



assessment report

The Standards-Referenced Interpretive Framework: Using Assessments for Multiple Purposes

Michael J. Young, Ph.D.

Sasha Zucker

March 2004



The Standards-Referenced Interpretive Framework: Using Assessments for Multiple Purposes

Introduction

From the outset of the development of a large scale assessment system, a central activity in the design process concerns determining the assessment’s *interpretive framework*—the way that its results are understood to convey meaningful information. It is important to understand an assessment’s interpretive framework because it determines how the assessment should be used—its purpose. This report reviews the *norm-referenced* and *criterion-referenced interpretive frameworks* traditionally associated with large-scale assessments and discusses how the current era of standards-based education reform has contributed to the emergence of the *standards-referenced interpretive framework*. Also discussed is a process relevant to all three frameworks—the *alignment* of assessment item content to content standards.

The Norm-Referenced Interpretive Framework

An assessment designed within the norm-referenced framework (Nitko, 2004) yields results that are primarily understood as measuring a student’s achievement relative to the performance of a well-defined reference group that has taken the same assessment. To make a comparison possible, the test publisher administers the assessment under standardized conditions to a nationally representative sample of students and analyzes the results using rigorous statistical methods. The reference population’s results indicate a range of student test scores—commonly referred to as the *norms*—that are used as a basis for comparison to the results of others who take the same test. An individual student’s results from a norm-referenced assessment are reported as a value that indicates a ranking relative to the students in the reference population.

To contribute to its validity as a measure of what students are actually learning, a norm-referenced assessment is designed to include a representative sample of content taught in schools nationwide. In developing Stanford 10, Pearson determined the “national curriculum” for each grade level by consulting

The Standards-Referenced Interpretive Framework

educators, examining textbooks, and, most important, reviewing content standards of educational systems around the country.

Norm-referenced assessment results are useful for purposes such as tracking the progress of a student's achievement over time and identifying a student's relative subject area strengths and weaknesses. However, because norm-referenced results only can be understood in relation to the reference population, they are less effective as an absolute measure of what the student knows and can do (Nitko, 2004).

The Criterion-Referenced Interpretive Framework

An educational assessment designed within the criterion-referenced interpretive framework measures a student's skill or level of performance in a subject area. The content of a criterion-referenced assessment is determined by specific criteria (content standards or instructional objectives) for what students are expected to know and be able to do in a subject area (Nitko, 2004). Typically, a criterion-referenced assessment yields results in the form of a percentage of the items with correct responses, or a *performance level*.

The most basic examples of performance levels are “pass” and “fail.” In an assessment that uses these performance levels, the student's results either *do* or *do not* indicate a satisfactory level of performance. The publisher of a criterion-referenced assessment sets performance levels using a rigorous, widely accepted method—such as the modified Angoff procedure (Angoff, 1984)—to arrive at one or more threshold or “cut scores.” With multiple cut scores, an assessment can use additional performance levels (for example, “basic,” “proficient,” and “advanced”) to provide a more sophisticated evaluation of student achievement. A student's performance level is determined by comparing the number of correct responses to the cut scores; scoring at or above a threshold score indicates that the student has reached the next performance level.

Criterion-referenced assessments are clearly useful for obtaining an absolute measure of what a student knows and can do in a subject area; however, they are less effective than norm-referenced assessments at indicating relative academic strengths or weaknesses (Nitko, 2004).

The Standards-Referenced Interpretive Framework

The standards-referenced interpretive framework, a recent innovation in assessment design that incorporates elements from both the norm- and criterion-

The Standards-Referenced Interpretive Framework

referenced frameworks, has emerged as a solution to the increasingly complex requirements of a series of educational reform initiatives.

Standards-Based Education Reform

Beginning with the *Elementary and Secondary Education Act* of 1965 (ESEA), each state has been required to monitor and assess the educational progress of its students. To accomplish this goal, states have adopted large-scale “high-stakes” assessments (both norm- and criterion-referenced) intended to hold both students and educators accountable to high levels of achievement in core academic subject areas (Education Commission of the States [ECS], 2002). As education reform evolved through the 1970s and 1980s, public leaders continued to search for ways to improve the national educational system. The most recent approach is the “standards-based” reform movement, which proposes the creation of broadly agreed-upon content and performance standards (ECS, 2002). Proponents of this reform movement assert that large-scale educational assessments based on clear and challenging academic expectations are an important means of improving both student performance and the nation’s educational system (Jennings, 1998; Ogawa et al., 2003; Thurlow, 2002).

No Child Left Behind and Standards

The most recent reauthorization of ESEA, the *No Child Left Behind Act* of 2001 (*NCLB*), adheres to the tenets of the standards-based reform movement by requiring states to develop challenging academic content standards for reading, mathematics, and science. To ensure that these reforms are implemented, *NCLB* holds school districts accountable for demonstrating the adequate yearly progress (AYP) of *all* students as measured by a set of high-quality, yearly academic assessments that are aligned with the state’s content standards (U.S. Department of Education, 2003). To encourage higher levels of student achievement and provide more useful feedback (ECS, 2002), the assessments required by *NCLB* must report results using at least three performance levels: basic, proficient, and advanced. Hence, educational agencies were required to use assessments both to track the academic progress of students over time, a task best suited to norm-referenced assessments, and to determine their performance levels in core subject areas, a task best suited to criterion-referenced assessments.

A New Interpretive Framework

The complex requirements of standards-based reform contributed to the development of the *standards-referenced interpretive framework* (Young, 2001). An assessment in this framework combines elements from the two traditional interpretive frameworks thereby yielding more information than either framework alone. As with criterion-referenced assessments, cut scores are established for an assessment developed in this framework, enabling a student’s achievement to be

The Standards-Referenced Interpretive Framework

reported with performance levels. Traditional norm-referenced scores are possible, but not needed, for tests developed under this interpretive framework. Additional types of norm-referenced measures, such as comparisons and rank orderings at the group level, can be obtained from the standards-referenced framework. An example of such a measurement is *achievement expectations*, which compares the progress over time of an individual or group's toward reaching an achievement target or goal—for example, the percentage of students who reach a certain performance level within a set period of time (Young, 2001). In this case, the results from the first year that the assessment is administered serve as the baseline, or point of comparison.

A standards-referenced assessment can potentially provide a more complete, efficient understanding of a student's educational achievement and progress. A suitable metaphor would be the measurements that could be obtained from a marathon. One can measure the running time of individual competitors, the order in which the runners finish the marathon, or the results of teams of runners taken together. Compared to the traditional types of assessments, a standards-referenced assessment can be used for a broader range of purposes than an assessment designed within one of the two traditional frameworks.

Differences From the Traditional Interpretive Frameworks

Although the similarities are clear, there also are some significant differences between the standards-referenced framework and the traditional frameworks. Because content standards vary from state to state, the reference population used to normalize a standards-referenced assessment is often limited to the students of the state for which the assessment is designed. In such a case, the students who take the assessment in the first year it is administered serve as the reference population. The performance levels that are developed within a standards-referenced framework also differ from those within a criterion-referenced framework. While a traditional criterion-referenced assessment's performance levels are developed with a narrower, more immediate focus on what the student is expected to know and be able to do, standards-referenced performance levels typically are associated with broader, long-term educational outcomes.

Alignment of Assessments to Content Standards

Alignment is a process for ensuring that an assessment matches content standards (U.S. Department of Education, 2003). For assessments designed within any of the three interpretive frameworks, a high degree of alignment is essential to make valid inferences about what the student knows and can do in a subject area (Ananda, 2003). Reinforcing the importance of this process, *NCLB* explicitly

The Standards-Referenced Interpretive Framework

requires a high degree of alignment between a state’s annual assessments and its content standards.

The Alignment Process

Test publishers have used several different methods to measure the degree to which an assessment aligns with a set of content standards. Pearson frequently relies on a method developed by Norman Webb (1999) which identifies four alignment criteria:

- *categorical concurrence*, the similarity between the content categories found in the assessment and the standards
- *depth-of-knowledge consistency*, the degree to which the assessment is as rigorous as the content standards specify
- *range-of-knowledge consistency*, the degree to which the range of knowledge measured by the assessment corresponds to the range required by the content standards
- *balance of representation*, the evenness of the distribution of items corresponding to the content standards

The process of alignment frequently reveals content standards that may pose challenges for measuring student achievement. For example, a content standard may specify a list of books for students to read, or might include more objectives than can be measured on a standardized assessment in a reasonable amount of time. During the assessment development and alignment process, states and test publishers can collaborate to resolve these issues and establish a comprehensive assessment system that satisfies *NCLB* requirements (Jorgensen, 2002).

Augmented Norm-Referenced Tests

The process of alignment can also be used to convert a test publisher’s norm-referenced assessment, such as Stanford 10, to the standards-referenced framework. Norm-referenced assessments can be expected to reflect a “national consensus curriculum,” but not necessarily to reflect completely and fully the curriculum as taught in a particular state. In an *augmented norm-referenced test*, differences between the assessment’s content and the state’s content standards are identified by the alignment process. These differences are resolved by augmenting the test with additional items that match any content standards not covered by the original test. Empirical research verifies that the norm-referenced interpretation of results still applies to the assessment, thereby allowing results to be understood in comparison to the reference group. The assessment is further augmented with the

The Standards-Referenced Interpretive Framework

setting of cut scores, allowing results to be reported as performance levels. Moreover, the U.S. Department of Education recognizes the use of augmented norm-referenced tests as a valid method of satisfying the assessment requirements of *NCLB* (Ananda, 2003; Roeber, 2003; U.S. Department of Education, 2003). Hence, the process of augmentation allows test publishers to leverage an existing, high-quality norm-referenced test for multiple purposes. This approach constitutes an affordable, compelling assessment solution for educational agencies.

Conclusion: Selecting an Interpretive Framework

When an educational agency requires an assessment, the intended purpose of the assessment often indicates the optimal framework for its design. If an agency only seeks a comparison between its student population and the rest of the nation, an assessment designed within the norm-referenced framework is appropriate. If the assessment needs to determine how well students perform compared to a set of content standards and objectives, an assessment designed within the criterion-referenced framework is appropriate. When both types of information are required, an educational agency may choose to administer two assessments, one in each of the traditional frameworks.

However, the standards-referenced framework is clearly a cost-effective design that enables one assessment to serve multiple purposes, including the measurement of broad, educational outcomes defined by content and performance standards. When a standards-referenced framework is identified as the optimal design for an assessment, augmentation of existing high-quality norm-referenced assessments represents a compelling and affordable solution. As educational reform demands an increasingly sophisticated understanding of student achievement, Pearson's measurement consultants and assessment specialists can assist educational agencies in adopting an assessment with the optimal interpretive framework for their purposes.

References

- American Education Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for education and psychological testing*. Washington, DC: American Education Research Association.
- Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd.
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

- Education Commission of the States. (2002). *No Child Left Behind issue brief: A guide to standards-based assessment*. Denver, CO: Author.
- Jennings, J. F. (1998). *Why national standards and tests?* Thousand Oaks, CA: SAGE Publications.
- Jorgensen, Margaret. (2002). Can the testing industry meet growing demand? *Issues in Science & Technology*, 19(2), 59.
- Nitko, A. J. (2004). *Educational assessments of students*. Englewood Cliffs, NJ: Prentice Hall.
- Ogawa, R.T., Sandholtz, J. H., Martinez-Flores, M., & Scribner, S.P. (2003). The substantive and symbolic consequences of a district's standards-based curriculum. *American Educational Research Journal*, 147.
- Roeber, E. D. (2003). *Assessment models for No Child Left Behind*. Denver, CO: Education Commission of the States.
- Thurlow, M. L. (2002). Positive educational results for all students. *Remedial & Special Education*, 195.
- U.S. Department of Education. (2003). *Standards and assessments: Non-regulatory guidance*. Retrieved December 18, 2003 from <http://www.ed.gov/policy/elsec/guid/saaguidance03.doc>
- Webb, N. L. (1999). *Alignment of science and mathematics standards and assessments in four states*. Washington, DC: Council of Chief State School Officers.
- Young, M. J. (2001). *Using assessments for multiple purposes: Norm-referenced, criterion-referenced, and standards-referenced interpretive frameworks*. Unpublished manuscript, San Antonio, TX: Pearson Inc.

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>