

“ALIGNING ASSESSMENTS WITH STANDARDS”

A Synthesis of Guidelines from Current Practice Adapted for use in Teacher Education And NCATE Accreditation

NCATE; EJE; last revised May 30, 2001

I. INTRODUCTION

NCATE standards for performance-based accreditation call for assessments that are “aligned” with standards, or assessments that “are appropriate” for the standards. For example, here are some excerpts from the supporting explanations for NCATE unit standard 1, *Candidate Knowledge, Skills, and Dispositions* and standard 2, *Assessment System and Unit Evaluation*:

- Institutions must submit program documentation, including candidate performance data, that responds to professional standards for national and/or state review prior to and during the on-site visit.
- The measures . . . must be sufficient and appropriate to inform the important aspects of . . . candidate performance.
- Professional, state, and institutional standards are reference points for candidate assessments.
- The unit administers multiple assessments in a variety of forms and aligns them with candidate standards.

Similarly, the *Program Standards for Elementary Teacher Preparation*, prepared as a “model” for shifting NCATE program standards and reviews to a performance-based process, anticipates evidence that:

- Is congruent with the knowledge and skills standards.
- Measures the different “attributes” of standards in appropriate and multiple ways.

Implicit in these references is that sometimes assessments are not aligned or appropriate. How might faculty think through, in a systematic and recurring way, whether and how well their assessments provide information to evaluate the knowledge and skills for effective teaching that are incorporated into the NCATE, specialty organization, or state standards?

This paper is an attempt to summarize guidelines for linking assessments with standards, based on the current state-of-the-art¹. Section II contains a brief narrative statement of “principles” for alignment, from Norman Webb of the National Institute for Science Education at the University of Wisconsin. Section III states guidelines for aligning assessments with standards, and section IV is a similar statement of guidelines for other features of assessment systems that are needed to

support alignment. Section V is a chart that suggests forms of assessments suited to measuring candidate proficiencies on standards of different types. The chart was prepared by Richard Stiggins, President of the Assessment Training Institute.

The audience for these guidelines includes:

- Faculty who are developing and applying assessments for candidates enrolled in teacher preparation programs;
- Program coordinators and reviewers from NCATE affiliated specialty organizations who must apply the results of candidate assessments in their judgments for national recognition;
- Board of Examiner and UAB members who will review results of candidate assessments in evaluating evidence for standard 1, *Candidate knowledge, skills, and dispositions*;
- NCATE state partners, and reviewers of state partnership applications from specialty organizations, for use in evaluating the alignment of state standards and specialty standards; and
- NCATE staff and presenters at orientation sessions.

II. PRINCIPLES FOR ALIGNMENT

A 1997 research monograph², by Norman Webb of the National Institute for Science Education at the University of Wisconsin-Madison, provides several introductory comments about “Alignment in Principle” from which the paragraphs below are excerpted. (The boldfaced type has been added to the original for emphasis. References in these paragraphs, endnotes 3 through 7, are all copied from Dr. Webb’s monograph and appear at the conclusion of this paper.)

Two or more system components are aligned if they are in agreement or match each other. In the past, the most common educational use of the concept of alignment referred to the match between an assessment instrument (or instruments) and a curriculum. Here, alignment is analogous to instructional or curricular validity of a test³. A legal ruling in the 1981 Florida case *Debra P. vs. Turlington* emphasized the importance of assuring agreement between a curriculum and tests. According to this ruling, for a test to be fair, both curriculum and instruction must match the content coverage of the test⁴. Legally, high stakes tests need to be fair by being “aligned” with curriculum and instruction. . . .

The form of an assessment can be as important as the content in judging alignment. “The content and form of an assessment task must be congruent with what is supposed to be

measured⁵.” For example, an assessment using a short-answer format is not aligned with an intended purpose of measuring students’ ability to frame questions for conducting scientific inquiry and to design an inquiry to address the questions.

Alignment does not only refer to a comparison . . . (of) one assessment instrument with a curriculum, but extends to a set of assessment instruments or the assessment system. “The term *alignment* is often used to characterize the congruence that must exist between an assessment and the curriculum. Alignment should be looked at over time and across instruments⁶.” A single assessment may not be well aligned with curriculum because it is too narrowly focused, but it may be part of a more comprehensive collection of assessments that is in full alignment with the curriculum. . . .

Alignment is intimately related to test “validity.” However, important and useful distinctions can be drawn between the two concepts. Validity refers to the appropriateness of inferences made from information produced by an assessment⁷ (Cronbach, 1971). **Alignment refers to how well all policy elements in a system work together to**

guide instruction and, ultimately, student learning. . . . a test, or tests, and a curriculum framework that are in alignment will work together to communicate a common understanding of what students are to learn, to provide consistent implications for instruction, and to be fair for all students.

III. GUIDELINES FOR ALIGNMENT OF ASSESSMENTS WITH STANDARDS

Twelve source materials were tapped to create these “state-of-the-art” guidelines. While there are unique aspects of each of these sources—reflecting either the particular use that authors had in mind, or the specialty field from which they come—it is the commonalities that are emphasized in the guidelines stated here. **These commonalities center on standards. This perspective contrasts, and intentionally so, with much of the current reform debate at the national level and in states in which the standards are virtually forgotten and, instead, “the test” is treated as the reform. The NCATE position is that standards form the foundation for discussion, action, and analysis in teacher preparation, and in accreditation as well.**

The guidelines address the content of standards and assessments, the forms of candidate performance, and the level of difficulty.

- **The content of assessments is congruent with the content of the standards.**
 - The same or consistent categories of content appear in both standards and assessments
 - The span or range of knowledge required in assessments is equivalent to that in the standards
 - Similar emphasis is given to different content topics, instructional activities, and tasks
 - Items in the assessments are balanced across objectives in the standards
- **The cognitive and skill demands of assessments closely match expectations defined in standards.**
 - There is congruence between assessments and standards in complexity, cognitive demands, and skill requirements
 - Forms of assessments are adapted to objectives within standards, such as knowledge and comprehension, ability to apply or practice, dispositions or attitudes, capacity to analyze or reflect, and effects on student learning
- **The level of difficulty of assessments is consistent with standards for a candidate completing teacher preparation.**

- The degree of challenge in assessments for each teaching specialty and teaching level is appropriate for candidates who are ready to teach

IV. GUIDELINES FOR SUPPORTING FEATURES OF ASSESSMENT SYSTEMS

Many of the source materials that have been synthesized for the “alignment” guidelines in section III, above, contain additional comments about the nature of assessments or assessment systems. In effect, the point is that “alignment” is not enough by itself. Assessments that are appropriate for standards also require: multiple measures of candidate proficiencies, written criteria that define successful performance, actions to evaluate the credibility of the assessments, and using results to improve teacher education. The assessment system guidelines and supporting statements from these sources are summarized in the points that follow.

- **Assessments evaluate candidates using multiple types of assessments across the multiple domains of knowledge, dispositions, and performances.**
 - Assessments include paper and pencil tests, performance measures, evidence of positive effects on student learning, external reviews, and candidate self-reports, some embedded in instruction and some summative in nature
 - Candidates demonstrate their knowledge and skills for teaching practice in a variety of instructional contexts (subject matters, socio-cultural, grade levels, locations)
 - Assessment data are collected through a variety of assessors (e.g., instructors, classroom teachers, candidates, peers, future employers, and their students)
 - Assessments reflect current knowledge and best practice in teaching of the respective specialty field
- **Rubrics or criteria define successful performance on assessments and are used to evaluate candidate work.**
 - Faculty describe levels of performances to indicate what is valued in a candidate’s response and what is expected in order for candidates to be successful
 - The expected levels of performances are written and publicly shared
 - Benchmark examples provide useful illustrations of faculty expectations for candidates’ performance levels
 - Candidate progress is evaluated with rubrics or criteria at various stages throughout a program to determine entry, retention, and exit or termination from the program
 - Rubrics or criteria performance levels are appropriate for candidates’ progress at the point each assessment is administered, and by completion should represent the institution’s expectation that candidates are fully prepared for licensure and initial teaching

- **Faculty evaluate the fairness of assessments on a continuing basis.**
 - Assessments are conducted only on knowledge and skills that candidates have opportunities to learn and practice
 - Assessments use appropriate terminology and avoid stereotyping, ethnocentrism, and biases
 - Assessments accurately represent the performances, competencies, and dispositions that are included in the standards
 - Assessments yield consistent results

- **Quality candidate assessment systems use results to advise candidates, improve teaching, and strengthen programs.**
 - Quality candidate assessment systems communicate performance expectations to students and provide systematic feedback on candidate products and performances in terms of their progress toward achieving proficiencies in standards
 - Assessments are also used to improve teaching and strengthen programs as well as to communicate with employers, governmental agencies, professional organizations, and the public.

V. APPLYING ALIGNMENT TO ASSESSMENT DESIGN

Richard Stiggins, President of the Assessment Training Institute, Portland, Oregon, has written a paper for NCATE⁸ in which he describes appropriate forms of assessments (quizzes, essays, performances, etc.) for particular types of standards or “achievement targets.” He introduces the chart as follows:

Since most assessment contexts call for student mastery of several different kinds of achievement and since no single assessment method can reflect them all, the assessor is always faced with the challenge of selecting from among a variety of methods. The available options include *selected response tests and quizzes*, (multiple choice, true/false, matching and fill in), *essay assessments*, *performance assessments* (based on observation and judgment), and direct *personnel*

communication with the student. Part of the challenge of building sound assessment systems is to consistently match methods to the intended targets. Strong and weak matches are depicted in [the figure on the next page]. Our challenge is to be sure that all who will design and use assessments to develop and then certify the competence of teacher candidates are sufficiently assessment literate to know and understand what methods to use when, and how to use them well.

The chart on the following page is a part of Dr. Stiggins’ response to matching “methods to the intended targets.”

Aligning Achievement Targets to Assessment Methods

<i>TARGET TO BE ASSESSED</i>	<i>ASSESSMENT METHOD</i>			
	SELECTED RESPONSE	ESSAY	PERFORMANCE ASSESSMENT	PERSONAL COMMUNICATION
KNOWLEDGE MASTERY	Multiple choice, true/false, matching, and fill-in can sample mastery of elements of knowledge	Essay exercises can tap understanding of relationships among elements of knowledge	Not a good choice for this target--Three other options preferred	Can ask questions, evaluate answers and infer mastery--but a time-consuming option
REASONING PROFICIENCY	Can assess understanding of basic patterns of reasoning	Written descriptions of complex problem solutions can provide a window into reasoning proficiency	Can watch students solve some problems and infer about reasoning proficiency	Can ask student to “think aloud” or can ask follow up questions to probe reasoning
SKILLS	Can assess mastery of the knowledge prerequisites to skillful performance--but cannot rely on these to tap the skill itself		Can observe and evaluate skills as they are being performed	Strong match when skill is oral communication proficiency; also can assess mastery of knowledge prerequisite to skillful performance
ABILITY TO CREATE PRODUCTS	Can assess mastery of knowledge prerequisite to the ability to create quality products--but cannot use these to assess the quality of products themselves		A strong match can assess; (a) proficiency in carrying out steps in product development, and (b) attributes of the product itself	Can probe procedural knowledge and knowledge of attributes of quality products--but not product quality
DISPOSITIONS	Selected response questionnaire items can tap student feelings	Open-ended questionnaire items can probe dispositions	Can infer dispositions from behavior and products	Can talk with students about their feelings

From Stiggins, Richard J., *Student-Involved Classroom Assessment*, 3rd ed. (Columbus, Ohio: Merrill Education, 2001)

¹ The first group of sources is derived from alignment at the K-12 level. Two publications were developed by the National Institute for Science Education at the University of Wisconsin-Madison for the Council of Chief State School Officers, the first in 1997 and the second in 1999. These are *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*, and *Assessment of Science and Mathematics Standards and Assessments in Four States. Academic Standards and Assessments Benchmarking Evaluation for Michigan* is from “Achieve” and describes the criteria employed in its piloting of “benchmarking” standards and assessments to “best in the practice” examples. This is supplemented with material, including assessment examples, from an Achieve paper currently pending publication that describes the alignment methodology developed by LRDC and Achieve, and currently employed in their state alignment studies.

Additional sources are AFT’s annual evaluation of state standards and assessments, specifically *Making Standards Matter: An annual fifty-state report on efforts to raise academic standards, 1995*, together with a note updating the AFT evaluation questions for 2001; and a National Education Goals Panel’s technical paper on high standards for student learning, *Promises to Keep: Creating High Standards for American Students*, 1993. [Additional information is pending on alignment of assessments with standards from NBPTS.] A final paper in this first group is directed at all levels of standards and assessments—the AERA/APA/NCME standards for educational testing, *Standards for educational and psychological testing*, 1999.

The second set of sources includes four NCATE-Specialty organization “assessment criteria” projects for social studies, science, English, and elementary teacher candidates. These four are: *Program Standards for Elementary Teacher Preparation; Assessment Criteria Project*, final edition, 2000; *Guidebook for Colleges and Universities Seeking to Meet NCSS/NCATE Program Standards for the Initial Preparation of Teachers of Social Studies*, 1999; *NCATE/NSTA Task Force Report*, 2000; and *NCTE/NCATE Research Project on the Assessment of the Preparation of Teachers of English Language Arts*, 2001. Finally, the chart illustrating forms of assessments that are particularly well suited to various standards or outcomes or “targets” to be assessed comes from a paper that NCATE commissioned by Richard Stiggins, of the Assessment Training Institute in Portland, Oregon, *Specifications for a Performance-based Assessment System for Teacher Preparation*, and the chart comes from his own book, *Student-Involved Classroom Assessment*, 3rd ed., 2001.

² Webb, Norman L., (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*, National Institute for Science Education, University of Wisconsin-Madison, Washington, DC, the Council of Chief State School Officers.

³ Harmon, M. (1991). Fairness in testing: Are science education assessments biased? In G. Kulm and S. M. Malcom (Eds.), *Science assessment in the service of reform* (pp. 31-54). Washington, DC: American Association for the Advance of Science.

⁴ Madaus, G. F. (1983). *The courts, validity, and minimum competency testing*. Boston: Kluwer-Nijhoff.

⁵ National Research Council. (1996). *National Science Education Standards*. p. 83. Washington, DC: National Academy Press.

⁶ Mathematical Sciences Education Board. (1993). *Measuring what counts. A conceptual guide for mathematics assessment*. P. 123. Washington, DC: National Academy Press.

⁷ Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

⁸ Stiggins, R. J. (2000). *Specifications for a Performance-based Assessment System for Teacher Preparation*. P.p 7 and 8. For the National Council for Accreditation of Teacher Education.