFF.* Washington, DC: National Institute for Literacy.

EFF Assessment Consortium. (2004). *Improving performance, reporting results: The guide to the Read With Understanding assessment prototype.* Washington, DC: National Institute for Literacy.

Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress.* New York, NY: National Academy of Education.

Madaus, G.F., West, M.M., Harmon, M.C., Lomax, R.G., & Viator, K.A. (1992). *The influence of testing on teaching math and science in grades 4-12.* Chestnut Hill, MA: Boston College, Center for the Study of Testing, Evaluation, and Educational Policy.

McGuire, P. (2000). *A performance framework for teaching and learning with the Equipped for the Future (EFF) framework. Adventures in Assessment 12* (Winter). (http://www.sabes.org/resources/adventures/vol12/12mcguire.htm)

Mislevy, R.J. (1997). Postmodern test theory. *In Transitions in work and learning: Implications for assessment.* Washington, DC: National Academy Press.

Mislevy, R.J., Steinberg, L.S. & Almond, R.G. (1999). *Evidence-centered assessment design.* Princeton, NJ: Educational Testing Service. (http://www.education.umd.edu/EDMS/EDMS738/papers/ECD_overview)

Mislevy, R.J., Steinberg, L.S., Almond, R.G., Haertel, G.D. & Penuel, W.R. (2001). *Leverage points for improving educational assessment. CSE Technical Report 534.* Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

National Research Council. (1999a). *Testing, teaching, and learning: A guide for states and school districts.* Committee on Title I Testing and Assessment, R.F. Elmore and R. Rothman, (Eds.), Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

National Research Council. (1999b). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress.* Committee on the Evaluation of National and State Assessments of Educational Progress, J.W. Pellegrino, L.R. Jones, and K.J.Mitchell, (Eds.), Commission on Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment, J.W. Pellegrino, N. Chudowsky, and R. Glaser (Eds.), Center for Education. Division on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

National Research Council. (2002). *Performance assessments in adult education: Exploring the measurement issues.* Committee for the Workshop on Alternatives for Assessing Adult Education and Literacy Programs, R.J. Mislevy and K.T. Knowles (Eds.), Center for Education. Division on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Resnick, L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B.R. Gifford and M.C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction.* Boston: Kluwer.

Romberg, T.A., Zarinnia, E.A., & Williams, S. (1989). *The influence of mandated testing on mathematics instruction: Grade 8 teachers' perceptions.* Madison, WI: University of Wisconsin, National Center for Research in Mathematical Science Education.

Rothman, R., Slattery, J.B., Vranek, J.L., & Resnick, L.B. (2002). *Benchmarking and alignment of standards and testing, CSE Technical Report 566*, Los Angeles, CA: Center for the Study of Evaluation, National Center for Research on Standards and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles.

Shepherd, L.A. (2000). *The role of classroom assessment in teaching and learning.* Santa Barbara, CA: Center for Research on Education, Diversity & Excellence, University of California, Santa Barbara.

Stein, S.G. (2000) *Equipped for the Future content standards.* Washington, D.C.: National Institute for Literacy.

Stites, R. (2002). *Assessing results that matter: Equipped for the Future's approach to assessment for adult basic education accountability and improvement.* Washington, DC: National Institute for Literacy.

Wiggins, G.P. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* New York, NY: Jossey-Bass.