2004

# Alignment of Classroom Assessments with the Virginia Standards of Learning (SOL)

Joan A. Rhodes
*Virginia Commonwealth Univer*, jarhodes2@vcu.edu

Valerie J. Robnolt
*Virginia Commonwealth University*, vjrobnolt@vcu.edu

Susan P. McKelvey
*Virginia Commonwealth University*, smckelvey@sunllc.com

# Alignment of Classroom Assessments with the Virginia Standards of Learning (SOL)

Joan A. Rhodes
Assistant Professor
Department of Teaching and Learning

Valerie J. Robnolt
Assistant Professor
Department of Teaching and Learning

Susan McKelvey
MERC Research Assistant

## Introduction

Parents, teachers, and school administrators all have one common goal – student success. Teachers have been trusted to impart knowledge to students, while hopefully fostering a love of learning, then assessing how much the students know. Assessment in the classroom, whether informal, formative, or summative, has been the major source of identifying student strengths and weaknesses. Today, many people believe that teachers must be accountable for what they do in the classroom. Within the educational system, students' success has been measured inconsistently from state-to-state, district-to-district, and even classroom-to-classroom. In this age of accountability, national and state standards have been developed, and through state-wide assessments, students are tested on what they have learned. Each state is accountable to the United States Department of Education, administrators are accountable to their school district and state, and teachers are accountable to the administrators. Today, all American students are expected to meet very specific standards. Teachers must assess their students' knowledge and abilities through testing in the classroom.

## Brief History of Standards-Based Education

Educational reform can be traced back to the 19th century, with the testing movement blossoming after World War II. During the 1960s, there was a great deal of focus on compensatory education because of the perceived disparities in educational opportunities as well as student performance. At this point, the Elementary and Secondary Education Act (ESEA), supported by the federal government, expanded testing because Congress needed to evaluate how funds were being distributed (Linn & Gronlund, 2000).

Standardized tests have been the "focus of controversy" because it has been said that, "they drive instruction in undesirable ways by overemphasizing factual knowledge and low-level

skills" (Linn & Gronlund, 2000, p. 2). Despite this controversy, *A Nation at Risk* (U.S. Department of Education, 1983) examined school-level accountability. Studies indicated that tests could be used to identify weaknesses of students, and because of this, they could be used to facilitate change. During this time, every state attempted to reform on some level, with many states using external assessments. Everyone, from administrators to students to teachers, felt the pressure of accountability. The research during this era reported that teachers were only teaching test-taking skills and the content covered on the assessments. They even went as far as using the external assessments to drive the format for their classroom tests (Linn & Gronlund, 2000).

Today, standards drive accountability through high-stakes testing to ensure that content is adequately covered by teachers. "Standards are statements that specify what should be taught and what students should learn" (Linn & Gronlund, 2000, p. 6). The most recent reform efforts fall under the *No Child Left Behind Act of 2001* (NCLBA), which was a reauthorization of the ESEA. The four main components of NCLBA are accountability, decision making based on scientific research, increased local control, and more options for parents. Reading during the elementary grades is emphasized under the premise that early intervention will prevent reading problems for children in later years (U.S. Department of Education, 2003).

In addition to improving reading skills, children will be tested in reading and math every year from grades 3 until 8. They will also be assessed a minimum of one time during high school. NCLBA requires alignment between state tests and the state standards. Schools are expected to make adequate yearly progress (AYP) on these state tests. Those that fail to make AYP will be subject to a series of improvement measures in order to meet the state requirements. If a school does not make AYP, parents will have the option of transferring their children to another school or receiving extra instruction. Teachers and teacher aides will be assessed as well.

Every professional and paraprofessional must meet the criterion of "highly qualified" by the year 2006. District and school-level "report cards," including information on teacher qualifications, will be available to the public. Fortunately, federal expenditures for schools have greatly increased to ensure that districts, schools, teachers and students meet the goals of the *NCLBA* (U.S. Department of Education, 2003).

## Alignment

What, exactly, is alignment? In reference to education, a single definition does not exist. However, Impara (2001) gives three possible definitions. A basic definition, which he quoted from LaMarca, Redfield, Winter, Bailey, and Despriet (as cited in Impara, 2001), is the "match between two or more things" (p. 3). Another definition is "the extent to which the items (or tasks/prompts) on a test match relevant content standards (or test specifications)" (p. 3). The final definition Impara provided is one developed by Webb (as cited in Impara, 2001) which states that alignment is "the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide the system toward students learning what they are expected to know and do" (p. 4). Impara outlined three types of alignment models, ranging from low complexity to high complexity, which will be detailed later.

With these three definitions, alignment can encompass content match, format match, and cognitive level match. The state of Virginia has developed Standards of Learning (SOL) as a curriculum framework in the four core areas of English, mathematics, history/social science, and science. Teachers are expected to align classroom instruction with the SOL so students possess specific knowledge and skills in these four areas. As teachers construct classroom assessments, there is an implication that teachers will align tests with the SOL (Virginia Department of Education, 2004). When aligning the content of their tests, teachers should take into account

whether each test item coincides with the standards the students are expected to know. In terms of format, the way the question is asked should be appropriate for the content of the standard (Linn & Gronlund, 2000). The cognitive level should reflect the complexity of the content. There are differing methods of analyzing the hierarchical nature of cognitive levels, beginning with the seminal work of Bloom.

*Bloom's Taxonomy*

Bloom (Distance Learning Resource Network, 2003) categorized learning into three domains: cognitive, affective, and psychomotor. The cognitive domain is of interest for the purpose of the present study. Within this domain, Bloom delineates six levels which are arranged from the lowest to the highest: (a) knowledge, (b) comprehension, (c) application, (d) analysis, (e) synthesis, and (f) evaluation.

The first level, knowledge, simply refers to facts or content students can learn through rote memorization. In the second level, comprehension, students would be asked to describe or identify something. The third level, application, expects students to apply concepts they have learned. For example, they might be asked to solve a problem using a new skill. At the analysis level, students might criticize an article or compare and contrast. In the fifth level of the cognitive domain, synthesis, students design, construct, or develop responses. For the last level, evaluation, students must support opinions or rate information (Distance Learning Resource Network, 2003).

*Revision of Bloom*

Anderson and Krathwohl (2001) believed that Bloom's Taxonomy was dated, and decided to take on the task of revising it. They made revisions with the intent of using the new taxonomy for aligning all aspects of instruction. In Bloom's original Taxonomy, the six major

categories were emphasized, even though he had developed sub-categories. The revised taxonomy pulls the sub-categories to the forefront, allowing for ease of use in the classroom. Because of the emphasis on utility of the revisions, the authors framed everything in a "verb-noun relationship" (p. 307).

The knowledge category has now become its own dimension with four categories: factual, conceptual, procedural, and metacognitive. The new cognitive domain still has six categories, with several sub-categories. The categories and sub-categories are: (a) "Remember – recognizing, recalling;" (b) "Understand – interpreting, exemplifying, classifying, summarizing, inferring, comparing, explaining;" (b) "Apply – executing, implementing;" (c) "Analyze – differentiating, organizing, attributing;" (d) "Evaluate – checking, critiquing;" and (e) "Create – generating, planning, producing" (p. 31). Anderson and Krathwohl believed that teachers can determine which level they are reaching by identifying the nouns and verbs in their objectives. The noun describes the category of knowledge they are assessing, and the verb defines the category and sub-category of the cognitive domain they have reached. Anderson and Krathwohl noted that teachers can create assessments by examining the categories covered in classroom instructional objectives.

## Alignment Research

Even before *NCLBA* and the requirement of scientific research, several studies were conducted concerning alignment with standards. Many of these studies focused on aligning the standards with state assessments, but some examined the relationship between the curriculum and the standards. Mid-continent Research for Education and Learning [McREL] (2000) gathered qualitative data from teachers for their study. The sample included beginning teachers as well as veterans who were involved in the reform efforts. They stated that they had been given

a list of standards, with no help or explanation. How then, were these teachers supposed to know the importance of aligning their instruction with the standards?

After being involved in the reform, the teachers realized how crucial it was to integrate all aspects of their instruction with the standards. The teachers also realized that standards needed to be written in understandable and measurable ways. This research suggested that once teachers have clear standards, they needed to use a variety of assessment types because "standards require that students have deeper levels of knowledge about content matter and are able to apply that knowledge" (McREL, 2000, p. 21). Typical assessment methods, such as true-false items and multiple-choice questions, "are not adequate for assessing this type of knowledge and its application" (McREL, 2000, p. 21). Teachers need to understand the type of knowledge they are assessing then choose an appropriate corresponding assessment. For example, facts may call for a matching exercise or multiple choice, while a question about generalizations may warrant a writing sample.

McREL (2000) also cited a study conducted in Philadephia by Simon et al. in 1998. When stating objectives in their classrooms, the teachers in this study referred more to the standardized state tests than the actual standards themselves. In this era of accountability, the one aspect that teachers are most concerned about is assessment. In their study, McREL found that the teachers emphasized the importance of being able to work together. The teachers in this study hoped that administrators would allow time for this, as well as, afford them opportunities to observe other teachers who were known to improve their instruction through the lens of state standards.

In a two-year study by Blank, Porter and Smithson (2001), the researchers examined the "enacted curriculum," which they defined as "the actual subject content and instructional

practices experienced by students in the classrooms" (p.i). Over 600 teachers from schools in 11 different states participated in this study. They decided that the K-12 curriculum could be divided into three parts: (a) the intended curriculum, (b) the enacted curriculum, and (c) the learned curriculum (student outcomes). They observed two teachers per grade level and subject in their classrooms. They also administered a 150 item survey which asked about instructional practices, subject content, and teacher characteristics.

Blank, Porter and Smithson (2001) found that classroom instruction did not align with the state assessments (standards were not examined) for the 11 states. On the state tests, only a few sub-topics were covered whereas many more topics were covered during instruction. In addition, while many sub-topics were covered in the classroom, they were covered with no depth. In a general sense, another result of their study was the creation of alignment analyses which teachers and policy makers could use.

In a third study, conducted by the Regional Educational Laboratory (2000), the researchers examined aspects of effective reform at the state, district and school levels. They found that Massachusetts had frameworks to align instruction and standards and a collaboration system between institutions of higher education and new teachers. Florida had a similar arrangement with the local universities so that new teachers developed skills in understanding standards, assessment and curriculum. Kentucky, Oregon and Maryland were also involved in reform. These three states focused on reforming only one or two subjects per year. The Regional Educational Laboratory researchers found that, at the school level, one of the key characteristics of reform was to closely monitor student learning through classroom assessments. The schools were relying heavily on test data.

Finally, a three-phase qualitative study was conducted by Hammrich (1999) with K-8 science teachers as they participated in a professional development program that lasted for one year. In Phase III, the teachers examined different tests to determine a match between the content and the standards. When it was over, the teachers stated that they recognized the importance of alignment; however, they believed that it was not worth the effort (Hammrich, 1999).

For all levels of reform (state, district, and school), three commonalities existed. They were "using student learning standards as a foundation for reform work, enhancing teachers' capacity to provide effective instruction, and providing meaningful assessment of student achievement" (Regional Educational Laboratory, 2000, p. 18). The researchers believed that each level greatly influenced the others. To implement effective reform in the educational system, one key issue is to align all aspects of curriculum, instruction and assessment.

Even before the *NCLBA*, many people had strong opinions concerning standards and testing. One very important aspect of having set learning standards is to make sure that they drive instruction. Bruner and Greenlee (2002) stated that, "By providing outlines for curriculum content and benchmarks that specify levels of performance and expectations, standards can turn educational goals into concrete and measurable criteria" (p. 23). But how should teachers use the standards to improve their instruction? Bruner and Greenlee believed that teachers should be given the opportunity to work with colleagues to discuss their grade-level standards. Administrators must support this by allowing time for collaboration, as well as giving teachers permission to change the order of content taught during the year. By having the chance to work with other teachers on aligning curriculum, they would be able to convert the standards into classroom activities. What teachers do in the classroom and how they assess students could make the difference in their standardized test scores.

Squires (1998) took a different approach by examining alignment both across and between grade levels. He called this horizontal and vertical alignment and suggested using state test results to improve this alignment. In agreement with Bruner and Greenlee (2002), Squires emphasized the importance of teacher collaboration. In his alignment work, Squires assessed each unit for both content and format. If the classroom assessments were aligned in both aspects, it was anticipated that student scores would improve on the state tests with teachers having to do little, if any, standardized test preparation.

Like Squires, Aviles (2001) discussed vertical and horizontal alignment. However, according to Aviles, vertical alignment pertains to cognitive level while horizontal alignment encompasses content. He concluded that to align a curriculum, all of the materials, including the test, need to be created before planning the unit.

In their article on alignment, La Marca, Redfield, and Winter (2000) focused on state-wide assessments, but suggested that their guidelines could be used for any type of evaluation. They emphasized the importance of matching the assessment to the standards. Their "General Organizing Principles" consisted of content match, depth match, emphasis, performance match and accessibility. To make these matches, the test developer must assume several things, including that instruction reflects the standards. A second assumption is that the assessments need to be aligned with the standards to use the data to make instructional decisions. LaMarca et al. (2000) suggested using a third party to develop an alignment checklist to evaluate the degree of the match between standards and assessments.

Tienken and Wilson (2001) discussed a program in New Jersey that helped teachers improve their instructional practices through their understanding of the standards. This was done through delineation (i.e., where teachers recognize all aspects of content, including state

standards), alignment and calibration. During the alignment phase, teachers created classroom tests, and addressed whether or not they examined the correct skills, matched the cognitive level of the standard, and, less importantly, duplicated the format. In the alignment phase, Tienken and Wilson (2001) stated that 100% alignment should not be the goal.

## Alignment Models

As previously mentioned, Impara (2001) outlined three types of alignment models. He stated that teachers should not be surprised when they receive their students' scores on state assessments. If teachers have been using appropriate classroom tests, they should be able to gauge how well their students will do. Again, the teachers' tests should be aligned with the state's standards. Impara gave three levels for aligning assessments, Low Complexity, Moderate Complexity, and High Complexity, and provided examples for each level. For the lowest level, content specialists are asked to look at an assessment and give a score for each item ranging from "no match" to "exact match." The content specialist may also be asked to give a score for both content and cognitive level. The next two levels are based on the low complexity model. The moderate level focuses on content match, depth match, emphasis, performance match, and accessibility, similar to the model discussed earlier. The highly complex model focuses on content, articulation across grades, equity and fairness, pedagogical implications, and system applicability. Impara stated that the most complex model could be applied to any testing situation; however, he also noted that there were problems with all of the models he outlined, which is discussed later in the review.

Impara (2001) cited Webb's alignment research as a good example of a highly complex model. To determine alignment of content, Webb (as cited in Impara, 2001) believed that examining the "depth of knowledge" was the first step (p. 5). He used a hierarchy similar to

Bloom's Taxonomy, which divided the cognitive level into four categories. In his research, Webb asked content specialists to use actual assessments and determine under which of his four cognitive levels each standard belonged. The content experts rated each assessment item using the same cognitive demand scale as the standard to see if they matched. Webb made the determination that at least one half of the test items had to be at the same or higher cognitive level than the one at which the standard was measured. If this one-half criterion was met, then the test measured the students at the same cognitive level as the standard. "If, for example, there were six items related to a standard and the passing level was four items answered correctly, then at least three of the items had to match the level of cognitive complexity for the indicators to have an acceptable depth of knowledge rating" (as cited in Impara, 2001, p. 5).

The last three steps for determining match were not as stringent as the first. The second step in Webb's research was to assess what he called categorical congruence. For example, if a standard calls for correct spelling, then the teacher (or scoring rubric on the state assessment) must consider a spelling score as part of the student's test score. The third step was to examine the "range of knowledge" (as cited in Impara, 2001, p. 5). One standard might have several dimensions or objectives that are related. For instance, students may be expected to utilize writing concepts, which would also include spelling, grammar, etc. Therefore, assessments should look at all aspects of writing to glean an alignment match. In the last step, the raters considered the objectives related to the standards and gave a score that represented whether or not the objectives had been equally weighted on the test.

The problems with any of the alignment models that Impara (2001) discussed in his paper are as follows. Impara believed that the terms, such as standard or objective, were too broad because the terms may have multiple dimensions. In determining alignment, one dimension may

be met, which led to the rater stating the item was aligned when only one aspect of the standard or objective was aligned. Another problem was that standards differ from state to state. Even if a standard looks very similar from one state to another, the curriculum requirements may not be the same. Thus, alignment in one state does not guarantee alignment in another state. The final problem that Impara considered with the alignment models was that states required students be classified into competency levels. To determine competency accurately it would take more than one test, resulting in a very expensive and time-consuming process.

## Guidelines and Standards

Linn and Gronlund (2000) developed some guidelines for creating assessments, without addressing alignment. Several questions should be answered, including "Is the format appropriate for the question?" and "Does the knowledge, understanding or thinking skill called forth by the item or task match the specific learning outcome and subject-matter content being measured?" (p. 346-347). They also developed "Standards for Teacher Competence in Educational Assessment of Students" of which three apply to the present study. They are: (a) "Teachers should be skilled in developing assessment methods appropriate for instructional decisions" (p. 538); (b) "The teacher should be skilled in administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods" (p. 539); and (c) "Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement" (p. 541).

From all of the articles and studies in this literature review, we can compile a list of guidelines for alignment. These guidelines are: (a) teachers should be knowledgeable about the standards with which they are aligning; (b) teachers should work collaboratively in developing

units, including classroom tests; (c) classroom tests should address the content as well as the cognitive level, possibly using Bloom's Taxonomy, for each standard; (d) teachers should have a variety of types of assessments including, but not limited to, multiple-choice; (e) teachers must create their assessments before teaching the units; and (f) teachers should think of teaching and testing together, not separately.

## Challenges

Regardless of models, standards and guidelines, challenges exist in aligning standards and assessments. McREL (2000) listed what they believe are the challenges to implementing effective standards-based education:

1. There is little agreement about what standards-based education looks like, so research and evaluation studies report vast differences in approaches.
2. The push for high-stakes accountability tends to limit the scope of standards-based curriculum and instruction.
3. There is no clear, comprehensive system of alignment between standards and benchmarks among local, state, and national standards.
4. Teachers do not have access to high-quality instructional resources that are aligned with local standards.
5. The assessment and record-keeping requirements for standards-based education place burdens on teachers.
6. There is limited information about standards-based approaches that can address the learning requirements of all students, especially students who do not meet standards.
7. More information is needed about the costs of implementing standards; support to cover the costs also is needed. (pp. 43-44)

Due to the implementation of *NCLBA*, the stakes of state assessment results are even higher for teachers in Virginia. There is a greater need to align all aspects of their instruction with the SOL and the state tests. Taking into account the research that has been reviewed regarding alignment, the following research questions will guide this study. To what extent do grades 3, 5, 8, and 10/11 English/Language Arts teachers, (a) align classroom assessment content with the Virginia SOL at their grade level, (b) align classroom assessment format with the English SOL assessments at their grade level, (c) align classroom assessment cognitive level with the English SOL assessments at their grade level and, (e) use SOL released items to align their classroom assessments with the English SOL assessments at their grade levels. In addition, to what extent do teachers of special populations align their classroom assessments similarly to regular education classroom teachers?

References

Anderson, L.W. & Krathwohl, D. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Addison Wesley Longman, Inc.

Aviles, C.B. (2001). *Curriculum alignment: Matching what we teach and test versus teaching to the test.* Buffalo, NY: Buffalo State College. (ERIC Document Reproduction Service No. ED448402)

Blank, R.K., Porter, A., & Smithson, J. (2001, July). *New tools for analyzing teaching, curriculum and standards in mathematics & science.* Results from Survey of Enacted Curriculum Project, Final Report. Washington, DC: Council of Chief State School Officers.

Bruner, D.Y., & Greenlee, B.J. (2002, January). Bringing standards from the state house to the schoolhouse. *Principal,* 23-25.

Distance Learning Resource Network (2003). *DLRN's technology resource guide* [Electronic Version]. Retrieved December 5, 2003 from http://www.dlrn.org/library/dl/guide4.html

Hammrich, P.L. (1999). *Science curriculum reform: What teachers are saying.* Philadelphia, PA: Temple University. (ERIC Document Reproduction Service No. ED432470)

Impara, J.C. (2001). *Alignment: One element of an assessment's instructional utility.* Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

La Marca, P.M., Redfield, D., & Winter, P.C. (2000). *State standards and state assessment systems:A guide to alignment. Series on standards and assessments.* Washington, DC: Council of Chief State School Officers.

Linn, R.L., & Gronlund, N.E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

Mid-continent Research for Education and Learning. (2000). *Noteworthy perspectives on implementing standards-based education.* Aurora, CO: Author.

Regional Educational Laboratory. (November 2000). *Implementing Education Reform: Strategies used by states, districts, and schools.* New York: Author.

Squires, D.A. (1998, Summer). Toward a balanced curriculum: Aligning standards, curriculum, and assessments. *ERS Spectrum,* 17-24.

Tienken, C., & Wilson, M. (2001). *State standards, assessments and instruction.* (ERIC Document Reproduction Service No. ED458215)

U.S. Department of Education. (1983, April). *A nation at risk.* Retrieved May 30, 2004, from http://www.ed.gov/pubs/NatAtRisk/risk.html

U.S. Department of Education. (n.d.). *No child left behind.* Retrieved December 5, 2003, from http://www.ed.gov/nclba/landing.jhtml

Virginia Department of Education (n.d.). *Standards of Learning currently in effect for Virginia Public Schools.* Retrieved June 1, 2004, from http://www.pen.k12.va.us