

Research Brief



Number 2

Report from the Department of Research, Evaluation, and Assessment

March 4, 2009

PERFORMANCE ASSESSMENT: A KEY COMPONENT OF A BALANCED ASSESSMENT SYSTEM

AUTHOR: Douglas G. Wren, Ed.D., Assessment Specialist
Department of Research, Evaluation, and Assessment

OTHER CONTACT PERSON: Jared A. Cotton, Ed.D., Assistant Superintendent
Department of Research, Evaluation, and Assessment

ABSTRACT

Performance assessment is used to evaluate higher-order thinking and the acquisition of knowledge, concepts, and skills required for students to succeed in the 21st century workplace. A review of relevant literature on performance assessment was conducted for this report, which includes a clarification of the term “performance assessment,” a comparison of traditional assessments and performance assessments, and a description of the procedures involved in developing performance assessments as well as the rubrics used to score the assessments.

Performance assessment is about performing with knowledge in a context faithful to more realistic adult performance situations, as opposed to out of context, in a school exercise.

--Grant Wiggins (2006)

A Strategy for Success

On October 21, 2008, the School Board of Virginia Beach adopted a new strategic plan for Virginia Beach City Public Schools (VBCPS). *Compass to 2015: A Strategic Plan for Student Success* includes five strategic objectives. The second objective states, “VBCPS will develop and implement a balanced assessment system that accurately reflects student demonstration and mastery of VBCPS outcomes for student success.” One of the key strategies of this objective is “Develop and/or adopt performance-based assessments and rubrics to measure critical thinking and other division outcomes for student success.”

What exactly are performance-based assessments? As is sometimes the case with professional jargon, expressions are used freely with the assumption that everyone is familiar with their meanings. This research brief will define and give examples of performance-based assessments, describe the process for their development, and explain how performance assessments fit within a balanced assessment system.

Contents:

<i>A Strategy for Success</i>	<i>p. 1</i>
<i>Performance-Based Assessment: Definitions</i>	<i>p. 2</i>
<i>Traditional vs. Performance Assessments</i>	<i>p. 3</i>
<i>Developing Performance Assessments</i>	<i>p. 4</i>
<i>A Balanced Assessment System</i>	<i>p. 8</i>
<i>Summary</i>	<i>p. 9</i>

Performance-Based Assessment: Definitions

The term “performance-based assessment” is frequently referred to as performance assessment, or by its acronym, PBA. Herrington and Herrington (1998) noted that the terms “performance assessment” and “authentic assessment” also tend to be used interchangeably. While performance assessment and PBA are simply shortened versions of performance-based assessment, there is a notable difference between authentic assessment and performance-based assessment. Gulikers, Bastiaens, and Kirschner (2004) cited previous literature to explain the difference between performance assessment and authentic assessment.

Some see authentic assessment as a synonym to performance assessment (Hart, 1994; Torrance, 1995), while others argue that authentic assessment puts a special emphasis on the realistic value of the task and the context (Herrington & Herrington, 1998). Reeves and Okey (1996) point out that the crucial difference between performance assessment and authentic assessment is the degree of *fidelity* of the task and the conditions under which the performance would normally occur. Authentic assessment focuses on high fidelity, whereas this is not as important an issue in performance assessment. These distinctions between performance and authentic assessment indicate that every authentic assessment is performance assessment, but not vice versa (Meyer, 1992).

There has been a considerable amount of information written on performance assessment over the past three decades. Consequently, there are numerous definitions of the term currently available. Palm (2008) observed that some of the definitions were extremely broad, while others were quite restrictive, and that most of the definitions of performance assessment were either response-centered (i.e., focused on the response format of the assessment) or simulation-centered (i.e., focused on the student performance observed during the assessment).

A publication from the Office of Educational Research and Improvement of the U.S. Department of Education (1993) provides an example of a response-centered definition of performance assessment.

Performance assessment . . . is a form of testing that requires students to perform a task rather than select an answer from a ready-made list. For example, a student may be asked to explain historical events, generate scientific hypotheses, solve math problems, converse in a foreign language, or conduct research on an assigned topic.

A simulation-centered approach with reference to “real-life contexts” is evident in the definition of performance assessment included in the glossary of the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Performance assessment – Product- and behavior-based measurements based on settings designed to emulate real-life contexts or conditions in which specific knowledge or skills are actually applied.

It is important to note that the word “emulate” is used in the *Standards* definition above. Educators should keep in mind that performance assessments are more meaningful when they

imitate real-life situations. Furthermore, Wiggins (1992) suggested that a well-designed performance assessment can be “enticing” to students, which seems plausible. In all probability, more students would rather participate in one of the following activities (examples of performance assessments) than take a paper-and-pencil test:

- Design and construct a model
- Develop, conduct, and report the results of a survey
- Perform a science experiment
- Write a mock letter to the editor of a newspaper

Traditional Assessments vs. Performance Assessments

The increase in popularity of performance assessments during the late 1980s and 1990s came about in part because of dissatisfaction with traditional, multiple-choice tests (Kahl, 2008). By the end of the 20th century, performance assessment had moved “from a trendy innovation to an accepted element of good teaching and learning” (Brandt, 1998). With the increase in standardized testing after the *No Child Left Behind Act of 2001* was signed into law, educators took a renewed interest in different types of alternative assessments, including performance assessment.

As seen in Table 1, performance assessment has a number of advantages over traditional assessment for evaluating individual students. Most notably, performance assessment has the capacity to assess higher-order thinking and is more student-centered than traditional assessment.

Table 1
Attributes of Traditional Assessments and Performance Assessments

Attribute	Traditional Assessment	Performance Assessment
Assessment Activity	Selecting a response	Performing a task
Nature of Activity	Contrived activity	Activity emulates real life
Cognitive Level	Knowledge/comprehension	Application/analysis/synthesis
Development of Solution	Teacher-structured	Student-structured
Objectivity of Scoring	Easily achieved	Difficult to achieve
Evidence of Mastery	Indirect evidence	Direct evidence

Sources: Liskin-Gasparro (1997) and Mueller (2008).

Advocates of performance assessment also emphasize that it is more in line with instruction than traditional assessment (Palm, 2008). While Popham (2001) and other experts (Haladyna, Nolen, & Haas, 1991; Mehrens, 1991) agree that teaching to the test—Popham refers to the practice as item-teaching—is highly unethical in preparation for traditional assessments, teaching to the test is actually encouraged when it comes to performance assessments (Mueller, 2008). With performance assessment, students have access to scoring rubrics in advance so they will know exactly how their performance (e.g., oral or written response, presentation, journal) will be evaluated. Teachers should also allow their students to preview examples of high-quality and poor performance products to use as models, provided the product cannot be mimicked.

In November 2008, the National Academy of Education (NAE) hosted the *Education Policy in Transition Public Forum* in Washington, D.C. The purpose of the forum was to facilitate a discussion between educational researchers, policy leaders, and advisers to Congress

and the new administration on the most critical issues in education policy. One of the main concerns mentioned by the panel on Standards, Accountability, and Equity in American Education was that:

accountability tests over-represent what is relatively easy to measure (such as basic skills) and under-represent highly valued reasoning skills such as problem solving. Because there are consequences for schools and school districts (and sometimes students and teachers) for how well students perform on the tests, the accountability system establishes strong incentives for schools to focus almost exclusively on what is tested.

The panel followed with this recommendation: “The federal government should support a program of research and development of the next generation of assessment tools and strategies for accountability systems” (NAE, 2008).

Palm (2008) maintained that performance assessment “is viewed as having better possibilities to measure complex skills and communication, which are considered important competencies and disciplinary knowledge needed in today’s society.” In short, performance assessments are better suited for measuring the attainment of 21st century skills than are traditional assessments.

However, Gewertz (2008) noted, “assessing students’ grasp of 21st century skills is tricky.” Critics of performance assessment routinely call attention to the fact that scoring performance assessments can be highly subjective (Liskin-Gasparro, 1997). Even though developing functional scoring rubrics or other standards for evaluating performance assessments is an achievable task, applying the standards consistently across a group of oral performances, research projects, or portfolios can be difficult. The task becomes Herculean when the group includes every student in a particular grade level across a large school division.

Developing Performance Assessments

The development of performance assessments involves a general process that has been described by a number of authors (Allen, 1996; Brualdi, 2000; Herman, Aschbacher, & Winters, 1992; Moskall, 2003). The three basic steps in this process—defining the purpose, choosing the activity, and developing the scoring criteria—will be explained in the next sections.

Defining the Purpose

The first step in developing performance assessments involves determining which concepts, knowledge, and/or skills should be assessed. The developer needs to know what type of decisions will be made with the information garnered from the assessment. Herman et al. (1992) suggested that teachers ask themselves five questions as they narrow down the myriad of possible learning objectives to be considered:

- *What important cognitive skills or attributes do I want my students to develop?* (e.g., communicate effectively in writing, employ algebra to solve real-life problems)
- *What social and affective skills or attributes do I want my students to develop?* (e.g., work independently, appreciate individual differences)

- *What metacognitive skills do I want my students to develop?* (e.g., reflect on the writing process, self-monitor progress while working on an independent project)
- *What types of problems do I want them to be able to solve?* (e.g., perform research, predict consequences)
- *What concepts and principles do I want my students to be able to apply?* (e.g., understand cause-and-effect relationships, use principles of ecology and conservation)

The initial step in developing performance assessments is analogous to the first stage in the “backward design” model espoused by Grant Wiggins and Jay McTighe (2005) in their book, *Understanding by Design*. The questions posed by Wiggins and McTighe in Stage 1 (Identify Desired Results) include these: “What should students know, understand, and be able to do? What content is worthy of understanding? What *enduring* understandings are desired?” For both backward design and performance assessment, the priority in the first step is establishing a clear focus for both instruction and assessment in terms of measurable objectives.

Choosing the Activity

The next step in the development of a performance assessment is to select the performance activity. Brualdi (2000) reminded teachers that they should first consider several factors, including available resources, time constraints, and the amount of data required to make an adequate evaluation of the student’s performance. In her synthesis of the literature on developing classroom performance assessments, Moskall (2003) made several recommendations:

- *The selected performance should reflect a valued activity* (i.e., a real-life situation).
- *The completion of performance assessments should provide a valuable learning experience*. Since performance assessments typically require a greater investment in time than traditional assessments, there should be a comparable payoff for students in terms of acquired knowledge and for teachers in their understanding of the students’ knowledge.
- *The statement of goals and objectives should be clearly aligned with the measurable outcomes of the performance activity*. The elements of the activity must correspond with the objectives that were specified in the first step (i.e., defining the purpose).
- *The task should not examine extraneous or unintended variables*. Students should not be required to possess knowledge that is not relevant to the activity’s purpose in order to complete the task.
- *Performance assessments should be fair and free from bias*. Activities that give some students an unfair advantage over other students should not be selected. (The example given by Moskall was an activity that included baseball statistics, which might penalize students who are not knowledgeable about baseball.)

The five recommendations above are inherently related to the validity of the performance assessment. Validity is defined as “the extent to which a test does the job for which it is used” (Payne, 2003). It is “the most important single attribute of a good test” (Lyman, 1998). Due to the increasingly popularity of performance assessments and their “potential benefits . . . validity issues need to be addressed through multiple lines of inquiry” (Randhawa & Hunter, 2001).

Publishers of nationally-normed, standardized tests go to great lengths to acquire validity evidence for their products. If the validity of a performance assessment is not established, then the interpretation and uses of the assessment’s results will be invalid. To obtain evidence of

content validity, assessments should be reviewed by qualified content experts. A content expert is “someone who knows enough about what is to be measured to be a competent judge” (Fraenkel & Wallen, 1996). Each content expert is then tasked with determining if the performance activity matches the learning objective(s) it was intended to measure. Rubrics designed to score performance tasks and products should also be reviewed for content validity. The development of rubrics will be discussed in the next section.

Developing the Scoring Criteria

The last step in constructing a performance assessment is developing the scoring criteria. While traditional assessments are comprised mostly of items for which the answer is either right or wrong, the difference is not as clear-cut with performance assessments (Brualdi, 2000). Rubrics are used to evaluate the level of a student’s achievement on various aspects of a performance task or product. A rubric can be defined as “a criterion-based scoring guide consisting of a fixed measurement (4 points, 6 points, or whatever is appropriate) and descriptions of the characteristics for each score point. Rubrics describe degrees of quality, proficiency, or understanding along a continuum” (Wiggins & McTighe, 2005).

Before creating or adopting a rubric, it must be decided whether a performance task, performance product, or both a task and product will be evaluated. Moskal (2003) explained that two types of rubrics are used to evaluate performance assessments: “Analytic scoring rubrics divide a performance into separate facets and each facet is evaluated using a separate scale. Holistic scoring rubrics use a single scale to evaluate the larger process.” Moskal’s six general guidelines for developing either type of rubric are as follows:

- *The criteria set forth within a scoring rubric should be clearly aligned with the requirements of the task and the stated goals and objectives.*
- *The criteria set forth in scoring rubrics should be expressed in terms of observable behaviors or product characteristics.*
- *Scoring rubrics should be written in specific and clear language that the students understand.*
- *The number of points that are used in the scoring rubric should make sense.*
- *The separation between score levels should be clear.*
- *The statement of the criteria should be fair and free from bias.*

When creating analytic scoring rubrics, McTighe (1996) has noted that teachers can allow students to assist, “based on their growing knowledge of the topic.” There are other practical suggestions to consider when developing rubrics. Stix (1997) recommended using “neutral words” (e.g., novice, apprentice, proficient, distinguished; attempted, acceptable, admirable, awesome) instead of numbers for each score level to avoid the perceived implications of good or bad that come with numerical scores. Another suggestion from Stix was to use an even number of score levels to avoid “the natural temptation of instructors—as well as students—to award a middle ranking.” For analytic rubrics, sometimes it is necessary to assign different weights to certain components depending on their importance relative to the overall score. Whenever different weighting is used on a rubric, the rationale for this must be made clear to stakeholders (Moskal, 2003).

Gathering evidence of content validity is critical for both performance assessments and rubrics, but it is also vital that rubrics have a high degree of reliability. Without a reliable rubric,

the interpretation of the scores resulting from the performance assessment cannot be valid. Herman et al. (1992) emphasized the importance of having “confidence that the grade or judgment was a result of the actual performance, not some superficial aspect of the product or scoring situation.” Scoring should be consistent and objective when individual teachers use a rubric to rate different students’ performance tasks or products over time. In addition, a reliable rubric should facilitate consistent and objective scoring when it is used by different raters working independently.

In order to avoid “capricious subjectivity” and obtain consistency for an individual rater as well as inter-rater reliability among a group of raters, extensive training is required for administering performance assessments and using rubrics within a school or across a school division. “Rater training helps teachers come to a consensual definition of key aspects of student performance” (Herman et al., 1992). Training procedures include several steps:

- *Orientation to the assessment task*
- *Clarification of the scoring criteria*
- *Practice scoring*
- *Protocol revision*
- *Score recording*
- *Documenting rater reliability*

Despite the fact that developing rubrics and training raters can be a complicated process, the ensuing rewards are worth the effort. Perhaps the greatest value of rubrics is in these two features: (1) “they provide information to teachers, parents, and others interested in what students know and can do,” and (2) “*promote learning by offering clear performance targets to students for agreed-upon standards*” (Marzano, Pickering, & McTighe, 1993).

Other Considerations

Performance assessments should always be field-tested before they are fully implemented in schools. As Wiggins warned, “Unpiloted, one-event testing in the performance area is even more dangerous than one-shot multiple-choice testing” (Brandt, 1992). Invaluable feedback from the persons who administer and score the assessments, as well as from the students themselves, can be obtained in pilot studies. Field-testing can provide evidence of whether the performance activity is biased or assesses any unintended variables. Additionally, Roeber (1996) maintained that, although writing the directions for performance assessments can be difficult, it is more easily facilitated after field-testing the assessment. Test administrators are instructed to note “areas of . . . confusion, responses that students provided which are vague and incomplete, and ways in which some or all of the students responded that were not anticipated.”

Performance assessments and rubrics have been used, revised, and reused by educators for over two decades, and the subsequent paper trail is extensive. There is a seemingly endless supply of performance assessments and rubrics that are available commercially or at no cost from various online sources. For example, the website for *Jay McTighe and Associates Educational Consulting* includes a webpage with numerous links to performance assessments and rubrics. To view these, go to <http://www.jaymctighe.com/ubdweblinks.html> and click on “Performance Assessments,” “Rubrics,” or “Subject Specific Rubrics.” A worthwhile resource that can be used for evaluating rubrics is *A Rubric for Rubrics* (Mullinax, 2003), which can be accessed at http://tltgroup.org/Mullinix/Rubrics/A_Rubric_for_Rubrics.htm. There is no need to

“reinvent the wheel” when it comes to performance assessments; however, the processes required to ensure valid and reliable results from the assessments involve a great deal of time and attention to detail.

A Balanced Assessment System

During a presentation at a recent conference, renowned testing authority Rick Stiggins (2008a) stated the following:

We have come to a tipping point in American education when we must change our assessment beliefs and act accordingly, or we must abandon hope that all students will meet standards or that the chronic achievement gap will close. The troubling fact is that, if all students don't meet standards—that is, if the gap doesn't close between those who meet and don't meet those standards—our society will be unable to continue to evolve productively in either a social or an economic sense. Yet, paradoxically, assessment as conceived, conducted, and calcified over the past has done as much to perpetuate the gap as it has to narrow it. This must change now and it can. As it turns out (again paradoxically), assessment may be the most powerful tool available to us for ensuring universal student mastery of essential standards.

In other words, assessment is not only part of the problem; it is also an important component of the solution. High-quality performance assessments have the potential to play a key part in American K-12 education's progress towards positive change.

According to Stiggins (2008a), Americans “have invested literally all of our resources in once-a-year testing for decades.” In 2003, the National Education Association (NEA) reported that most assessment systems in the U.S. were “out of balance.” More recently, education professionals and policy makers have recognized the importance of “the appropriate and effective use of a variety of assessment processes, all of which should serve student learning” (Redfield, Roeber, & Stiggins, 2008). This realization has led to a call for school divisions to implement balanced assessment systems to guide educational improvement in the 21st century.

A balanced assessment system is comprised of formative and summative assessments administered on both a large scale and at the classroom level. In this context, “balanced” does *not* refer to assessments that are of equal weight (Redfield, Roeber, & Stiggins, 2008). A balanced assessment system is founded on the belief that the primary purpose of K-12 education is to maximize achievement for all students, and that different types of assessment can be used to support instruction. Traditional assessments and performance assessments that yield accurate information in a timely manner all have a place in a balanced assessment system (NEA, 2003). In his *Assessment Manifesto*, Stiggins (2008b) explained, “Truly productive assessment systems within schools and districts serve the information needs of a wide variety of assessment users.”

Performance assessment tasks and products not only inform educators of students' progress towards predetermined objectives, they also provide students and parents with meaningful feedback about the ability of these students to perform successfully in real-life situations. Although there is justifiable concern that policy makers and the general public—as well as some educators—will not readily accept performance assessment as a viable complement to more traditional methods of assessment, there are indications that positive attitudes toward

performance assessment can be procured. Meisels, Xue, Bickel, Nicholson, and Atkins-Burnett (2001) cited evidence of parental support for performance assessment in previous research, and reported that the results of their own study indicated most parents preferred performance assessment summaries to traditional report cards. These researchers further stated that if performance assessment is “ever to become more generally accepted by parents and policy makers, it is essential that parents’ reactions be taken into account and shaped through positive and informative interactions with teachers and other educators.”

Summary

Performance assessment can be defined as a method of evaluating students’ knowledge, concepts, or skills by requiring them to perform a task designed to emulate real-life contexts or conditions in which students must apply the specific knowledge, concepts, or skills (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; U.S. Department of Education, 1993).

Not only does performance assessment allow students to demonstrate their abilities in a more genuine context than is required by other types of assessment, performance assessment has other advantages over the traditional assessments that are more commonly used in schools today. Students are able to recognize real-life connections with performance assessments. Additionally, students are generally more motivated by high-quality performance assessments, which have the capacity to measure higher-order thinking skills and other abilities needed to achieve success in the contemporary workplace.

However, a great deal of time and effort must be invested to ensure that performance assessments and the rubrics used to score them are reliable and yield valid results. Additional time must be devoted to professional development for educators and efforts to familiarize parents with this innovative assessment concept. Although performance assessments will never completely replace traditional tests, they can be effectively utilized by schools and divisions to complement other types of assessment within the framework of a balanced assessment system.

References

- Allen, R. (1996). *Performance Assessment*. Wisconsin Education Association Council. Retrieved January 13, 2009, from <http://www.weac.org/resource/may96/perform.htm>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Association for Supervision and Curriculum Development. (2006). Education Topics: Performance Assessment – What Is Performance Assessment? Retrieved November 6, 2008, from http://www.ascd.org/research_a_topic/Performance_Assessment/Performance_Assessment_-_Expert_1.aspx
- Brandt, R. (1992). On Performance Assessment: A Conversation with Grant Wiggins. *Educational Leadership*, 49(8), 35-37.
- Brandt, R. (1998). Foreword. In G. Wiggins and J. McTighe, *Understanding by Design* (pp. v-vi). Alexandria, VA: Association for Supervision and Curriculum Development.
- Brualdi, A. (1998). Implementing Performance Assessment in the Classroom. *Practical Assessment, Research & Evaluation*, 6(2). Retrieved August 7, 2008, from <http://PAREonline.net/getvn.asp?v=6&n=2>
- Fraenkel, J. R., & Wallen, N. E. (1996). *How to Design and Evaluate Research in Education* (3rd ed.). New York: McGraw-Hill.
- Gewertz, C. (2008). States Press Ahead on '21st-Century Skills.' *Education Week*, 28(8), 21-23.
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). Perceptions of Authentic Assessment: Five Dimensions of Authenticity. Paper presented at the Second Biannual Joint Northumbria/European Association for Research on Learning and Instruction SIG Assessment Conference, Bergen, Norway. Retrieved December 7, 2008, from <http://www.ou.nl/Docs/Expertise/OTEC/Publicaties/judith%20gullikers/paper%20SIG%202004%20Bergen.pdf>
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising Standardized Achievement Test Scores and the Origins of Test Score Pollution. *Educational Researcher*, 20(5), 2-7.
- Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Herrington, J., & Herrington, A. (1998). Authentic Assessment and Multimedia: How University Students Respond to a Model of Authentic Assessment [Electronic version]. *Higher Education Research and Development*, 17(3), 305-322. Retrieved November 6, 2008, from http://edserver2.uow.edu.au/~janh/assessment/authentic%20assessment_files/herdsa.pdf

- Kahl, S. (2008). *The Assessment of 21st Century Skills: Something Old, Something New, Something Borrowed*. Paper presented at the Council of Chief State School Officers 38th National Conference on Student Assessment, Orlando, FL.
- Liskin-Gasparro, J. (1997). *Comparing Traditional and Performance-Based Assessment*. Paper presented at the Symposium on Spanish Second Language Acquisition, Austin, TX. Retrieved December 30, 2008, from http://sedl.org/loteced/comparing_assessment.html
- Lyman, H. B. (1998). *Test Scores and What They Mean* (6th ed.). Boston: Allyn and Bacon.
- Marzano, R. J., Pickering, D., & McTighe, J. (1993). *Assessing Student Outcomes: Performance Assessment Using the Dimensions of Learning Model*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McTighe, J. (1996). What Happens Between Assessments? *Educational Leadership*, 54(4), 6-12.
- Mehrens, W. A. (1991). Defensible/Indefensible Instructional Preparation for High Stakes Achievement Tests: An Exploratory Dialogue. Paper presented at the Annual Meetings of the Educational Research Association and the National Council on Measurement in Education, Chicago, IL.
- Meisels, S. J., Xue, Y., Bickel, D. D., Nicholson, J., & Atkins-Burnett, S. (2001). *Parental Reactions To Authentic Performance Assessment*. Ann Arbor, MI: University of Michigan, Center for the Improvement of Early Reading Achievement. Retrieved December 31, 2008, from <http://www.ciera.org/library/archive/2001-06/0106prmx.pdf>
- Moskal, B. M. (2003). Recommendations for developing classroom performance assessments and scoring rubrics. *Practical Assessment, Research & Evaluation*, 8(14). Retrieved January 13, 2009, from <http://PAREonline.net/getvn.asp?v=8&n=14>
- Mueller, J. (2008). Authentic Assessment Toolbox: What is Authentic Assessment? Retrieved November 6, 2008, from <http://jonathan.mueller.faculty.noctrl.edu/toolbox/whatisit.htm>
- Mullinax, B. B. (2003). *A Rubric for Rubrics*. The TLT Group. Retrieved December 30, 2008, from http://tltgroup.org/Mullinax/Rubrics/A_Rubric_for_Rubrics.htm
- National Academy of Education. (2008). *Recovering the Promise of Standards-Based Education*. Education Policy Briefing Sheet presented at the National Academy of Education, Education Policy in Transition Public Forum, Washington, DC. Retrieved December 4, 2008, from http://naeducation.org/White_Papers_Project_Standards_Assessments_and_Accountability_Briefing_Sheet.pdf
- National Education Association. (2003). *Balanced Assessment: The Key to Accountability and Improved Student Learning* (Student Assessment Series). Retrieved November 6, 2008, from <http://www.assessmentinst.com/forms/nea-balancedassess.pdf>
- Palm, T. (2008). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment, Research & Evaluation*, 13(4), 1-11. Retrieved December 29, 2008, from <http://pareonline.net/pdf/v13n4.pdf>

- Payne, D. A. (2003). *Applied Educational Assessment* (2nd ed.). Belmont, CA: Wadsworth.
- Popham, W. J. (2001). Teaching to the Test? *Educational Leadership*, 58(6), 16-20.
- Randhawa, B. S., & Hunter, D. M. (2001). Validity of Performance Assessment in Mathematics for Early Adolescents [Electronic version]. *Canadian Journal of Behavioural Science*, 33(1), 14-24. Retrieved November 6, 2008, from http://findarticles.com/p/articles/mi_qa3717/is_200101/ai_n8945122
- Redfield, D., Roeber, E., & Stiggins, R. (2008). *Building Balanced Assessment Systems to Guide Educational Improvement*. Paper presented at the Council of Chief State School Officers 38th National Conference on Student Assessment, Orlando, FL. Retrieved June 24, 2008, from <http://www.ccsso.org/content/PDFs/OpeningSessionPaper-Final.pdf>
- Roeber, E. D. (1996). Guidelines for the Development and Management of Performance Assessments. *Practical Assessment, Research & Evaluation*, 5(7). Retrieved January 13, 2009, from <http://PAREonline.net/getvn.asp?v=5&n=7>
- Stiggins, R. J. (2008a). *Assessment FOR Learning, the Achievement Gap, and Truly Effective Schools*. Presentation given at the Educational Testing Service and College Board Conference, Educational Testing in America: State Assessments, Achievement Gaps, National Policy and Innovations, Washington, DC. Retrieved December 31, 2008, from http://www.ets.org/Media/Conferences_and_Events/pdf/stiggins.pdf
- Stiggins, R. J. (2008b). *Assessment Manifesto: A Call for the Development of Balance Assessment Systems*. Portland, OR: Educational Testing Service, Assessment Training Institute.
- Stix, A. (1997). *Empowering Students Through Negotiable Contracting*. Paper presented at the National Middle School Initiative Conference, Long Island, NY. (ERIC Document Reproduction Service No. ED411274). Retrieved January 15, 2009, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/14/f3/c8.pdf
- U.S. Department of Education, Office of Educational Research and Improvement. (1993). *Consumer Guide: Performance Assessment* (ED/OERI 92-38). Retrieved November 6, 2008, from <http://www.ed.gov/pubs/OR/ConsumerGuides/perfasse.html>
- Virginia Beach City Public Schools. (2008). *Compass to 2015: A Strategic Plan for Student Success*. Retrieved November 6, 2008, from http://www.vbschools.com/strategic_plan/index.asp
- Wiggins, G. (1992). Creating Tests Worth Taking. *Educational Leadership*, 49(8), 26-33.
- Wiggins, G., & McTighe, J. (2005). *Understanding by Design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.