



# Alignment in Educational Assessment



Betsy J. Case, Ph.D.

Margaret A. Jorgensen, Ph.D.

Sasha Zucker

December 2004

# Alignment in Educational Assessment

## Introduction

In the context of education, *alignment* can be broadly defined as the degree to which the components of an education system—such as standards, curricula, assessments, and instruction—work together to achieve desired goals (Ananda, 2003; Resnick, Rothman, Slattery, and Vranek, 2003; Webb, 1997b). Most recently, alignment studies examine the degree to which standards and assessment address the same content (Webb, 1997a; Webb, 1999). Porter’s work on the enacted curriculum extends this to standards, assessments, and curriculum. Criterion-referenced tests focus on instructional objectives—small, discrete improvements of learning—not standards (Porter, 2002, 2004). With the passage of the *No Child Left Behind Act* of 2001 (NCLB) and the resulting era of accountability in education, the attention of education policymakers and researchers has recently turned to alignment. Conceptions of alignment have become increasingly sophisticated to meet goals for strengthening education systems and to satisfy rigorous requirements for accountability assessments mandated by NCLB. To provide an overview of this wide-ranging topic, this report discusses the background of alignment in educational assessment, traditional alignment methodologies, the three most frequently used alignment models, and broader issues of alignment in education.

## Alignment Background History

Alignment is not particularly new to the field of educational assessment (Bloom, Madaus, Hastings, 1981; Impara, 2001; Tyler, 1949; Webb, 1999). The alignment between an assessment and a set of content standards in a subject area has long been recognized as evidence of the assessment’s *validity*—the degree to which the interpretations of an assessment’s results can be considered accurate (AERA, APA, and NCME, 1999; Ananda, 2003; Impara, 2001; Resnick et al., 2003; Webb, 1997b). Hence, since NCLB, the process of establishing the relationship between a test and content standards is required for any assessment to be considered rigorous, high quality, and valid. Previous to the current era of accountability in education, the method for establishing alignment was not carefully defined. However, provisions in the two most recent reauthorizations of the *Elementary and Secondary Education Act* of 1965 (ESEA)—the *Improving Americas School Act* of 1994 (IASA) and NCLB—set the stage for the rigorous models of alignment in use today.

## **The Standards-based Reform Movement**

The central tenet of the standards-based reform movement is that stakeholders in an education system agree upon certain content standards—the expectations for what students should know and be able to do in a subject area (Jennings, 1998). Standards delineate the priorities for an education system and serve as a unifying guide for educators and students (Webb, 1997a). Adherents to the movement assert that establishing standards makes the system more effective and coherent, thereby supporting student learning and improving achievement (Ananda, 2003). Moreover, standards-based reform is promoted as more equitable because, theoretically, all students are taught the same academic content. However, the extent to which instructional equity exists in classrooms is not guaranteed by the adoption of this reform.

Standards-based accountability assessments have come to serve an important role in this reform movement. These assessments concretely embody the standards, providing an impetus for educators to teach the content to which students will be held accountable. Over the past three decades, the results from assessments have increasingly been used to evaluate education systems and make high-stakes decisions (Pehuniak, 2004). Over this period of time states have moved from norm-referenced or interim-referenced to standards-based assessment. Today, these assessments serve as a principal method for holding educators and students accountable to them.

## **NCLB and the Era of Accountability in Education**

In the 1990s, standards-based reform became the subject of high-profile federal legislation. Prior to NCLB, IASA and the *Goals 2000: Educate America Act* encouraged almost every state to adopt content standards in core subject areas, such as mathematics, English language arts, and science (Linn and Gronlund, 2000; Webb, 1997a; Webb, 1999). While this early legislation motivated many states to adopt content standards, few states developed assessments aligned to the standards to measure student achievement against the standards.

Building upon these previous reforms, NCLB includes a mandate for each state to develop assessments tightly aligned to established content standards so that schools can be held accountable to expectations for high levels of student achievement. To be eligible for federal funds under Title I of NCLB, states must adopt content standards in reading, mathematics, and science. By the 2005–2006 school year, each state must also implement rigorous annual testing programs in reading and mathematics for students in grades 3 through 8 and in one high school grade. By the 2007–2008 school year, students must be assessed in science at least once in grades 3 through 5, once in grades 6 through 9, and once in grades 10 through 12 (Department of Education, 2003; NCLB, § 6311). Moreover, the



results from these assessments must demonstrate that the overall student population and designated subgroups are meeting the state’s requirement for Adequate Yearly Progress (AYP) toward proficiency in each subject area. Sanctions are placed on schools that repeatedly fall short of the annual AYP goal.

To meet the testing mandate of NCLB, states frequently develop their own assessments. These assessments must be aligned with the state’s content standards for each corresponding subject and grade level, and the state must provide evidence of the alignment from a study carried out by an objective third party (Ananda, 2003). NCLB allows a state to fulfill testing requirements using pre-existing assessments, such as a published norm-referenced test (NRT), provided that it has been augmented as necessary to align with state standards. When using a pre-existing assessment, the state uses the results from an alignment study to identify the content gaps between the norm-referenced assessment and the content standards (Jorgensen and McBee, 2003). Items that fill these gaps are added to the test, augmenting the assessment so that it can be used to measure student achievement against the standards. Items can also be removed to improve alignment.

### **Traditional Methodologies for Alignment**

There are three traditional methodologies for systematically evaluating and documenting the alignment between standards and assessments: sequential development, expert review, and document analyses (Webb, 1997a; La Marca, Redfield, Winter, Bailey, and Despriet, 2000). These methodologies can be adapted to compare diverse qualities of standards and assessments, ranging from content match to depth and comprehensiveness. The models of alignment in use today (discussed in this report) incorporate these alignment methodologies.

#### **Sequential Development**

Sequential development, in which the standards and assessments are developed in a serial manner, is perhaps the most straightforward and common (although time-consuming) method of alignment. In this method, the standards are developed first and are then used by the test developers as a blueprint for the structure and content of the assessment. This method has “the advantage of proceeding in a logical process and, after the development of the first document, having known criteria for the development of subsequent documents” (Webb, 1997a, p. 9). During the development process, test developers can identify and record the items that correspond to each standard, thereby providing thorough evidence of alignment. Moreover, the standards and assessments can be presented throughout the process for broad public review and input, ensuring that the alignment of the



assessments to the standards is safeguarded by the scrutiny of educators, experts, and the public (La Marca et al., 2000; Resnick et al., 2003; Webb, 1997a).

### **Expert Review**

Expert review, as the name implies, relies on the opinions of specialists who are knowledgeable about the content covered by the standards and about the development of educational assessments (Webb, 1997a). Generally, this method is used to analyze the alignment between assessments and standards when both have already been developed. The format of the process varies but typically consists of the systematic item-by-item review of an assessment by a committee of content specialists who have been specially trained to judge the alignment of items with standards (Ananda, 2003; Webb, 1997a). The review process may also vary in degrees of formality, sometimes including educators, administrators, parents, and other members of the public in addition to content specialists. Expert review may be integrated into a sequential development process with content specialists evaluating the alignment of each newly developed item.

### **Document Analyses**

In document analyses, the standards and assessment documents are analyzed using a system for encoding their content and structure. Once encoded, the alignment of the documents can be quantified and systematically compared. This approach has been used for complex alignment studies, such as the comparison of the curricula of different nations that was required in developing the Third International Mathematics and Science Study (TIMSS) (Webb, 1997a). Other documents used in education systems, such as textbooks, can be subject to this method of alignment. As with expert review, the reliability of the specialists in accurately encoding the content of the assessments and standards determines the quality of the alignment study's outcomes. The Survey of Enacted Curriculum can be used for this purpose (Porter, 2002, 2004).

## **Leading Alignment Models**

In the past, alignment has been generally understood as a one-to-one match between an item and a content standard in a subject area. The non-regulatory guidance for NCLB gives alignment a more sophisticated definition, asserting that assessments must “measure the depth and breadth of the state academic content standards for a given grade level” (U.S. Department of Education, 2003, p. 12) as well as a range of other qualities, such as comprehensiveness and clarity for users. Hence, an alignment methodology used to satisfy NCLB must include qualitative analysis of alignment in addition to evaluating the content match between each item and standard. Researchers have developed models which use the methodologies discussed earlier to enable more sophisticated alignment analysis.



Today, the most frequently used alignment models are the Webb Model, the Surveys of Enacted Curriculum (SEC) Model, and the Achieve Model.

### **Webb Model**

Norman Webb's model of alignment (1997a; 1997b; 1999) has been very influential in the field (Ananda, 2003; Impara, 2001; La Marca et al., 2000). Webb's criteria for alignment include content focus, articulation across grades and ages, equity and fairness, pedagogical implications, and system applicability (Webb, 1999). *Content focus* concerns the development of student knowledge of the subject matter. *Articulation across grades and ages* concerns the way a student's knowledge grows and changes over time. Evaluation of the articulation should be based on research about how students learn. *Equity and fairness* corresponds to issues of diversity in the student population. Strong alignment between standards and assessments must account for diversity and make high levels of achievement possible for all students. *Pedagogical implications* include the factors that influence student learning, such as attitudes toward subject areas and the technology, materials, and tools that are used in classroom instruction. For example, if the standards require a student to know how to use a calculator or computer, the assessment should allow students to demonstrate the use of such technology. *System applicability* requires standards and assessments to be aligned in realistic, manageable, and credible ways so that educators, administrators, and parents can feasibly use them in the day-to-day activities of the education system.

This range of criteria is thorough but may exceed the needs of an alignment study carried out for NCLB compliance. In his widely cited alignment study of four states (Webb, 1999), Webb pares down his model to evaluate the four aspects of an assessment's content focus: categorical concurrence, depth of knowledge consistency, range of knowledge correspondence, and balance of representation. *Categorical concurrence* evaluates the similarity between the categories of content in the standards and assessments. *Depth of knowledge consistency* compares the content complexity required by the standards and measured by the assessments. *Range of knowledge correspondence* compares the span of knowledge required by the standards in a subject area to that of the assessments. *Balance of representation* compares the emphasis given to certain topics and objectives in the standards to the assessment's corresponding emphasis. To carry out an alignment study using the Webb model, a panel consisting of four to six educators and content specialists is trained to identify the DOK levels and to identify the correct standards. Each criterion is rated numerically, allowing the results to be objectively quantified, calculated, and reported. Then, from the coding of items to the standards, values are computed for all four criteria.

The Webb model's reliability has been well established by its use in alignment studies for more than 10 states (Council of Chief State School Officers [CCSSO],



2002). Leading test publishers also rely on Webb’s model to augment norm-referenced assessments for NCLB compliance. The comprehensiveness of Webb’s complete model makes it similarly adaptable to other purposes for which an alignment study may be required (Impara, 2001).

### **Surveys of Enacted Curriculum (SEC) Model**

The SEC model, developed by Andrew Porter and John Smithson, categorizes the standards and assessments according to *content topics* and *cognitive demand* (CCSSO, 2002). Cognitive demand is described using categories that are specific to each subject area. For example, increasing levels of cognitive demand in mathematics can be described with the following categories: “memorize,” “perform procedures,” “communicate understanding,” “generalize/prove,” and “solve non-routine problems.” This categorization produces a matrix that enables a comparison of the standards and assessments of different states.

To carry out an alignment study using the SEC model, four reviewers (typical case) categorize the content of the standards and assessments according to content topic and cognitive demand. Once the categorization has been completed, the degree of alignment between the standards and assessments can be quantified using statistical computations. To convey the analyses more effectively, the results are also visually presented using charts and graphs. This alignment model can also be adapted to evaluate other elements of an education system, including classroom instruction. By 2002, the SEC model had been used in 11 states and four large urban school districts (CCSSO, 2002).

### **Achieve Model**

Achieve, Inc., has developed an alignment model that can be used to compare a state’s standards to those of other states or nations, provide professional development to state educators, and perform an audit of a state’s education reform. It also can be used for compliance with NCLB by providing evidence of alignment or analyzing a test’s content gaps to guide item development for augmentation (CCSSO, 2002). The Achieve model uses five criteria for alignment: content centrality, performance centrality, challenge, balance, and range (CCSSO, 2002; Resnick et al., 2003). *Content centrality* compares the content of each test item to the corresponding standard. *Performance centrality* compares the difficulty (cognitive demand) of each item to the difficulty required by the corresponding standard. *Challenge* examines whether a set of items considered together expresses the degree of proficiency required by the standards. *Balance* and *range* provide a quantitative and qualitative evaluation of the emphasis placed on topics in the assessment compared to the emphasis placed on the same topics in the standards.



To perform an alignment study using the Achieve model, a panel of content experts uses numerical scales to judge the degree of alignment on each criterion. This quantitative data is computed to provide an objective evaluation of the alignment. This process verified the state mapping of the assessment to the standards. Then reviewers judged the mapping on the above criteria. On content centrality and performance centrality, the group reached consensus on the mapping of each item. The panel members then use written commentary to illuminate patterns that are apparent in the quantitative analysis of the test (CSSO, 2002).

Table 1 provides additional details on the three alignment models concerning their key features and the time required for analysis and for training reviewers.

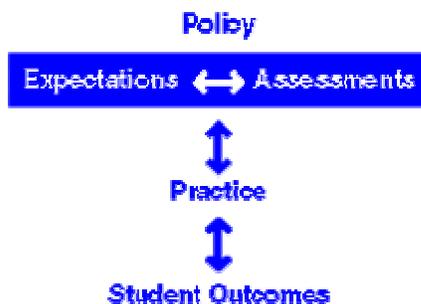
**Table 1. Summary of Three Alignment Models**

<b>Model</b>	<b>Key Features</b>	<b>Review and Analysis Time</b>	<b>Training Time</b>
Webb	<ol style="list-style-type: none"> <li>1. Qualitative ratings</li> <li>2. Quantitative results</li> <li>3. Can measure inter-rater reliability and variation in alignment statistics</li> </ol>	<p>1 day per team and match items, depth of knowledge (Multiple grades); 1 month turnaround for analysis and report. Using the Webb Alignment Tool, alignment analysis is a two-part process. In the first part, reviewers reach consensus on the DOK levels for the objectives under the standards. This takes about 2 hours. In part 2, reviewers code a test to the standards by identifying the DOK for each item and the corresponding objective. This takes from 60 to 90 minutes. We are now able to generate 11 tables of analyses almost instantaneously. We produce complete reports in one to two weeks.</p>	<p>½ day to train reviewers</p>

SEC	<ol style="list-style-type: none"> <li>1. Content matrix</li> <li>2. Measure of alignment highly predictive of student achievement scores</li> <li>3. Information can be applied to help educators and schools</li> </ol>	1 day per team for coding items and benchmarks in matrix; ½ day for readers to complete survey on instruction; 1 week for analysis and report	½ day to train reviewers
Achieve	<ol style="list-style-type: none"> <li>1. Reviewers need to make inferences</li> <li>2. In-depth review</li> <li>3. Provides technical reports</li> </ol>	Alignment review takes 1 day per test; Report and analysis takes 1 – 1½ months.	Has pool of highly trained reviewers

### Broader Issues of Alignment

In a standards-based education system, the curriculum used to guide classroom instruction is based on the same academic standards as the accountability assessments. When one more broadly considers the elements of an education system that are affected by the standards, the scope of alignment extends beyond assessments. Education policymakers distinguish between *horizontal alignment*, the alignment within one level of an education system—policy, program, or performance, and *vertical alignment*, the agreement between levels of an education system (see Figure 1).



**Figure 1. Horizontal and vertical alignment within an education system (Webb, 1997b)**

Horizontal alignment is the degree to which an education system’s accountability assessments match content and performance standards (Porter, 2002; Webb, 1997a; Webb, 1997b). When they are tightly aligned, standards and assessments strengthen the education system by providing a coherent set of expectations to students and educators. Moreover, for an accountability assessment to provide useful (valid and reliable) data about student achievement, it must be aligned to



the standards established by the stakeholders and decision makers (La Marca et al., 2000).

Standards and assessments represent one level, albeit an important one, of the elements in an education system. Examples of other elements include classroom instruction; educational philosophies; curricula at the classroom, district, and state levels; textbook content; the opinion of stakeholders (such as parents); and student achievement outcomes (La Marca et al., 2000; Porter, 2002; Webb, 1997b). When vertically aligned, these elements work in concert to improve student achievement as measured by accountability assessments (Porter, 2002; Webb, 1997b). Vertical alignment can occur at broad or specific levels of an education system. For example, standards and assessments must be vertically aligned with one another so that they reflect the logical, consistent order for teaching the content in a subject area from one grade level to the next. Webb considers both standards and assessments (large-scale) to be policy documents. Therefore, they should be horizontally aligned. These documents should then be vertically aligned with curriculum and teaching. Standards and assessment can serve as the main goals. Ultimately, the strong alignment of this broad range of elements, guided by rigorous standards, is fundamental to the success of a standards-based education system (Ananda, 2003; Anderson, 2002; Porter, 2002; Webb, 1997a; Webb, 1997b).

## Conclusion

Alignment is a complex but increasingly important topic in education. Because of the NCLB mandate for alignment between standards and accountability assessments, policy makers must consider the advantages and disadvantages of available alignment models. The three models presented in this report represent the most sophisticated and comprehensive models for alignment available. Each model has been used successfully in education systems around the nation. In choosing a model, policy makers must weigh the time and resources required to complete the study against the information and outcomes that will be provided by the results.

Ultimately, high levels of student achievement depend on vertical and horizontal alignment within an education system. The models discussed have already been used to evaluate the alignment of other elements of an education system, such as instructional practices and course materials. As the practical applications of the alignment models become more refined and additional resources become available, the scope of alignment studies may extend further. In an optimal standards-based assessment system, every method of gathering information about students, including classroom assessments, must be aligned to the academic standards (McGehee and Griffith, 2001). For example, researchers cite alignment



ASSESSMENT REPORT  
**Alignment in Educational Assessment**

---

as an important factor in the success of online, real-time assessment systems currently being developed to inform instruction (Impara, 2001). By continuing to collaborate on alignment studies, educators, researchers, and policymakers contribute to the advances in education begun in the current era of accountability and standards-based reform.

## References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: Author.
- Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd.
- Anderson, L. W. (2002, Autumn). Curricular alignment: A re-examination. *Theory into Practice*, 41(4), 255–260.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to Improve Learning*. New York: McGraw-Hill.
- Council of Chief State School Officers (CCSSO). (2002, September). *Models for alignment analysis and assistance to states*. Washington, DC: Author. Retrieved from <http://www.ccsso.org/content/pdfs/AlignmentModels.pdf> on September 24, 2004.
- Impara, J. C. (2001). *Alignment: One element of an assessment's instructional unity*. Paper presented at the 2001 annual meeting of the National Council on Measurement in Education, Seattle, WA. Retrieved from <http://www.unl.edu/BIACO/NCME/Alignment%20revised.pdf> on September 21, 2004.
- Jennings, J. F. (1998). *Why national standards and tests? Politics and the quest for better schools*. Thousand Oaks, CA: Sage Publications, Inc.
- Jorgensen, M. A. & McBee, M. (2003). *The new NRT model*. San Antonio, TX: Pearson Inc.
- La Marca, P. M., Redfield, D., Winter, P. C., Bailey, A., & Despriet, L. (2000). *State standards and state assessment systems: A guide to alignment*. 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.
- McGehee, J. J. & Griffith, L. K. (2001, Spring). Large-scale assessments combined with curriculum alignment: Agents of change. *Theory into Practice*, 40(2), 137–144.
- No Child Left Behind Act of 2001*, 20 U.S.C § 6311 *et seq.* (2001).

- Pehuniak, P. (2004). Educational assessment in an era of accountability. In Wall, J. E. & Walz, G. R. (Eds.). *Measuring up: Assessment issues for teachers, counselors, and administrators*. Greensboro, NC: CAPS Press.
- Porter, A. C. (2002, October). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3–14.
- Porter, A. C. (2004, June). *Curriculum Assessment*. Vanderbilt University.
- Resnick, L. B., Rothman R., Slattery, J. B, & Vranek, J. L. (2003). Benchmarking and alignment of standards and testing. *Educational Assessment*, 9(1 & 2), 1–27.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago.
- U.S. Department of Education. (2003, March 10). *Standards and assessments: Non-regulatory guidance*. Washington, DC: Author.
- Webb, N. L. (1997a, April). *Research monograph No. 6. Criteria for alignment of expectations and assessments in mathematics and science education*. Washington, DC: Council of Chief State School Officers.
- Webb, N. L. (1997b, January). Determining alignment of expectations and assessments in mathematics and science education. *NISE Brief 1(2)*. Retrieved from <http://facstaff.wcer.wisc.edu/normw/1997alignmentbrief.htm> on April 23, 2004.
- Webb, N. L. (1999). *Research monograph No. 18. Alignment of science and mathematics standards and assessments in four states*. Washington, DC: Council of Chief State School Officers.

Additional copies of this and related documents are available from:  
Pearson Inc.  
19500 Bulverde Road  
San Antonio, TX 78259  
1-800-211-8378  
1-877-576-1816 (fax)  
<http://www.pearsonassess.com>