THE EFFECTIVENESS OF THE STUDENT ANNUAL NEEDS DETERMINATION INVENTORY (SANDI) WITHIN THE SPECIAL EDUCATION CONTEXT

Prepared for Riverside County Office of Education

In this report update, Hanover Research links secondary research findings on special education assessments to evaluations of the value, validity, and effectiveness of RCOE's proposed tool for special education assessment, SANDI.

HANOVER RESEARCH
www.hanoverresearch.com

TABLE OF CONTENTS

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

The Riverside County Office of Education (RCOE) is in the process of preparing an article regarding its new alternate assessment for students with significant disabilities, called the Student Annual Needs Determination Inventory (SANDI). The SANDI is a comprehensive online system that captures student data through meaningful summative and formative assessment and delivers access to standards based, data-driven instruction. It supports and guides standards-based IEP goal writing, and documents educational benefit over time.

At the request of RCOE, Hanover Research (Hanover) has contributed to this effort through a series of projects confirming the validity of SANDI's items in assessing relevant content in reading/ELA, writing, communication, and math, and preparing a literature review to situate the SANDI in its broader research context. Work completed for this effort by Hanover as of February 2017 includes studies of:

- **correlation**, which examines the relationship between students' Student Annual Needs Determination Inventory (SANDI) content-area test scores and their performance on three other assessments: California Alternate Performance Assessment (CAPA), Woodcock-Johnson-III (WJ), and Vineland-II (Vineland);
- **content validity**, which evaluates the extent to which items on the Student Annual Needs Determination Inventory (SANDI) assess the relevant content area; and
- **inter-rater reliability**, which examines the level of agreement between special education teachers on subject items from the SANDI assessment.

The present report seeks to unify these efforts in a single document, supported with additional literature review and analysis. It is organized in three sections as follows:

- **Section I: Alternate Assessment in Special Education** describes the history of alternate assessment within the context of national educational trends and explores the components of quality alternate assessments based on empirical research studies.
- **Section II: Current Options for Alternate Assessments** examines practices for choosing from among present options available to states and districts, reviews two prominent options, and presents Hanover's analyses of the SANDI.
- **Section III: Best Practices in Alternate Assessment** situates the use of alternate assessment within the larger practice of supporting students with disabilities. In particular, it draws attention to the SANDI's innovations in professional development over other options available to states and districts.

Across these sections, Hanover demonstrates that **the full value of the SANDI goes beyond that of other current alternate assessment options to provide a larger framework for guiding the kinds of communication and collaboration that effective alternate assessment**

requires: defining appropriate standards, differentiating instruction to meet those standards, and accurately observing the success of such efforts in providing students access to those standards. Below we summarize key findings of our work.

KEY FINDINGS

- **Alternate assessment was introduced with the 1997 reauthorization of the IDEA, and was followed by a gradual intensification of expectations for students with severe disabilities**. It is intended for a small audience of a state's or district's total population, with a cap on using alternate assessments with only 1 percent of students. As a result of subsequent legislation, all states now have an alternate assessment plan in place.
  - *The federal government funded development of two widely-used national alternate assessment models aligned to educational standards and policy*: National Center and State Collaborative/Multi-State Alternate Assessment (NCSC/MSAA), and Dynamic Learning Maps (DLM). Twenty-seven states explicitly rely on one of these two systems for their own alternate assessment program. Both systems are constructed around valid and reliable alternate achievement standards and offer digital, adaptive technologies to facilitate testing.
  - *The SANDI presents an important innovation in this assessment space* by offering not just an assessment instrument, but a larger framework for implementing the assessment effectively and appropriately for students with significant disabilities.
- **As with all assessments, technical quality is the most important characteristic of an effective alternate assessment**. It can be difficult to develop technical quality in an alternate assessment, however, because the population for an alternate assessment is both small and diverse. The population size makes it difficult to follow standard quantitative investigation of validity and reliability, while the population's diversity challenges typical assumptions about how to operationalize these assessment characteristics.
  - *The SANDI has a demonstrated technical quality in terms of correlation with other key alternate assessment instruments, internal item content validity, and inter-rater reliability*. Students' SANDI outcomes are highly and significantly correlated with their performance on CAPA, WJ, and Vineland assessments. Likewise, the level of inter-rater agreement is high, denoted by a kappa of 0.70 ("substantial agreement"). There is variation in the level of agreement among teachers in their ratings of different subjects' items, but in general the level of agreement among teachers is significantly higher than random agreement. Finally, experts rated most content area items very highly in terms of content representativeness, with little variation among participating scorers.
- **Designating a student as eligible for an alternate assessment is a complex choice that influences his or her education, as well, because it allows for a lowering of academic achievement expectations**. Alternate assessments must consider issues of

inclusion, appropriateness, meaning, and cultural relevance, so as to maintain as high a standard as is appropriate for these students. Researchers consider this a second essential quality to examine, referred to as "consequential validity."

- **Communication is central to effective implementation of alternate assessment systems**. Parents, specialists, and other stakeholders must receive adequate training and support to advocate effectively for their child or student. Such collaboration is facilitated through the IEP development process, and through professional development opportunities for administrators and teachers.
  - *The SANDI uses professional learning community (PLC) as a central component of its alternate assessment model to analyze student data, explicitly led through accompanying professional development modules*. Specifically, modules have been developed and implemented through consistent and ongoing teacher input, teacher and administrator training, leadership team and administrative coaching, and feedback cycle. Each module may be customized depending on the site staff availability and needs. To ensure consistent administration of the SANDI assessment, professional development modules are delivered by site-level leadership teams, and implementation is supported and monitored by district leadership. All modules are available online 24/7 for review.

SECTION I: ALTERNATE ASSESSMENT IN SPECIAL EDUCATION

This section reviews empirical literature on the history, purpose, and policy of alternate assessment in special education, and describes efforts to manage the technical quality of alternate assessments.

HISTORY OF SPECIAL EDUCATION ASSESSMENT

Alternate assessment was introduced through modification of the Individuals with Disabilities Education Act (IDEA) in 1997, and was followed by a gradual intensification of expectations for students with severe disabilities in the 2000s to present. IDEA 1997 "required states to create and implement alternate assessment systems by July 1, 2000, and include the performance of students participating in alternate assessments in public accountability reporting." [1] At the time, "only one state, Kentucky, had widespread implementation of this process."[2] Once states began implementing their own systems, expectations were expanded with the introduction of NCLB's (2001) "adequate yearly progress" mandate, which included students with disabilities in measuring a district's quality via academic performance, as well as its "demands to improve the reliability and validity of inferences based on alternate assessment results."[3] NCLB marked an intensification of focus on "accountability" in education.[4] Roach, Elliott, and Webb describe the impact of this legislation in terms of its sudden shift from previous approaches to special education:

> This emphasis on attaining academic achievement represents a dramatic departure from the curriculum and inclusion practices that traditionally have been implemented with many students with significant disabilities. Early considerations of mainstreaming and least restrictive environment (LRE) often focused on the socialization and self-esteem benefits for students with significant disabilities. More recent practices have maintained the focus on relationships and self-concept while adding an emphasis on exposure to the general curriculum and the broader school experience. IDEA, however, demands even greater access to the general education curriculum. Students must have instruction and accommodations that promote their progress, no matter how modest, toward the education expectations of the larger student population.[5]

[1] Roach, A.T., S.N. Elliott, and N.L. Webb. "Alignment of an Alternate Assessment with State Academic Standards: Evidence for the Content Validity of the Wisconsin Alternate Assessment." *Journal of Special Education* 38:4 (Winter 2005): p. 218. ProQuest.
[2] Browder, D.M., et al. "What We Know and Need to Know About Alternate Assessment." *Exceptional Children* 70:1 (Fall 2003): p. 45. EBSCO.
[3] Roach, Elliott, and Webb, "Alignment of an Alternate," Op. cit., p. 218.
[4] Lemons, C.J., et al. "Implementing an Alternate Assessment based on Modified Academic Achievement Standards: When policy meets practice." *International Journal of Disability Development and Education* 59:1 (March 2012): p. 67. EBSCO.
[5] Roach, Elliott, and Webb, "Alignment of an Alternate," Op. cit., p. 219.

Researchers emphasize that IDEA 1997's intent was that alternate assessment be used with only a small portion of the larger group of students with disabilities. Towles-Reeves et al. (2009) write, "alternate assessments are designed for a very small percentage of the student population, for whom traditional assessments, even with appropriate accommodations, would be inappropriate measures of progress within the general education curriculum."[6] Furthermore, "the USED [U.S. Department of Education] has placed a cap on the number of proficient AA scores that a school district or state may include in AYP calculations at one percent."[7] Under the new guidelines of the Every Student Succeeds Act (ESSA), the caps will be somewhat relaxed as states can apply for a waiver and are given authority to define "severe cognitive disabilities" independently, though it will still be difficult to meet the eligibility requirements for that waiver.[8]

DEVELOPMENT OF EFFECTIVE ALTERNATE ASSESSMENTS

The National Center on Educational Outcomes (NCEO) classifies alternate assessments for students with disabilities into three groups: those based on alternate achievement standards (AA-AAS), those based on modified academic achievement standards (AA-MAS), and those based on grade-level achievement standards (AA-GLAS).[9] Elliot and Roach note that each

> […] in general requires the *collection of evidence samples* (e.g., classroom work products, videotapes, interviews, structure[d] observations, students' responses to on-demand tasks) to characterize students' knowledge and skills that are determined to be *aligned or "linked" to states' grade level content standards*. The evidence samples must then be evaluated and scored to yield data that can be *summarized by a proficiency level descriptor* based on a rigorously established set of grade-level achievement standards.[10]

The focus of the alternate assessment conversation is on the first of these, AA-AAS, being designed "for students with the most significant cognitive disabilities."[11] Meanwhile, AA-MAS are seen as optional and being "phased out," and AA-GLAS are for students "who need testing formats or procedures that are not included in the general assessment or not addressed with use of accommodations" but otherwise are held to standard grade-level achievement expectations.[12]

---

[6] Towles-Reeves, E., et al. "An Analysis of the Learning Characteristics of Students Taking Alternate Assessments Based on Alternate Achievement Standards." *Journal of Special Education* 42:4 (February 2009): p. 242. ProQuest.
[7] Streagle, K., and K.W. Scott. "The Alternate Assessment Based on Alternate Achievement Standards Eligibility Decision-Making Process." *The Qualitative Report* 20:8 (2015): p. 1290. EBSCO.
[8] Klein, A. "ESSA Panel Hammers Out New Testing Regulations." *Education Week*, April 26, 2016. http://www.edweek.org/ew/articles/2016/04/27/essa-panel-hammers-out-new-testing-regulations.html
[9] "Alternate Assessments for Students with Disabilities." National Center on Educational Outcomes, 2016. https://nceo.info/Resources/publications/TopicAreas/AlternateAssessments/altAssessTopic.htm
[10] Elliot, S.N., and A.T. Roach. "Alternate Assessments of Students with Significant Disabilities: Alternative Approaches, Common Technical Challenges." *Applied Measurement in Education*, 20:3 (2007), p. 305. EBSCO.
[11] Towles-Reeves, E., H. Kleinert, and M. Muhomba. "Alternate Assessment: Have We Learned Anything New?" *Exceptional Children* 75:2 (Winter 2009), p. 235. http://www.naacpartners.org/publications/2009Towles-ReevesKleinertMuhomba.pdf
[12] "Alternate Assessments," NCEO, Op. cit.

"The decision for a student to take an AA-AAS is a complex decision with consequences beyond simply what academic achievement test a student will take at the end of a school year," warn Streagle and Scott. Designation as eligible for an alternate assessment "influences the academic instruction a student will receive and the performance expectations to which that student will be held […], reduced in depth and complexity from that of their peers without significant ID."[13] However, advocates see the alternate assessment movement as offering several "promises," as outlined by Browder et al. (2003):

- To foster greater consideration of students with disabilities in school and state policy decisions
- To increase overall expectations for people with disabilities
- To increase access for all students to the same curriculum and standards
- To improve instructional programs at the teacher and classroom level[14]

The wide range of individual needs that must be addressed through alternate assessment, despite the small target percentage, complicates both its practice and the study of its effectiveness. Changes made to create an alternate assessment addressing these varied needs can impact any of four areas: "the levels of support or […] the breadth, depth, or complexity of the standards being assessed," illustrated in Figure 1.1 below.[15]

Figure 1.1: Change Types in Creating an Alternate Assessment

IMPACT	DEFINITION
Levels of support	Scaffolds, prompts, and assistive technologies used in the administration of the assessment
Breadth	Comprehensiveness of content and skills embodied in the assessment
Depth	The level of cognitive processing (i.e., recognition, recall, problem solving, analysis, synthesis, and evaluation) required for success relative to the performance standards
Complexity	The language used to express the item's content

Source: Tindal et al., "Documenting Reading Achievement," Op. cit., p. 322

Creating these changes "require[s] educators to decide what they expect students with severe disabilities to achieve" while considering the broad scope of needs to be governed by this alternate.[16] This is a difficult process for several reasons, and one further hampered by the limited research conducted on the quality of these alternate assessments. Tindal et al. write that, "given the complexities of the population and the measurement systems, and the difficulties associated with both aligning to standards and establishing technical adequacy, it is not surprising that little information is available on the achievement growth

[13] Streagle and Scott, "The Alternate Assessment…Process," Op. cit., p. 1309.
[14] Browder et al., "What We Know," Op. cit., pp. 46, 49.
[15] Tindal, G., et al. "Documenting Reading Achievement and Growth for Students Taking Alternate Assessments." *Exceptional Children* 82:3 (2016): p. 322. EBSCO.
[16] Browder, D., et al. "How States Implement Alternate Assessments for Students with Disabilities." *Journal of Disability Policy Studies* 15:4 (Spring 2005): p. 217. ProQuest.

of this population."[17] Wyse et al. echo this, writing that the population of students eligible for AA-AAS is so "small and often unique" as to limit the "amount of research [that] has directly investigated many of the technical aspects of these assessments and specifically approaches for equating them."[18]

Additionally, educators disagree about the purpose of alternate assessment, falling into two main argumentative camps: those who feel alternate assessment should focus on functional life skills, and those who feel it should focus on academic achievement. Browder et al. describe this as a debate between "authentic skills assessed in real life environments" and "access to the same curriculum."[19] Recent legislation has sided, to an extent, with the access and academic achievement camp, given the intensification of interest in accountability. But observers note that, still, "many teachers opt to maintain a difficult balance between academic skills and nonacademic skills in their classroom instruction, in a good-faith effort to provide what they deem to be essential tools for these students to live successful lives outside school."[20]

Beyond these philosophical debates, researchers identify three specific needs in crafting an alternate assessment, including "(a) ensuring content validity (that indicators proposed to be math are really math); (b) using scientifically based research to define the indicators; and (c) continuing to honor such values as inclusion, self-determination, and age-appropriate functional skill instruction."[21] Some terms for this process of developing rigorous, appropriate alternate assessments include "equating," "calibrating," "projecting," and "moderating," which assume that there is a standard form used with the general population that is being adapted for a subpopulation of students with disabilities.[22] The end goal of this process, regardless of terminology, can be summarized as "technical quality," which Kettler et al. define as "reliability, validity, accessibility, objectivity, and consistency."[23] Presumably this encompasses much of the same rigor and quality expectations required of any assessment.[24] Kettler et al. specify four main components associated with an alternate assessment's technical quality:

- An explicit structure
- Guidelines for determining which students may participate
- Clearly defined scoring criteria and procedures

[17] Tindal et al., "Documenting Reading Achievement," Op. cit., p. 322.
[18] Wyse, A.E., et al. "Considerations for Equating Alternate Assessments: Two Case Studies of Alternate Assessments Based on Alternate Achievement Standards." *Applied Measurement in Education* 26 (2013): p. 51. EBSCO.
[19] Browder et al., "What We Know," Op. cit., p. 49.
[20] Kettler et al., "What Do Alternate Assessments," Op. cit., pp. 471-472.
[21] Browder et al., "How States Implement," Op. cit., p. 217.
[22] Wyse et al., "Considerations for Equating," Op. cit., p. 53.
[23] Kettler, R.J., et al. "What Do Alternate Assessments of Alternate Academic Achievement Standards Measures? A Multitrait-Multimethod Analysis." *Exceptional Children* 76:4 (Summer 2010): p. 458. EBSCO.
[24] Wyse et al., "Considerations for Equating," Op. cit., p. 53.

- A report format that communicates student performances in terms of academic achievement standards[25]

Wyse et al. argue that "standardized scoring is a particular challenge" because changes to the assessment can affect multiple points in the item, such as how the directions are written or administered. Similarly, AA-AAS must be "flexibl[e] to meet the often unique disability manifestations of individual students."[26] Furthermore, the complications of administering the assessment appropriately challenge our operationalization of qualities like reliability and validity. Taylor and Pastor write, "traditional statistics typically used for capturing reliability may not clearly and comprehensively model the measurement error associated with scores" since students might be given multiple opportunities for assessment, be observed or rated by multiple staff according to unique sets of expectations, and with outputs like portfolios that might not easily match the referenced "normal" assessment.[27]

[25] Kettler et al., "What Do Alternate Assessments," Op. cit., p. 458.
[26] Wyse et al., "Considerations for Equating," Op. cit., p. 52.
[27] Taylor, M.A., and D.A. Pastor. "An Application of Generalizability Theory to Evaluate the Technical Quality of an Alternate Assessment." *Applied Measurement in Education* 26 (2013): p. 280. EBSCO.

SECTION II: CURRENT OPTIONS FOR ALTERNATE ASSESSMENT

This section examines practices for choosing from among present options for alternate assessment and compares two prominent options to the SANDI.

OVERVIEW OF ALTERNATE ASSESSMENT INSTRUMENTATION

In 2009, Towles-Reeves et al. wrote that "as a field, alternate assessment for students with disabilities is in its infancy."[28] Kettler et al. (2010) echo this observation, writing that "very little published research examines the constructs measured by AA-AASs, […] partially attributable to the challenges of assessing the student population for whom alternate assessment are intended and in some states because of the lack of adequate sample sizes to conduct [empirical] studies."[29]

Yet, **driven largely by federal education mandates like IDEA and NCLB, the field of alternate assessment options has grown substantially since Kentucky's initial offering in 1992**. Kohl, McLaughlin, and Nagle (2006) conducted a "descriptive investigation" of the alternate assessments in 16 randomly-sampled states to determine "how states understand and implement federal policy relating to the assessment of students with the most significant cognitive disabilities."[30] At the time of the study, the authors found, "all 16 states ha[d] developed standards in content areas beyond those required under NCLB. In addition to reading, math, and science, most of the states had content standards in writing and social studies, history, or civics. Slightly less common were standards in foreign language, fine arts, health/physical education, and technology."[31] Importantly, at the time of the study (2006), "thirteen states ha[d] generated statewide alternate assessments and the three remaining states allow[ed] local districts to select which alternate assessments will be administered […] One of the three states provides a performance rubric of functional domains and indicators, but allows districts to choose which target indicators will be assessed."[32]

In 2009 the U.S. Department of Education prepared profiles of all 50 states and the District of Columbia, likewise using 2006-2007 data, noting that "a number of states had begun reworking their alternate assessments at that time."[33] Similarly, in its own review of state Department of Education websites in late 2016, Hanover finds that many states are currently "reworking" their systems once again in response to the transition from NCLB to ESSA as well as to the emergence of two federally-supported alternate assessment efforts

[28] Towles-Reeves et al., "An Analysis of the Learning Characteristics," Op. cit., p. 241.
[29] Kettler et al., "What Do Alternate Assessments," Op. cit., pp. 469-470.
[30] Kohl, F.L., M.J. McLaughlin, and K. Nagle. "Alternate Achievement Standards and Assessments: A Descriptive Investigation of 16 States." *Exceptional Children*, 73:1 (2006), p. 110. EBSCO.
[31] Ibid., p. 111.
[32] Ibid., p. 116.
[33] Cameto, R., et al. "State Profiles on Alternate Assessments Based on Alternate Achievement Standards: A Report from the National Study on Alternate Assessments." U.S. Department of Education, August 2009, p. 3. https://ies.ed.gov/ncser/pdf/20093013.pdf

described later in this section. Opportunities for new alternate assessment options remain to be determined.

Figure 2.1 below catalogues the state alternate assessment options currently available in each of the 50 states and the District of Columbia. **Note that each state has an alternate assessment option in place.** This figure also indicates whether the state explicitly links its AA-AAS system to one of the two prominent national AA-AAS, discussed later in this section.

Figure 2.1: Overview of State Alternate Assessments

State	State Assessment Option(s)	NCSC/ MSAA	DLM
Alabama	Alabama Alternate Assessment (AAA)		
Alaska	*Dynamic Learning Maps*		✓
Arizona	Arizona Instrument to Measure Standards-Alternate (AIMS-A); *Multi-State Alternate Assessment*	✓	
Arkansas	Arkansas Alternate Portfolio Assessment (AAPA) in science, and *Multi-State Alternate Assessment* in English and math	✓	
California	California Alternate Assessments (CAA)	?	
Colorado	Colorado Alternate Assessment (CoAlt), using *Dynamic Learning Maps* for English and math		✓
Connecticut	Connecticut Alternate Assessment (CTAA) for English and math, Connecticut Mastery Test (CMT) and Connecticut Academic Performance Test (CAPT) for science	✓	
Delaware	DCAS-Alt1		
District of Columbia	*Multi-State Alternate Assessment* in English and math, and DC Science Alternate Assessment (DC Science Alt)	✓	
Florida	Florida Standards Alternate Assessment (FSAA)		
Georgia	Georgia Alternate Assessment (GAA)		
Hawaii	Hawaii State Alternate Assessment (HSA-Alt)		
Idaho	Idaho Alternate Assessments (IAA)		
Illinois	*Dynamic Learning Maps*		✓
Indiana	Indiana Standards Tool for Alternate Report (ISTAR)		
Iowa	*Dynamic Learning Maps*		✓
Kansas	*Dynamic Learning Maps* and Essential Elements		✓
Kentucky	Kentucky Alternate Assessment Program (Alternate K-PREP)		
Louisiana	Louisiana Educational Assessment Program Alternate Assessment (LAA)		
Maine	*Multi-State Alternate Assessment*	✓	
Maryland	Alternate Maryland School Assessment (ALT-MSA), being replaced by *NCSC Alternate Assessments*	✓	
Massachusetts	Massachusetts Comprehensive Assessment System Alternate Assessment (MCAS-Alt)		
Michigan	MI-Access		
Minnesota	Minnesota Test of Academic Skills (MTAS)		
Mississippi	Mississippi Assessment Program-Alternate (MAP-A)		✓
Missouri	MAP-Alternate Assessment using *Dynamic Learning Maps*		✓
Montana	*Multi-State Alternate Assessment* for English and math, and Science Criterion-Referenced Test– Alternate (CRT-ALT)	✓	
Nebraska	Nebraska State Accountability Alternate Assessment (NeSA-AA)		

State	State Assessment Option(s)	NCSC/ MSAA	DLM
Nevada	Nevada Alternate Assessment (NAA)		
New Hampshire	*Dynamic Learning Maps* for English and math, and Alternate Learning Progressions (ALP) Science Assessment		✓
New Jersey	*Dynamic Learning Maps* for English and math, and Alternate Proficiency Assessment (APA) in science		✓
New Mexico	New Mexico Alternate Performance Assessment (NMAPA)		
New York	*Dynamic Learning Maps* for English and math, and New York State Alternate Assessment (NYSAA) in science and social studies		✓
North Carolina	NCEXTEND1 Alternate Assessments		
North Dakota	*Dynamic Learning Maps* for English and math, and North Dakota Alternate Assessment (NDAA) for Science		✓
Ohio	Alternate Assessment for Students with Significant Cognitive Disabilities (AASCD)		
Oklahoma	Oklahoma Alternate Assessment Program (OAAP) using *Dynamic Learning Maps*		✓
Oregon	Oregon Extended Assessment		
Pennsylvania	Pennsylvania Alternate System of Assessment (PASA)		
Rhode Island	*Multi-State Alternate Assessment* for English and math, Rhode Island Alternate Assessment (RIAA) in science	✓	
South Carolina	*Multi-State Alternate Assessment* for English and math, and South Carolina Alternate Assessment (SC-Alt) in Science and Social Studies	✓	
South Dakota	*Multi-State Alternate Assessment*	✓	
Tennessee	*Multi-State Alternate Assessment* for English and math, and TCAP/Alternate (TCAP/Alt) Assessment for Social Studies and Science	✓	
Texas	STAAR Alternate 2		
Utah	*Dynamic Learning Maps*		✓
Vermont	*Dynamic Learning Maps* for English and math, Vermont Alternate Assessment Portfolio (VTAAP) in science		✓
Virginia	Virginia Alternate Assessment Program (VAAP)		
Washington	Washington Access to Instruction & Measurement (WA-AIM)		
West Virginia	*Dynamic Learning Maps*		✓
Wisconsin	*Dynamic Learning Maps*		✓
Wyoming	Wyoming Alternate Assessment for Students with Significant Cognitive Disabilities (Wy-ALT)		
Total Adopters		**11-12**	**16**

Source: State Departments of Education

Of the three key qualities in an effective alternate assessment described in Section I — construct validity, research base, and appropriateness — **construct validity is most securely established across evaluations of various state options**. Elliot and Roach (2007) describe three key "technical issues that confront developers and users of these assessments […:] (a) alignment of content (knowledge and skills) expected to be taught and learned with the

content assessed, (b) scoring of students' knowledge and skills, and (c) standards for determining the proficient performances of students."[34]

Kettler et al. examine the state alternate assessments in Indiana, Arizona, Nevada, Idaho, Mississippi, and Hawaii in terms of the construct validity, following on the work of the limited prior research[35] and in response to the sense of a research gap shared with Towles-Reeves et al. In examining construct validity of the AA-AAS in these sample states in 2007, the authors find that state AA-AAS are generally representative of "single, unitary constructs" such as reading or mathematics ability. However, these AA-AAS instruments "often measure a number of constructs" beyond basic academic skills: "Interwoven in these measures for students performing at the extreme extensions to the grade-level standards are features of academic readiness and functional skills."[36] Elliott, Compton, and Roach (2007) had similarly found that evidence for construct validity related to the Idaho AA-AAS was "mixed, yet promising,"[37] as had Roach, Elliott, and Webb (2005) in a separate evaluation of the Wisconsin AA-AAS.[38]

In 2006 Flowers, Browder, and Ahlgrim-Delzell analyzed alignment of three anonymous states' alternate assessments to the general curriculum. Unlike other researchers, they find that "none" of their sample of state alternate assessments "met the recommended level for any of the alignment criteria established for general assessments" — but neither do general education assessments.[39]

REVIEW OF ALTERNATE ASSESSMENT OPTIONS

The subsections below briefly profile two widely-available alternate assessment systems developed by the National Center and State Collaborative (NCSC) and Dynamic Learning Maps (DLM), respectively. The NCSC and DLM options were described as "the two federally-funded consortia that created tests aligned with the Common Core State Standards for students with severe cognitive disabilities."[40] **Given growth of options in response to federal and state educational mandates, the U.S. Department of Education was encouraged that these two grant-winners would "move the field forward" in alignment with the PARCC and Smarter Balanced general student academic assessments related to Common Core.[41]**

---

[34] Elliot and Roach, "Alternate Assessments…Challenges," Op. cit., p. 302.
[35] Elliot, S.N., E. Compton, and A.T. Roach. "Building Validity Evidence for Scores on a State-Wide Alternate Assessment: A Contrasting Groups, Multimethod Approach." *Educational Measurement: Issues & Practice*, 26:2 (June 2007). EBSCO.
[36] Kettler et al., "What Do Alternate Assessments," Op. cit., pp. 470-472.
[37] Elliott, Compton, and Roach, "Building Validity Evidence," Op. cit.
[38] Roach, Elliott, and Webb, "Alignment of an Alternate Assessment," Op. cit.
[39] Flowers, C., D. Browder, and L. Ahlgrim-Delzell. "An Analysis of Three States' Alignment Between Language Arts and Mthematics Standards and Alternate Assessments." *Exceptional Children* 72:2 (Winter 2006), p. 211. ProQuest.
[40] Samuels, C. "ESSA Panel Weighs Rules for Testing for Those With Severe Cognitive Disabilities." *Education Week*, March 25, 2016. http://blogs.edweek.org/edweek/speced/2016/03/essa_testing_severe_disabilities.html
[41] "U.S. Education Department Awards Grants to Improve Assessments for Students with Disabilities." U.S. Department of Education, October 4, 2010. http://www.ed.gov/news/press-releases/us-education-department-awards-grants-improve-assessments-students-disabilities

Per Hanover's review of state Departments of Education websites in late 2016, 28 states have explicitly adopted either the NCSC/MSAA or DLM models for assessment, in whole or in part (see Figure 2.1). Some adopt the foundational academic achievement indicators to develop their own systems administered through third-party companies. Mississippi, for example, uses the DLM system through the current academic year (2015-16), but beginning in 2016-17 students will be tested using a new alternate assessment based on DLM but administered by Questar.[42] Others use the NCSC/MSAA or DLM model for assessment in English language arts and mathematics, while continuing to use state-developed alternate assessments for other subject areas, particularly science, such as New Hampshire.[43]

Note that states have other options beyond these two models, including other research-based, criterion-referenced tests developed by major research organizations like American Institutes for Research (AIR).[44] Hanover reviews the SANDI as one such option available to states and districts.

National Center and State Collaborative/Multi-State Alternate Assessment

The National Center and State Collaborative (NCSC), a consortium of "five centers and 24 states,"[45] created an alternate assessment fully rolled out in the 2015-16 academic year[46] "that is built […] on powerful validity arguments linked to clear learning outcomes and defensible assessment results."[47] The alternate assessment (AA-AAS) developed by NCSC is considered the foundational offering of a planned larger "system intended to support educators, which includes formative assessment tools and strategies, professional development on appropriate interim uses of data for progress monitoring, and management systems to ease the burdens of administration and documentation."[48]

The field testing report from February 2015 identifies core components of the NCSC AA-AAS, available for Grades 3-8 and 11 mathematics and English language arts, as follows:

- Around **30-35 items for each subject**, mostly selected response; one writing prompt per grade that accommodates multiple modes of expression
- Direct student interaction with **online testing** program or the teacher may print out testing materials and enter student responses into the computer

[42] "Mississippi Assessment Program-Alternate (MAP-A)." Mississippi Department of Education. http://www.mde.k12.ms.us/OSA/SP/mississippi-assessment-program-alternate-(map-a)
[43] "New Hampshire's Alternate Assessment Programs." New Hampshire Department of Education. http://education.nh.gov/instruction/assessment/alt_assess/
[44] "Alternate Assessments." American Institutes for Research. http://www.air.org/page/alternate-assessments
[45] Home page. National Center and State Collaborative. http://www.ncscpartners.org/
[46] "Project Timeline." National Center and State Collaborative. http://www.ncscpartners.org/project-timeline
[47] "About." National Center and State Collaborative. http://www.ncscpartners.org/about
[48] Ibid.

- Approximately **1.5 to 2 hours for each assessment** (math and ELA), permitting smaller time slots over a 6- to 8-week period to meet the student's needs[49]

NCSC's alternate assessments are delivered through a "comprehensive technology system" contracted through CTB/McGraw-Hill via a procurement grant in 2013.[50] State licenses became available in December 2015, and agency for distribution was transferred to edCount Management in late 2016 following the conclusion of the NCSC federal grant period.[51] Per edCount's management of NCSC intellectual property rights related to this assessment, "entities requesting use of the NCSC system, the system code, and/or test content or to view the system, code, and/or test content, must follow appropriate licensing procedures" by contacting edCount Management directly via email or telephone.[52] With this transfer, the nomenclature for this system has changed (somewhat inconsistently across states) to Multi-State Alternate Assessment (MSAA).

As a participating member state in NCSC, California chose to pilot and promote alternate assessments linked to the NCSC AA-AAS crosswalk of "essential understandings" simplifying requirements of the Common Core State Standards, referred to as Core Content Connectors on the state Department of Education webpage dedicated to alternate assessment. Alternate assessments in mathematics and ELA will be available in January, as will pilot alternate assessments for Grade 5 and 8 science. The Department notes, "[i]n January, new grade-specific practice tests will be available for ELA and mathematic which will allow teachers and students to become familiar with using the various embedded and non-embedded resources to establish the best possible setting for the student's testing experience."[53] However, it appears that the state is no longer a direct participant in the NCSC/MSAA program,[54] and it continues to battle U.S. DoE regulations on alternate assessment pilot testing periods.[55]

[49] Bulleted items taken verbatim from: "Inclusive Assessment Design, Development, Evaluation, and Data; Accessibility Strategies; and Voices from the Field." National Center and State Collaborative, February 2015, p. 13. http://www.ncscpartners.org/Media/Default/PDFs/Resources/NCSC-Accessibility-Information-from-Design-through-Field-Testing-v3.pdf

[50] "Procurement." National Center and State Collaborative, October 9, 2013. http://www.ncscpartners.org/procurement

[51] [1] "Licenses for Use of NCSC System and Content Now Available for States." National Center and State Collaborative, December 9, 2015. http://www.ncscpartners.org/news/licenses-for-use-of-ncsc-system-and-content-now-available-for-states
[2] "NCSC Project Transition to the Multi-State Alternate Assessment States." National Center and State Collaborative. http://www.ncscpartners.org/news/ncsc-project-transition-to-the-multi-state-alternate-assessment-states

[52] Home page. edCount Management LLC. http://www.edcountmanagement.com/

[53] "California Alternate Assessments." California Department of Education, November 28, 2016. http://www.cde.ca.gov/ta/tg/ca/altassessment.asp

[54] Udesky, L. "Alternate assessments for special education students delayed." EdSource, May 11, 2015. https://edsource.org/2015/alternate-assessments-for-special-education-students-delayed/79309

[55] Maio, P. "California appeals to federal officials to administer one statewide science test, not two." EdSource, November 30, 2016. https://edsource.org/2016/california-appeals-to-federal-officials-to-administer-one-statewide-science-test-not-two/573545

DYNAMIC LEARNING MAPS

Dynamic Learning Maps is an online, adaptive alternate assessment available to partner states for integrated (formative) or end-of-year (summative) testing.[56] Like NCSC, DLM development was funded through a federal grant and managed by "the Center for Educational Testing and Evaluation at the University of Kansas, a part of the Achievement and Assessment Institute, in partnership with the Center for Literacy and Disability Studies at the University of North Carolina, Chapel Hill." (California is not currently a partner state with DLM, either for integrated or end-of-year assessment.) DLM is now a "self-funded" and ongoing operation.[57]

DLM assessments are available for "mathematics, English language arts, and science," and administered through its proprietary KITE Client online system. While DLM assessments are designed primarily as online, adaptive "testlets" with a variety of "unique accessibility tools and supports," the design incorporates certain elements of teacher interaction "outside the [testing] system" as additional options "to fit each student's needs and preferences."[58] On a page associated with DLM's set of "released testlets," it highlights the following digital functionalities as "common test features," including:

- **Educator directions**: Instructions to help the test administrator deliver testlets to students. Only found in teacher-administered testlets.
- **Next/back buttons**: For navigation forward and backward through a testlet before submitting answers.
- **Exit Does Not Save button**: Stops the testlet without saving answers. The student will start at the beginning of that testlet when logging back in.
- **Review screen**: For reviewing answers and making changes before ending the testlet.[59]

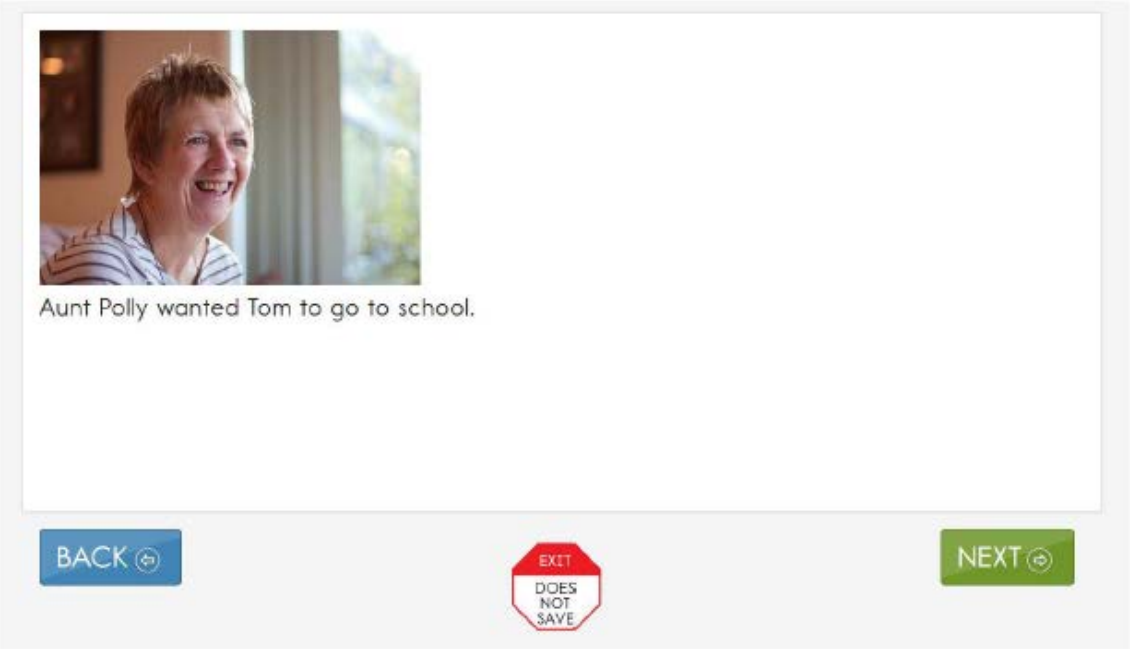
As illustrated in a sample released testlet, testlets are designed around specific "Essential Elements" of content learning and target a specific "node" within that element — for example, that the student "can make judgments about the meaning of word(s)" as a node within the Grade 8 element "determine connotative meanings of words and phrases in a text." The sample testlet features a series of single, illustrated sentences composing a larger narrative. The teacher is instructed to read these statements with the student and "maximize your interaction" through prompting such as "pointing to objects in the pictures" and using "leading comments," "sounds," and "actions when appropriate."[60]

Then, prior to a second reading of the same story, the teacher is given context for the assessment before the assessment questions begin: "Because this testlet addresses

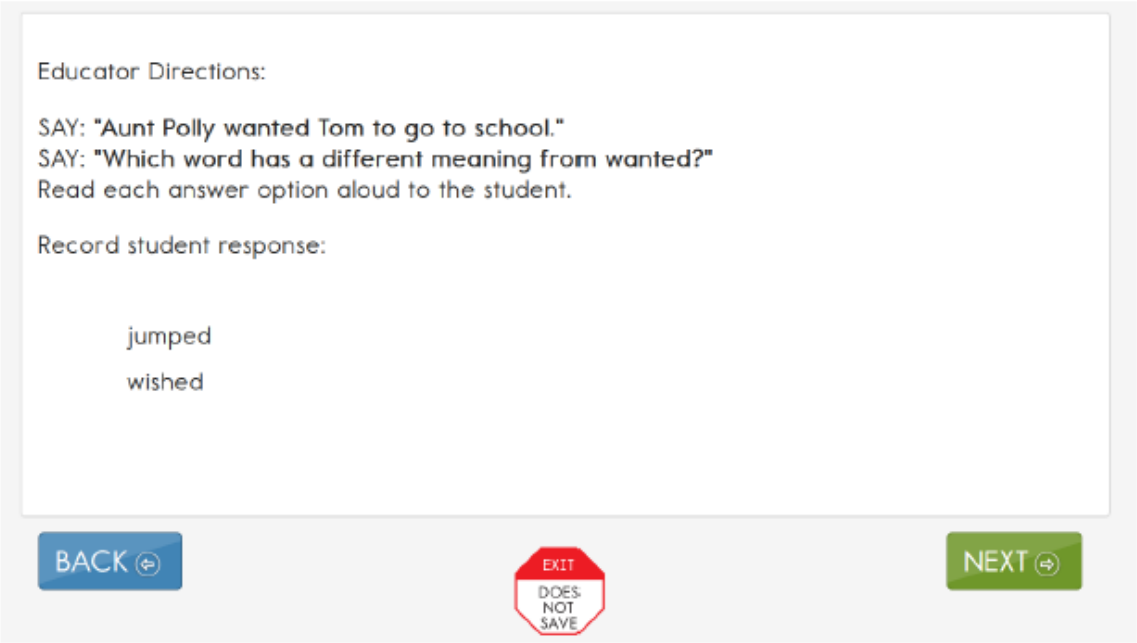
[56] "About DLM Tests." Dynamic Learning Maps. http://dynamiclearningmaps.org/about/tests
[57] "About the Consortium." Dynamic Learning Maps. http://dynamiclearningmaps.org/about/consortium
[58] "About DLM Tests," Op. cit.
[59] Bulleted text taken verbatim from: "Released Testlets." Dynamic Learning Maps. http://dynamiclearningmaps.org/about/tests/releasedtestlets
[60] "Released Testlet ELA.RL.8.4.IP." Dynamic Learning Maps. dynamiclearningmaps.org/sites/default/files/documents/ELA.RL.8.4.IP.pdf

foundational skills, the questions focus on the student's ability to make judgments about the meaning of words."[61] A sample question from this testlet is reproduced in the figure below.

Figure 2.2: Sample Released Testlet Item from DLM, Grade 8 English Language Arts



[61] Ibid., p. 9.



Source: DLM[62]

STUDENT ANNUAL NEEDS DETERMINATION INVENTORY

Riverside County Office of Education (RCOE) developed the Student Annual Need Determination Inventory (SANDI) as a comprehensive district-wide system for guiding both summative and formative assessment for students with intellectual disabilities (ID), and to prepare students for the rigor of the state Alternate Assessment. Reference in the work of DuFour, DuFour, and Eakter (2008),) [63], the SANDI uses the Professional Learning Community (PLC) model. The PLC model serves as a vehicle to: identify what students are expected to learn, whether Common Core or new state standards; identify how we will know if they have learned it through meaningful authentic reliable assessment; and determine how the learning community can support struggling students using evidence-based practices. The first edition of the SANDI was introduced to teachers in Riverside County in 2002 as an assessment of pre-academic/academic, motor, daily living and behavioral skills for students with intellectual disabilities. It was aligned to the California State Standards and the California Alternate Performance Assessment (CAPA) blueprint and served as a tool for assessing the need areas of each student ages 5-22, for the purpose of writing and implementing standards-based IEP goals. Now in its third edition, the SANDI is an online assessment system aligned to Common Core State Standards (CCSS), with the goal of providing access to new state standards and meaningful summative and formative assessment for students with significant disabilities.

[62] Ibid., pp. 13-14.
[63] DuFour, R., R. DuFour, R., and R. Eaker. *Revisiting professional learning communities at work: New insights for improving schools*. Bloomington, IN: Solution Tree, 2008.

The SANDI is currently administered to over 25,000 students with intellectual disabilities across the United States. Consistent use of the SANDI assessment system has demonstrated student achievement for students with disabilities on state alternate assessments. The CAPA Adequate Yearly Progress (AYP) data for RCOE from 2005-2013 demonstrate continuous growth in student achievement in both ELA and Math: from 55 percent proficient in English Language Arts in 2005 to 92 percent proficient in 2013; and from 42 percent proficient in Math in 2005 to 82 percent proficient in 2013. In independent evaluations of the SANDI, Hanover Research has demonstrated:

- the correlation of the SANDI to three other assessments — California Alternate Performance Assessment (CAPA), Woodcock-Johnson-III (WJ), and Vineland-II (Vineland) —;
- high inter-rater reliability of items in reading, writing, communication, and math; and
- strong content validity within individual items, particularly in terms of alignment, representativeness, and clarity.

Each of these three studies is described briefly below.

SANDI Correlation Study

For the correlation study, RCOE provided Hanover with three types of student-level data: SANDI outcomes; CAPA Outcomes; and Vineland/WJ outcomes (Figure 2.3). All the data are available in 2012-13 and 2013-14.[64]

Figure 2.3: Correlation Study Data Overview

Data Type	Data Content	Year
SANDI	▪ Reading, Mathematics, Writing, and Communication outcomes	2012-13 and 2013-14
CAPA	▪ Reading and Mathematics outcomes	2012-13 and 2013-14
Vineland and WJ	▪ Woodcock-Johnson III W scores: Brief Reading and Brief Mathematics ▪ Vineland-II standard scores: Communication and Socialization ▪ Vineland-II raw scores: Communication domain: Receptive, Expressive, and Written; Socialization domain: Interpersonal Relationship, Play and Leisure Time, and Coping Skills	2012-13 and 2013-14

This study examines the correlations between students' SANDI outcomes and their performance on the three other assessments (CAPA, WJ, and Vineland). We use combined two-year assessment outcomes because the total number of students for Vineland and WJ assessments in 2013-14 is only between 2 to 16.

To estimate the correlation between students' SANDI outcomes and their performance on other assessments by assessment area, we provide separate comparisons for each SANDI content-area. Figure 2.4 presents the CAPA, Vineland, and WJ outcomes that we use to calculate the correlations for each SANDI content-area. For SANDI Reading and Writing outcomes, we use corresponding assessment area outcomes in CAPA and WJ. For SANDI Writing outcomes, we use Vineland-II Written raw scores. Finally, for SANDI Communication outcomes, we use both Vineland-II standard scores and Vineland-II raw scores in Communication domain and Socialization domain, except Written raw score. Even though the Written subdomain is under Vineland Communication, its content is not similar to that of the SANDI Communication section. In the Vineland-II assessment, the Written subdomain measures "[what] the individual understands about how letters make words, and what he or she read and writes," and we do not find similar items in the SANDI Communication section.[65]

Figure 2.4: Correlation Analysis Summary

SANDI Content-Area	CAPA or VWJ Subject	Assessment Type

[64] SANDI data is also available in 2010-11, 2011-12, and 2014-15. As CAPA data and VWJ data are only available in 2012-13 and 2013-14, we only include these two years' data in analysis.

[65] "Review of the Vineland Adaptive Behavior Scales-Second Edition (Vineland-II)." University of Alberta, Community-University Partnership for the Study of Children, Youth, and Families, 2011, p. 3. https://d1pbog36rugm0t.cloudfront.net/-/media/ualberta/faculties-and-programs/centres-institutes/community-university-partnership/resources/tools---assessment/vinelandjune-2012.pdf

SANDI Content-Area	CAPA or VWJ Subject	Assessment Type
Reading	Reading	CAPA
	Brief Reading	Woodcock-Johnson III
Mathematics	Mathematics	CAPA
	Brief Mathematics	Woodcock-Johnson III
Writing	Written	Vineland-II Subdomain Raw Scores
Communication	Communication	Vineland-II Domain Standard Scores
	Receptive	Vineland-II Subdomain Raw Scores
	Expressive	Vineland-II Subdomain Raw Scores
	Socialization	Vineland-II Domain Standard Scores
	Interpersonal Relationship	Vineland-II Subdomain Raw Scores
	Play and Leisure Time	Vineland-II Subdomain Raw Scores
	Coping Skills	Vineland-II Subdomain Raw Scores

In Figure 2.5, we present summary statistics for each assessment. Fewer than 25 students have WJ and Vineland outcomes in 2012-13 and 2013-14. Because of such low counts, the correlation results between SANDI and WJ outcomes and correlation results between SANDI and Vineland outcomes may not be accurate. Score ranges and average scores vary greatly by assessment, but these differences do affect correlations between any pair of outcomes. Correlation estimates depend on the degree to which outcome values move together and the direction of movement, and not strictly on the proximity of the actual values.

Figure 2.5: Assessment Outcome Summary

Assessment	N	Mean	SD	Minimum	Maximum
SANDI					
Reading	1,079	173.56	110.06	2	435
Writing	1,041	96.01	63.08	1	264
Mathematics	1,040	106.24	96.25	1	368
Communication	1,022	152.70	82.12	1	324
CAPA					
Reading	365	45.12	9.40	15	60
Mathematics	365	39.91	7.98	15	60
Woodcock-Johnson III					
Brief Reading	28	383.25	40.40	316	463
Brief Mathematics	13	437.38	20.47	414	476
Vineland-II Domain Standard Scores					
Communication	24	60.79	17.32	23	99
Socialization	24	62.13	12.17	39	96
Vineland-II Subdomain Raw Scores					
Communication					
Receptive	24	13.17	8.19	1	40
Expressive	23	22.35	18.01	1	75
Written	19	11.58	8.03	1	28
Socialization					
Interpersonal Relationship	23	14.91	12.69	1	50
Play and Leisure Time	24	12.92	9.87	1	38
Coping Skills	23	12.26	8.80	1	30

Overall, as presented in Figure 2.6, we find that **students' SANDI outcomes are highly and significantly correlated with their performance on CAPA, WJ, and Vineland assessments**. In the Reading and Mathematics assessment areas, SANDI outcomes exhibit correlation that is much higher with WJ outcomes than with CAPA outcomes. This result is especially clear since we find higher correlation estimates between SANDI and WJ, even though very few students took the WJ assessments. For Reading, the correlation between SANDI and WJ is about 0.88, higher than the correlation between SANDI and CAPA outcomes (0.43). For Mathematics, the correlation between SANDI and WJ is about 0.71, higher than the correlation between SANDI and CAPA outcomes (0.36).

Furthermore, SANDI Writing and SANDI Communication are highly and significantly correlated with Vineland outcomes. SANDI Writing score is highly correlated with Vineland Written raw score, and the correlation coefficient is about 0.77. Among Vineland raw score outcomes, SANDI Communication has the highest correlation with Vineland Coping Skills (0.79). Among Vineland standard score outcomes, SANDI Communication has higher correlation with Vineland Socialization (0.77) than with Vineland Communication (0.74), though the difference is small. Between the two subdomains of Vineland Communication that we correlate with SANDI Communication, students' outcomes on the Expressive subdomain exhibit greater correlation (0.76) than do their Receptive subdomain outcomes (0.63). Among the three subdomains of Vineland Socialization that we correlate with SANDI

Communication, students' outcomes on the Coping Skills subdomain has higher correlation (0.79) than do either Interpersonal Relationship (0.50) or Play and Leisure Time (0.47).

Figure 2.6: Correlation Results

Assessment	Correlation Coefficients	N
SANDI Reading		
CAPA Reading	0.4309***	336
Woodcock-Johnson III Brief Reading	0.8770***	26
SANDI Mathematics		
CAPA Mathematics	0.3616***	328
Woodcock-Johnson III Brief Mathematics	0.7132***	13
SANDI Writing		
Vineland Written	0.7673***	15
SANDI Communication		
Vineland Communication	0.7409***	17
Receptive	0.6315***	17
Expressive	0.7564***	16
Vineland Socialization	0.7731***	17
Interpersonal Relationship	0.4978**	17
Play and Leisure Time	0.4740*	17
Coping Skills	0.7871***	16

*** p<0.01, ** p<0.05, * p<0.1.

SANDI Inter-Rater Reliability Study

The inter-rater reliability study examines the level of agreement between special education teachers on subject items from the SANDI. Teachers in Riverside County and New York City were asked to rate 24 items in four subject areas after viewing a set of video clips. The purpose of these ratings is to ensure that a teacher who administers the SANDI to students is able to assign ratings that coincide with other teachers' scores. We compute inter-rater reliability measures and apply other descriptive analyses to teachers' ratings to quantify the level of agreement in each item.

Overall, the teachers' ratings information is collected from 171 special education teachers. Teachers provide these ratings, ranked on a scale of 0 to 4, for 24 SANDI items across four subject areas: Reading, Writing, Communication, and Math. A final set of 159 teachers rated each of the 24 items. Seven teachers did not provide a rating for one item, another teacher submitted an invalid rating of "12" for M5, and five teachers did not provide ratings for more than one item. These missing values only have a small impact on the analysis of the inter-rater agreement, which we elaborate below.

To quantify the level of agreement between SANDI item ratings, Hanover computes Fleiss's kappa values for all ratings overall and for ratings in each subject area. Fleiss's kappa is the appropriate measure of inter-rater agreement or reliability (IRR) for this study since we observe

ratings on the same scale from more than two raters. Hanover also performs significance testing to determine whether the level of agreement among teachers is statistically different from random agreement, denoted by a kappa of zero. While kappa values can be negative, they typically range from 0 to 1, where higher values represent higher levels of agreement. Figure 2.7 presents a kappa value categorization that is used prevalently in the field of psychology.

Figure 2.7: Interpretation of IRR Measure (Kappa)[66]

Kappa	Interpretation
< 0	Less than chance agreement[67]
0.01 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 0.99	Almost perfect agreement

While some researchers have criticized this categorization as subjective,[68] the criticism applies primarily to situations with very few (e.g., binary rating) or very many rating categories. Since teachers provided ratings for SANDI items on a five-point scale, we believe that traditional interpretation of kappa values is still helpful for this study.

Figure 2.8 summarizes the level of agreement in ratings on all items overall and for items within each subject area. Overall, the level of agreement is high, denoted by a kappa of 0.70 ("substantial agreement"). There is variation in the level of agreement among teachers in their ratings of different subjects' items, from a kappa of 0.52 in Communication ("moderate agreement") to 0.82 in Math ("almost perfect agreement"). The level of agreement among teachers, in each subject and overall, is significantly higher than random agreement.

Figure 2.8: Fleiss's Kappa of Teachers' Ratings by Subject Area

Subject Area	Kappa	Z-Statistic	P-Value	Interpretation
Overall	0.70	702.7	<0.0001	Substantial agreement
Reading	0.72	301.9	<0.0001	Substantial agreement
Writing	0.65	332.3	<0.0001	Substantial agreement
Communication	0.52	267.3	<0.0001	Moderate agreement
Math	0.82	341.24	<0.0001	Almost perfect agreement

SANDI Content Validity Study

In the content validity study, experts in special education evaluated each item in the four different content areas of the SANDI: reading/ELA, writing, communication, and math.

[66] Landis, J.R., and G. Koch. "The Measurement of Observer Agreement for Categorical Data." *Biometrics*, 33:1, March 1977, pp. 159-74.
[67] A kappa of exactly 0 indicates the same level of agreement as would be expected had the ratings been given randomly.
[68] Gwet, K. L. *Handbook of Inter-Rater Reliability*. Fourth ed. Gaithersburg, Maryland: Advanced Analytics, LLC, 2014.

These expert ratings were provided to Hanover by RCOE. Seven to nine experts evaluated each content area, assessing items in terms of the following criteria:

- **Representativeness** – Ability of the test item to represent the content of that section based on incorporating the functional skill and the Common Core State Standard (CCSS) for the item; this item is scored from 1 to 4, with 1 indicating the lowest level of representativeness and 4 indicating the highest level of representativeness.
- **Clarity** – How clearly the item is worded; this item is scored from 1 to 4, with 1 indicating the lowest level of clarity and 4 indicating the highest level of clarity.
- **Factor Structure**[69] – How well the item demonstrates the skill of the content CCSS; this item is scored as 0 or 1, with 0 indicating incorrect standard alignment and 1 indicating correct standard alignment.

To analyze the scores provided by the experts, Hanover calculates an average for each of the three criteria (representativeness, clarity, and factor structure) for each item within the four content areas. For the representativeness and clarity scores, in addition to the average, the average absolute deviation (AAD) is also calculated. This measure calculates the average deviation of each expert's score from the mean score. As such, it provides a measure of the variation among experts in the scores they gave for a particular item. Since the factor structure metric is binary (taking only values of 0 or 1), a measure of spread is not needed for these scores.

In addition to calculating these averages and measures of variation for each individual assessment item, overall averages, across all items within a given content area, are also calculated. These overall averages are specified in Figure 2.9 below.

Figure 2.9: Overall Averages Calculated for Each Content Area

Composite Average	Mean	Avg. Abs. Dev.
Overall Inter-Rater Agreement	The average of all **representativeness** and **clarity** scores within a given content area	The average of all AADs calculated across the **representativeness** and **clarity** scores within a given content area
Overall Content Validity	The average of all **representativeness** scores within a given content area	The average of all AADs calculated across the **representativeness** scores within a given content area

[69] Note that "factor structure" and "alignment" are used interchangeably throughout this report.

Composite Average	Mean	Avg. Abs. Dev.
Overall Factor Structure Ratings	The average of all factor structure scores within a given content area. Because the **factor structure** score is binary, with a 1 indicating the item is aligned with state standards, this average represents the percentage of experts who believe the item is aligned with state standards.	-- *

*Note: The AAD is not calculated for the factor structure scores since the measure is binary.

Note that because the representativeness and clarity measures are on a scale of 1 to 4 – a relatively small range – the average absolute deviation tracks very closely with the mean. The higher the average, the less the deviation will be since scores are fluctuating within a smaller range. Because of this, the discussion presented in this report focuses primarily on the average scores rather than the variation in the scores. Furthermore, when interpreting the data presented in this subsection it is important to keep in mind that a small number of experts (seven to nine) rated each item. Thus, caution must be taken when comparing scores for individual assessment items. With such a small number of data points, there is a higher probability that the difference in scores is due to random chance.

The experts rated most content area items very highly in terms of representativeness, with little variation among participating scorers (Figure 2.10). The average of the representativeness and clarity scores ("overall inter-rater agreement") across all reading items is 3.92, with an average absolute deviation of 0.14. Likewise, the average overall inter-rater agreement scores across all writing items is 3.84, with an AAD of 0.22; for math, an overall inter-rater agreement score of 3.83, AAD of 0.24; and for communication, an overall inter-rater agreement score of 3.81, AAD of 0.32. Content validity scores across the four content areas ranged from 3.65 to 3.87.

Figure 2.10: Composite Scores by Content Area

Composite Metric	Reading (N=109)		Writing (N=69)		Communication (N=81)		Math (N=100)	
	Mean	AAD	Mean	AAD	Mean	AAD	Mean	AAD
Overall Inter-Rater Agreement[70]	3.92	0.14	3.84	0.22	3.81	0.32	3.83	0.24
Overall Content Validity[71]	3.87	0.20	3.73	0.36	3.65	0.49	3.76	0.28
Overall Factor Structure Ratings[72]	91%	- -	95%	- -	98%	- -	92%	- -

NB: Total items indicated in parentheses in content area header.

SELECTION MODELS AND CONSIDERATIONS

From their review of assistive technology literature, Watts, O'Brian, and Wojcik (2004) provide a summary of critical features of special education assessments. They note that,

[70] These scores take into account both the representativeness and clarity ratings.
[71] These scores only take into account the representativeness ratings.
[72] These scores consider the percent of experts rating item as aligned with state standards.

"[d]espite the multitude of components for quality assessment, the assessment literature is extremely consistent in addressing certain features as critical aspects of many types of assessment processes and tools." [73] These features include aspects such as a "comprehensive ecological approach," meaning that assessments consider a variety of data; and "documentation," meaning that assessments create records of the process and outcomes (see Figure 2.11).

Figure 2.11: Critical Features of Special Education Assessments

FEATURE	DEFINITION
Comprehensive Ecological Approach	Data from many potential environments and all factors within a given environment
Emphasis on Individual Supports	Data specifies individual supports for students
Technical Adequacy	Assessment process has reliability, validity, and absence of bias
Team Problem-Solving Model	A group of professionals, parents, and student work collaboratively to identify issues and plan assessment
Student Involvement	The student is systematically included in the assessment
Documentation	Recording of process and data is included in model
Student Outcomes	Assessment process provides clear indication of student achievement
Program Outcomes	A systematic method of evaluating program effectiveness is included in the model
Strength-based Model	Areas of strength are used as basis for assessment
Consistency between Framework and Process	Connection between the assumptions about knowledge and learning with the assessment process
On-going Longitudinal Approach	Assessment process allows for data collection over the course of the student's development

Source: Watts, O'Brian, and Wojcik, "Four Models," Op. cit., p. 46.

Some prominent models for evaluating general assessments may be less useful for describing the value of alternate assessment models. Flowers, Browder, and Ahlgrim-Delzell (2006) write, for example, about the misalignment of purpose related to Webb's model:

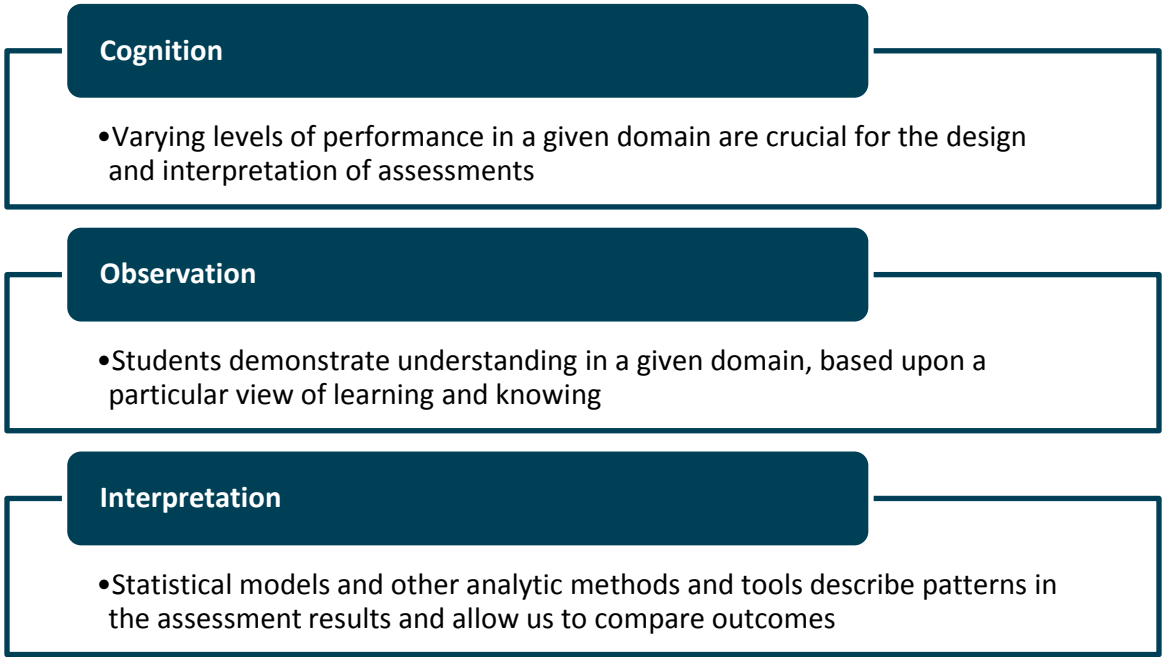
> Webb's recommendations are based on his work with general education mathematics and science assessments; these assessments tend to have many more items than alternate assessment and are usually given in a paper/pencil format. Given that alternate assessments were designed using modified academic standards (i.e., extended standards or expanded benchmarks) and alternate achievement standards, we may not find alignment statistics that meet criteria established by Webb.[74]

[73] Watts, E.H., M. O'Brian, and B.W. Wojcik. "Four Models of Assistive Technology Consideration: How Do They Compare to Recommended Educational Assessment Practices?" *Journal of Special Education Technology*, 19:1, Winter 2004, p. 45. ProQuest.

[74] Flowers, Browder, and Ahlgrim-Delzell, "An Analysis of Three States'," Op. cit., p. 211.

Pellegrino et al.'s 2001 triangle model for evaluating validity according to *cognition*, *observation*, and *interpretation*. The "assessment triangle" grounds understanding of specific development decisions made to accurately assess the academic achievement of students with significant cognitive disabilities (Figure 2.13).[77] Towles-Reeves et al. note: "Specifically, the validity evaluation of an assessment should consider two questions. First, it is necessary to know whether the assessment is appropriate for the intended population. Second, in high-stakes accountability environments, it is important to ensure that the appropriate population is in fact the population being assessed."[78]

Figure 2.13: Pellegrino et al.'s Assessment Triangle



Source: Marion and Pellegrino, "A Validity Framework," Op. cit., p. 49.

Section III of this report offers additional considerations for districts selecting among alternate assessment options.

[77] Marion, S.F., and Pellegrino, J.W. "A Validity Framework for Evaluating the Technical Quality of Alternate Assessments." *Educational Measurement: Issues & Practice*, 25:4 (Winter 2006). EBSCO.
[78] Towles-Reeves et al., "An Analysis of the Learning Characteristics," Op. cit., p. 243.

SECTION III: BEST PRACTICES IN ALTERNATE ASSESSMENT

This section situates the use of alternate assessment within the larger practice of supporting students with disabilities. It describes best practices in the use of alternate assessments and addresses considerations related to individual instructional plans (IEPs) and educator professional development systems such as professional learning communities (PLCs).

PREPARATION

Development of alternate assessment begins, in most cases, with an examination of the characteristics of the learner population who will be the audience for the new system. Towles-Reeves et al. (2009) summarize the following characteristics as "typical" of students eligible for alternate assessments: they "(a) have individualized education programs, (b) have cognitive disabilities, (c) require instruction under multiple conditions to generalize learning, and (d) may receive 'functional curricula.'"[79] These conditions might be captured by "special education labels such as autism, mental retardation, and/or multiple disabilities."[80] However, as Towles-Reeves et al. note, local conditions vary and — at the time of their writing — are not "consistently monitored" by states administering these tests.[81] Additionally, "not all students with these labels will require an alternate assessment, and students with other special education labels may also qualify for an AA-AAS."[82]

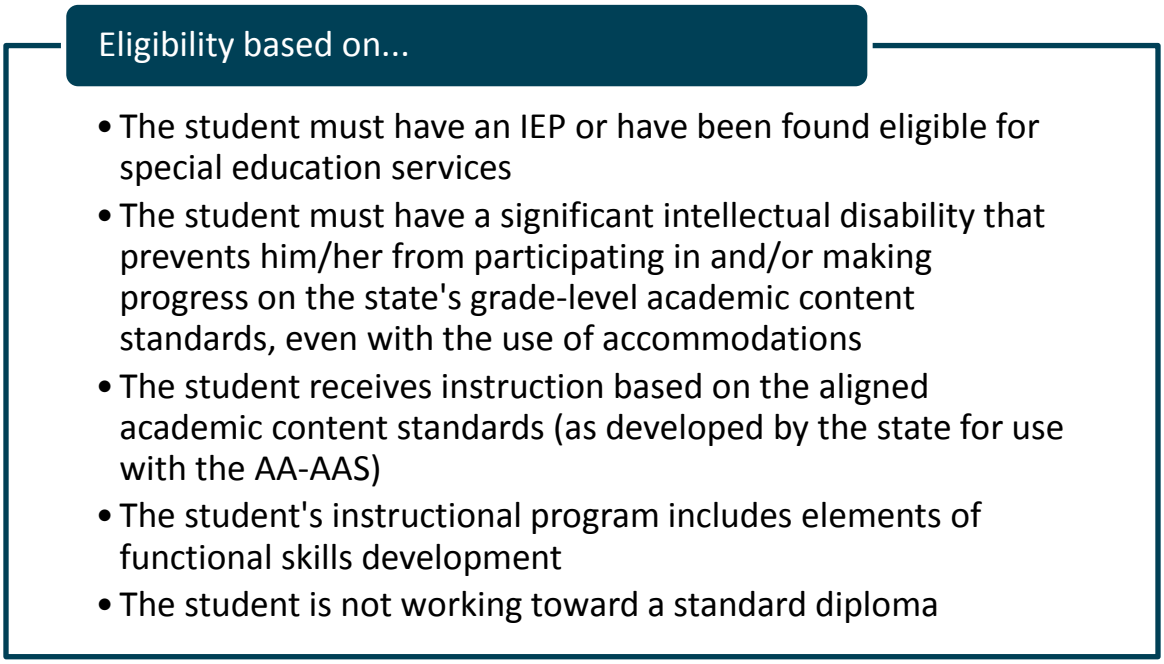
In their examination of three states, Towles-Reeves et al. (2009) describe the range of needs and accommodations that must be considered in terms of "levels of communication," "social engagement," and "functional skill." They write:

> The first set of students (and the majority of the students in our sample) have either symbolic or emerging symbolic levels of communication, evidence social engagement, and possess at least some level functional reading and math skills. The second set of students in our sample (10% to 25% of our students, depending on the measure and the state) have not yet acquired formal, symbolic communication systems, do not initiate, maintain, or respond to social interactions, and have no awareness of print, Braille, or numbers. Between these two sets of students are those who most likely represent skills and abilities characteristic, in part, of each of these groups. States must consider the educational needs of all these students in designing their AA-AAS. Most important, states will need to thoughtfully consider, especially for students at a presymbolic level of communication, how to ensure linkage to grade-level content standards in ways that provide meaningful and useful educational targets for those students.[83]

[79] Towles-Reeves et al., "An Analysis of the Learning Characteristics," Op. cit., p. 243.
[80] Towles-Reeves, Kleinert, and Muhomba, "Alternate Assessment…New," Op. cit., p. 234.
[81] Towles-Reeves et al., "An Analysis of the Learning Characteristics," Op. cit., p. 243.
[82] Towles-Reeves, Kleinert, and Muhomba, "Alternate Assessment…New," Op. cit., p. 234.
[83] Towles-Reeves et al., "An Analysis of the Learning Characteristics," Op. cit., p. 253.

Given the 1 percent cap for alternate assessment within a given state or district,[84] most states and districts have guidelines on when it would be appropriate to use an alternate assessment rather than the general assessment with accommodations. Previous work as reported in Streagle and Scott (2015) identified "common AA-AAS participation criteria" through reviews of state websites in 2007 and 2010, which collectively define a sequence of eligibility requirements demonstrating alternate assessment as the only reasonable option for testing that student's academic progress (Figure 3.1).

Figure 3.1: Common Criteria for Alternate Assessment Participation

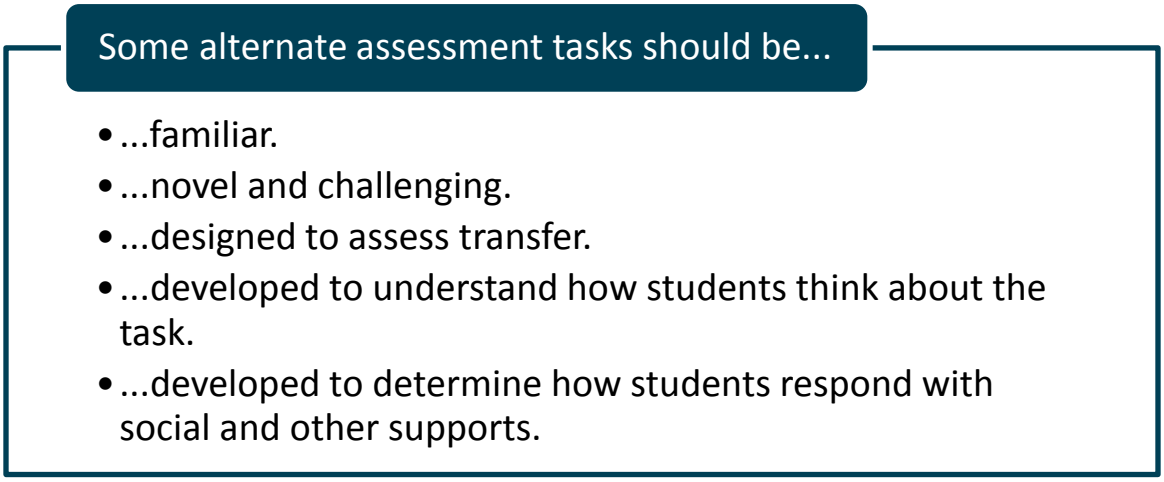


Source: Streagle and Scott, "The Alternate Assessment…Process," Op. cit., p. 1295.

Given the variety of needs served, local districts must choose from among the options available to them to best serve their specific set of students eligible for alternate assessment. In an early white paper for the National Alternate Assessment Center (NAAC), Kleinert, Browder, and Towles-Reeves (2005) propose a set of "guidelines" for identifying cognitively-appropriate "instructional and assessment practices for students with significant cognitive disabilities,"[85] illustrated in Figure 3.2. These guidelines broadly describe the variety of item types that should be represented across a given alternate assessment.

[84] Kleinert, H.L., D.M. Browder, and E.A. Towles-Reeves. "Models of Cognition for Students With Significant Cognitive Disabilities: Implications for Assessment." *Review of Educational Research*, 79:1 (March 2009), p. 303. ProQuest.
[85] Kleinert, H., D. Browder, and E. Towles-Reeves. "The assessment triangle and students with significant cognitive disabilities: Models of student cognition." National Alternate Assessment Center, White paper #1, 2005, p. 31. http://www.naacpartners.org/publications/whitePapers/18000.pdf

Figure 3.2: Characteristics of Cognitively-Appropriate Alternate Assessments



Source: Kleinert, Browder, and Towles-Reeves, "The assessment triangle," Op. cit., pp. 31-33.[86]

Elliot and Roach (2007) caution that test security is a challenge given the absence of multiple iterations of an alternate assessment. "Teachers matter" in that teachers deliver the assessment to individual students and must professionally interpret the student's response. Figure 3.3 summarizes the shared attributes of alternate assessments, which highlight the influential role of teachers in this system. The authors write, "Alternate assessments require teachers to make many judgments and consequently place teachers at the center of the resulting accountability decisions."[87] **Promisingly, Browder et al. (2005) find that "training teachers in instructional interventions [related to AA-AAS] seemed to override any potential influence of students' characteristics."[88]**

Figure 3.3: Shared Attributes of Alternate Assessments

[86] See also: Kleinert, Browder, and Towles-Reeves, "Models of Cognition," Op. cit., pp. 319-321.
[87] Elliot and Roach, "Alternate Assessments…Challenges," Op. cit., p. 312.
[88] Browder, D.M., et al. "The Impact of Teacher Training on State Alternate Assessment Scores." *Exceptional Children*, 71:3 (Spring 2005), p. 278. ProQuest.

- The knowledge and skills assessed are less complex and generally characterized as entry-level or prerequisite to the knowledge and skills outlined in grade-level content standards.
- Students' IEPs and their related instruction are acknowledged and may influence the array of classroom-based evidence collected, but the focus of the assessment is contextualized in and "controlled" by approved extended content standards in reading and mathematics (and soon science); thus, the IEP is not directly assessed.
- Teachers play a major role in the development, collection, and organization of evidence about a student's entry level or prerequisite skills.
- The assessment is managed by a student's teacher, untimed, and embedded in the student's classroom(s).
- The assessment occurs over a broader timeframe than the general education achievement testing, but concludes at nearly the same time to facilitate an integrated data management and reporting process.
- Teacher support of the student is often an explicit aspect of the assessment that must be documented in the evaluation of the collected evidence.
- The evidence is summarized using a criterion (or standard)-referenced scoring rubric that features at least three characteristics: (a) correctness of responses, (b) generalization of the response, and (c) support needed.
- The reliability of the summary score is supported by using a multi-rater agreement method.
- The translation of the student's score or performance to a proficiency level for the corresponding content areas is guided by alternate achievement standards.
- The assessment is not secure; that is, teachers know the nature and focus of the assessment tasks in advance of administration of the assessment. In some cases, states have tried to secure performance tasks by collecting assessment materials after the testing event, but because alternate forms of the assessment are not available, the same tasks are used for several years and become known to many teachers.

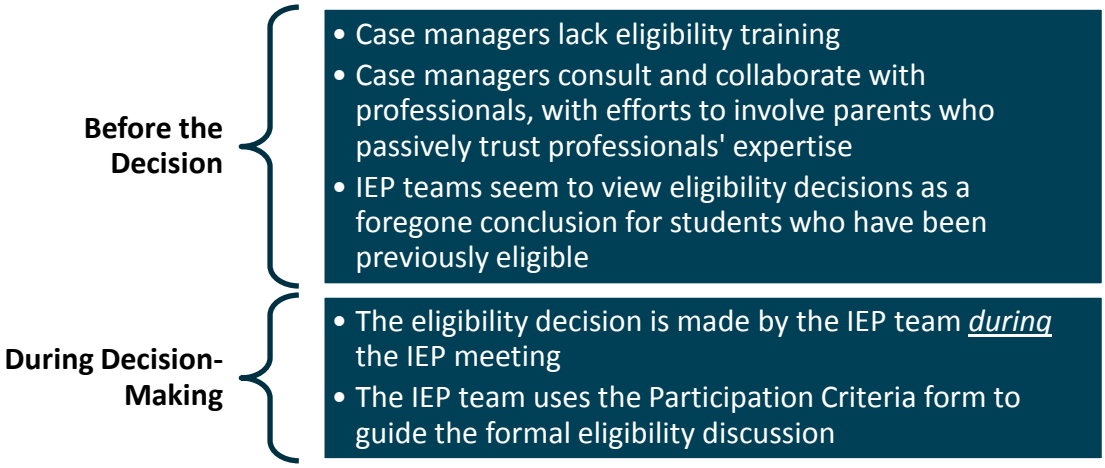
Source: Elliot and Roach, "Alternate Assessments…Challenges," Op. cit., p. 306.

IMPLEMENTATION

While Kohl, McLaughlin, and Nagle (2006) found widespread development of alternate assessments and systems, the authors were concerned about "how states fail to exert quality control over the administration of the alternate assessments."[89] For example, Streagle and Scott (2015) evaluated the appropriateness of one state's eligibility decisions, finding that students are "usually appropriately, unusually inappropriately" found eligible to take the Virginia AA-AAS.[90] However, there were several negative influences that contributed to the rare incorrect decisions regarding eligibility for some students (see Figure 3.4). **In general, these factors demonstrate opportunities for improved communication with various stakeholders in the eligibility decision-making process that will allow for more voices and more information to contribute.**

[89] Kohl, McLaughlin, and Nagle, "Alternate Achievement Standards…States," Op. cit., p. 120.
[90] Streagle and Scott, "The Alternate Assessment…Process," Op. cit., p. 1301.

Figure 3.4: Factors Influencing Accuracy of Alternate Assessment Eligibility Decisions

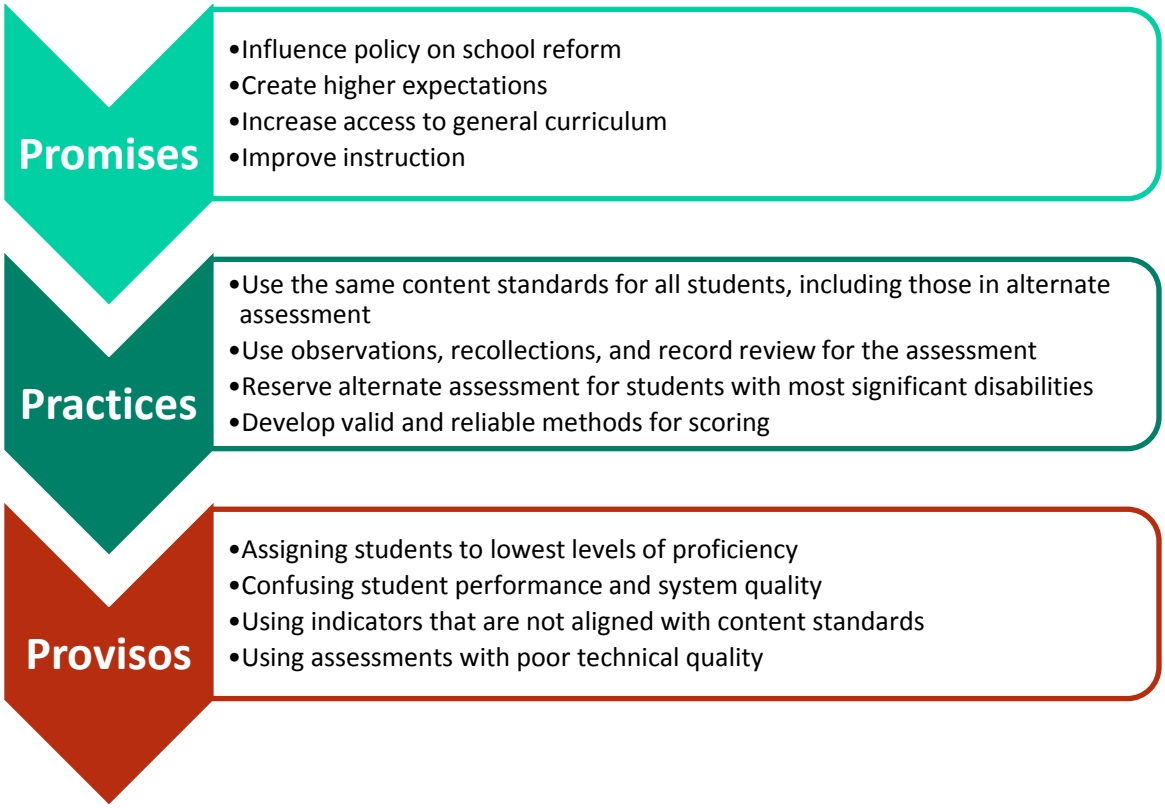


Source: Streagle and Scott, "The Alternate Assessment…Process," Op. cit., p. 1302.

Through examination of "19 data-based studies on alternate assessment,"[91] Browder et al. (2003) propose an early best practice model for understanding the value of alternate assessment, reproduced in Figure 3.5 below. While these authors, like many proponents of AA-AAS, express interest in the "promises" and "practices" of alternate assessment that seek to strengthen the educational opportunities for students with significant disabilities, they caution (in "provisos") that alternate assessment can send mixed messages about what a student is capable of achieving. In particular, "[t]he threat of states assigning alternate assessments to the lowest proficiency level when using them for school accountability provides a disincentive to promote student achievement […] and could discourage inclusion of students in their neighborhood schools."[92]

[91] Browder et al., "What We Know," Op. cit., p. 46.
[92] Ibid., p. 51.

Figure 3.5: Organizing Themes for Evaluating Alternate Assessment



Source: Browder et al., "What We Know," Op. cit., p. 50.

IEP Development

To mitigate this risk, alternate assessment must connect to a student's Individual Education Plan (IEP) "so students and parents can participate in setting the level of expectation."[93] Indeed, such collaboration is essential to the IEP development process. Effective collaboration provides a "sturdy foundation" of preparation and planning to allow all stakeholders to support the process and the student. Moody (2010-11) suggests the following strategies for parents to advocate for their child in an IEP meeting:

- Consider the significant role of facial expressions, body language, voice pitch, intonation, and timing in communicating information, particularly when making requests and suggestions.
- Use strengths as a foundation for planning, not just weaknesses and goals, to help establish a common ground and to open lines of communication.
- Ask questions to gain clarity around certain points and to better understand documentation.
- Invite advocates who can speak on behalf of your child's needs and experiences.[94]

[93] Towles-Reeves, Kleinert, and Muhomba, "Alternate Assessment…New," Op. cit., p. 239.
[94] Adapted from: Moody, A.K. "Empowering Families To Be Collaborative Participants in IEP Meetings." *Childhood Education*, 87:2 (Winter 2010/2011), p. 130. ProQuest.

Geltner and Leibforth (2008) similarly argue for the inclusion of school counselors who, like students and parents, have the opportunity to emphasizes the strengths of the student as well as highlight environmental strengths and promote strength development within the context of the school's resources and structures.[95] Towles-Reeves, Kleinert, and Muhomba (2009) similarly note that the "10 studies [that] characterize the status of the field […] suggest the need for experts in both significant cognitive disabilities and academic content domains to establish alignment criteria, as well as the necessary amount of alignment, for students participating in AA-AAS."[96]

Such expertise and system opportunities must in turn must be effectively communicated to all school personnel involved in IEP planning and alternate assessment delivery. Towles-Reeves and Kleinert (2006) find in a survey of teachers that many feel "that they had always been doing what was required by the alternate assessment, so they had not made any changes to their daily instruction or IEP development." This finding frustrates the authors, who note that it contributes to a sense "that the expenditures of these resources [for training and implementation] have resulted in little or no impact upon the instruction of their students."[97] Yet even researchers find it difficult to establish a clear relationship between IEP contents and alternate assessment scores: there are too many contextual variables at play.[98]

Thus, Towles-Reeves, Kleinert, and Anderman (2008) cautioned that, in addition to studying the technical quality of developed AA-AAS, researchers and practitioners need to explore the impacts of alternate assessment on special education. They write, "As the inclusion of all students in school accountability and the inclusion of students with significant cognitive disabilities in grade-level academic content are both relatively recent developments, more work must be done to understand the impact of the assessment (and school reform) on teachers' daily instruction."[99] Elsewhere, Towles-Reeves, Kleinert, and Muhomba (2009) refer to this as assessing the "consequential validity" of an assessment, in terms of "the consequences – both intended and unintended – of that assessment on student instruction and learning."[100]

[95] Geltner, J., and T. Leibforth. "Advocacy in the IEP Process: Strengths-Based School Counseling in Action." *Professional School Counseling*, 12:2 (Dec 2008). ProQuest.
[96] Towles-Reeves, Kleinert, and Muhomba, "Alternate Assessment…New," Op. cit., p. 238.
[97] Towles-Reeves, E., and H. Kleinert. "The Impact of One State's Alternate Assessment Upon Instruction and IEP Development." *Rural Special Education Quarterly*, 25:3 (Summer 2006), pp. 36-37. EBSCO.
[98] Karvonen, M., and H. Huynh. "Relationship between IEP Characteristics and Test Scores on an Alternate Assessment for Students with Significant Cognitive Disabilities." *Applied Measurement in Education*, 20:3 (2007). EBSCO.
[99] Towles-Reeves, E., H. Kleinert, and L. Anderman. "Alternate Assessments Based on Alternate Achievement Standards: Principals' Perceptions." *Research & Practice for Persons with Severe Disabilities*, 33:3 (2008), p. 124. EBSCO.
[100] Towles-Reeves, Kleinert, and Muhomba, "Alternate Assessment…New," Op. cit., p. 241.

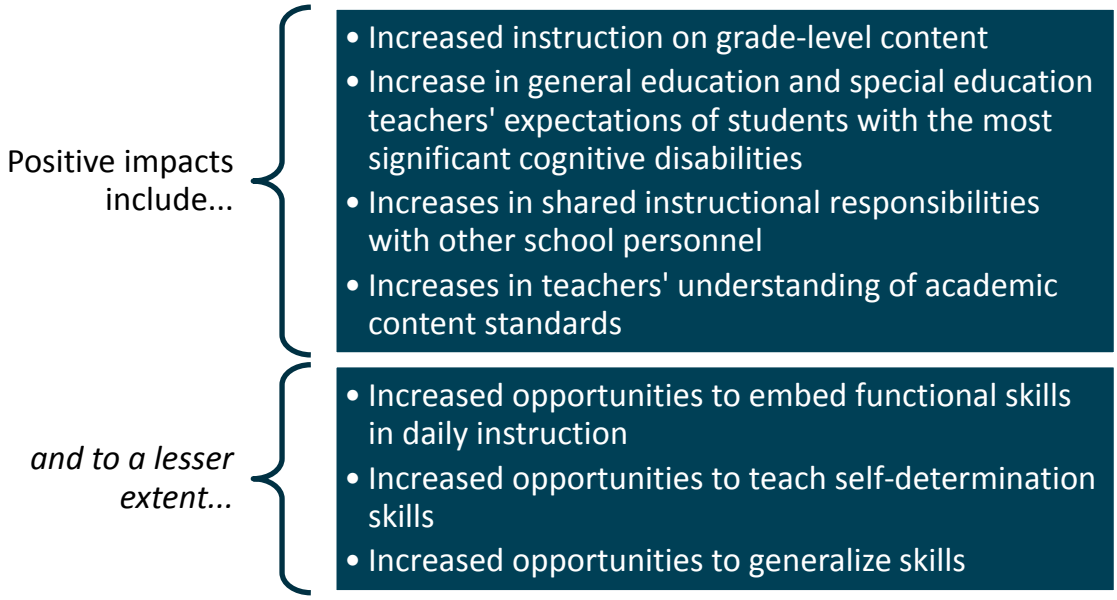
INSTRUCTIONAL CHANGES AND PROFESSIONAL DEVELOPMENT

Goldstein and Behuniak (2012) do explore the impact of alternate assessment on instructional practice in Connecticut, in a period governed by the Second Generation Connecticut Mastery Test/Connecticut Academic Performance Test (CMT/CAPT) Skills Checklist.[101] (Note some changes to current practice as the state has partially adopted NCSC/MSAA alternate assessment; see Figure 2.1.) They argue that "there was an expectation that AA-AAS would change the education system for students with severe disabilities" by "moving toward including more academic content in their performance indicators for these students." However, at least initially, functional skills were also central components, and "scoring criteria varied widely."[102] In particular, Goldstein and Behuniak's project addresses concerns over "appropriateness," given findings that "approximately half of the students have scores at or close to zero (*does not demonstrate skill*) in both Mathematics and Reading."[103] What they found was that a zero score "is more often used to indicate the teacher had not introduced the described skill rather than to indicate that the student does not demonstrate the described skill."[104]

Towles-Reeves, Kleinert, and Anderman similarly emphasize that "many teachers are finding it difficult to teach and to assess students on grade-level academic content standards." But principals also "appear to have struggled with understanding the intent of the AA-AAS as well as the overall importance of including these students in school accountability indices."[105] Nonetheless, principals participating in their survey identified several positive impacts on instructional practice, summarized in Figure 3.6 below.

[101] Goldstein, J., and P. Behuniak. "Can Assessment Drive Instruction? Understanding the Impact of One State's Alternate Assessment." *Research & Practice for Persons with Severe Disabilities*, 37:3 (2012). EBSCO.
[102] Goldstein and Behuniak, "Can Assessment Drive Instruction," Op. cit., p. 199.
[103] Ibid., p. 202.
[104] Ibid., p. 205.
[105] Towles-Reeves, Kleinert, and Anderman, "Alternate Assessments…Perceptions," Op. cit., p. 124.

Figure 3.6: Benefits of Shift to Alternate Assessment Systems



Source: Towles-Reeves, Kleinert, and Anderman, "Alternate Assessments…Perceptions," Op. cit., p. 129.

To resolve any persistent issues, Goldstein and Behuniak encourage clear "downward extensions" (written alternate standards) that can make instructional practice easier to manage. Specifically, the authors write, "downward extensions that were more likely to be introduced in an instructional setting were either described clearly with explicit direction for teachers or were representative of less cognitively demanding academic skills." [106] Additionally, they "urge continued professional development for educators on the scaffolding of academic content for this population as well as improved access to augmentative communication for students."[107]

For example, Kleinert, Browder, and Towles-Reeves (2005) summarize considerations related to the cognition element that link to specific "supports and strategies needed to compensate for these challenges" facing students with significant cognitive disabilities.[108] As illustrated in Figure 3.7, **their recommendations emphasize specific changes in teacher supports and teacher practice**. Among other approaches, teachers might need to use more "compensatory strategies" and emphasize "mnemonic strategies" with students to cope with this population's generally limited short-term memory. Meanwhile, teachers need support in preparing an appropriate instructional context that accounts for "ongoing instruction of specific target skills," "ongoing assessment" of progress toward those skills, and finding appropriate "social supports."[109]

[106] Goldstein and Behuniak, "Can Assessment Drive Instruction," Op. cit., p. 207.
[107] Ibid., p. 208.
[108] Kleinert, Browder, and Towles-Reeves, "The assessment triangle," Op. cit., p. 16.
[109] Ibid., pp. 14-16.

Figure 3.7: Implications for Alternate Assessment Based on Components of Cognition

Variable that Relates to Cognition	Characteristic of Students with Significant Cognitive Disabilities	Support and Opportunities Needed During Instruction and Alternate Assessment
Working or Short-Term Memory	Limitations in short-term memory	▪ Compensatory strategies (e.g., picture cues) ▪ Systematic instruction of mnemonic strategies & chunking with ongoing, frequent opportunities to use new learning
Long-Term Memory	Less clear that deficits are in long-term memory, but are rather affected by short-term memory challenges	▪ Need opportunities to continue using priority skills (e.g., in daily routines) ▪ Need opportunities to apply skills reflecting two types of long-term memory: the "way the world is" and "how things are done"
Metacognition	Students can use self-determination skills that require "thinking about thinking," but may need instruction to do so	▪ Opportunities and instruction to learn to problem-solve, self-evaluate, and self-correct ▪ Need to be planful about giving students opportunities to negotiate novel tasks and communicate about their learning
Practice and Feedback	While all learners need practice and feedback, this needs to be much more explicit and more frequent for students with significant cognitive disabilities	▪ Need many opportunities to practice tasks with feedback ▪ Need feedback that goes beyond accuracy (e.g., instructive feedback)
Transfer of Knowledge	While generalization cannot be assumed for any students unless they are provided with opportunities to experience the concept with multiple representations, transfer/ generalizability of concepts and skills cannot be assumed for students with significant cognitive disabilities unless explicitly taught and assessed	▪ Will need instruction in multiple contexts, materials, example to generalize ▪ Need opportunities to show both near and far transfer and to show degree of conceptual generalization, not just generalization across people and settings
Microgenetic Analysis	Need to consider how students learn on an intensive, ongoing trial-by-trial basis to understand the process of learning, as this process may be subtle and gradual	▪ Need for ongoing instruction of specific target skills with systematic instruction (prompting and feedback) ▪ Need for ongoing assessment of acquisition of target skills with data-based decisions about progress
Social and Situative Context of Learning	Learning is mediated by social and situational context	▪ Need to consider how social supports impact learning (e.g., opportunities to learn with typical peer in inclusive contexts) ▪ Need for opportunities to learn and apply skills in "real world" context

Source: Kleinert, Browder, and Towles-Reeves, "The assessment triangle," Op. cit., pp. 14-16.

In summary, Towles-Reeves, Kleinert, and Anderman emphasize the importance of the following three offerings:

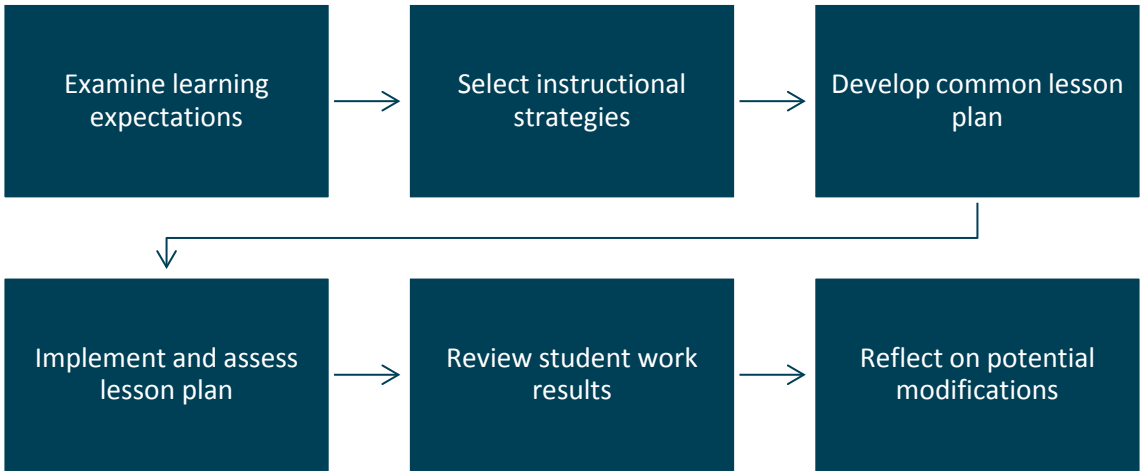
- **Professional development for principals** (including those in the field for longer periods) related to several topics:
  - the development of AA-AAS at the state level and approaches used for the state's AA-AAS;
  - the access to the general curriculum for students with the most significant cognitive disabilities; and
  - the instruction and assessment of these students on grade-level academic content standards paired with active instructional leadership practices.
- **Collaboration between general and special education teachers** to develop excellent adapted approaches and materials for students with the most significant cognitive disabilities
- **Inclusion of parental preferences and voices** in the development of the IEP[110]

PROFESSIONAL LEARNING COMMUNITIES AND THE SANDI

One way to support the rich collaboration recommended by research is the professional learning community (PLC) model of professional development. The SANDI assessment tool, for example, uses PLC as a central component of its assessment model to analyze student data. PLC describes a method of semi-structured conversation among educators involving "three important elements […]: focus on learning, collaborative culture, and results-oriented thinking."[111] Effective PLCs emphasize supportive and shared leadership, shared values and vision, collective learning, shared practice, and supportive conditions for the maintenance of the group.[112] Educators work through six key steps, illustrated in Figure 3.8.[113]

[110] Bulleted items loosely adapted from: Towles-Reeves, Kleinert, and Anderman, "Alternate Assessments…Perceptions," Op. cit., p. 132.
[111] Jessie, L.G. "The Elements of a Professional Learning Community." *Leadership Compass*, 5:2 (Winter 2007), p. 1. https://www.naesp.org/resources/2/Leadership_Compass/2007/LC2007v5n2a4.pdf
[112] Ferger, S. and Arruda, E. "Professional Learning Communities: Key Themes from the Literature." The Education Alliance, Brown University. Spring 2008. http://www.misalondon.ca/PDF/BIP/SupportMaterials/Professional_Learning_Communities.pdf
[113] Provini, C. "Best Practices for Professional Learning Communities." *Education World*, 2012. http://www.educationworld.com/a_admin/best-practices-for-professional-learning-communities.shtml

Figure 3.8: PLC Process



Source: Provini, "Best Practices," Op. cit.

The professional development model used with the SANDI is based on implementation research and embeds the principles that "implementation appears most successful when:

- carefully selected practitioners receive coordinated training, coaching, and frequent performance assessments;
- organizations provide the infrastructure necessary for timely training, skillful supervision
- and coaching, and regular process and outcome evaluations;
- communities and consumers are fully involved in the selection and evaluation of programs and practices; and
- state and federal funding avenues, policies, and regulations create a hospitable environment for implementation and program operations."[114]

Beginning in 2004-2005, district leadership teams, formed of representative teacher and administrative leaders from school sites, participated in professional development to build a shared, district wide vision of high expectations for the achievement of all students, especially those students with intellectual disabilities. Contractual agreements with the teachers' association (NEA/RCOTA) were negotiated, insuring that all teachers countywide would participate in PLC during the contractual school day. Teachers and administrators participated in professional development to build shared knowledge of the process of understanding the foundational pieces of summative and formative assessment, access to core content standards, collaboration, and data based decision making.

[114] Fixsen, D.L., et al. "Implementation Research: A Synthesis of the Literature." Tampa, Florida: University of South Florida, 2005, p. vi. http://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

To ensure consistent administration of the SANDI assessment, professional development modules are delivered by site-level leadership teams, and implementation is supported and monitored by district leadership. Each module may be customized depending on the site staff availability and needs. All modules are available online 24/7 for review.

The current professional development model consists of three modules. Part 1 includes an in-depth analysis of the CCSS and explicit direct instruction of the consistent implementation of SANDI online across school sites and districts, including connections between the SANDI and the IEP writing process. In Part 2, teachers and administrators learn methods of analyzing SANDI student assessment data reports to improve instruction and increase student achievement. Master schedules, structured learning environments, effective instructional groups, identifying standards and skills to be targeted, and lesson planning are the end goals of Part 2 Professional Development. The assessment further provides consistent data across a district, allowing teachers to collaborate and analyze the data in an inquiry process that informs instruction by identifying evidence based practices that align with student learning needs. Part 3 covers a range of teaching practice topics such as the PLC Data Team Process, writing SMART goals for a PLC, identifying proficiency levels and learner strengths in students, and related practical management and instruction.

Professional development modules have been developed and implemented through consistent and ongoing teacher input, teacher and administrator training, leadership team and administrative coaching, and feedback cycle. This has given teachers the skills and tools to interpret and use SANDI assessment information to write and implement standards based IEP goals, examine student data as an inquiry process, and inform classroom instruction including student grouping, and evidence based practices.

TECHNICAL DOCUMENTATION

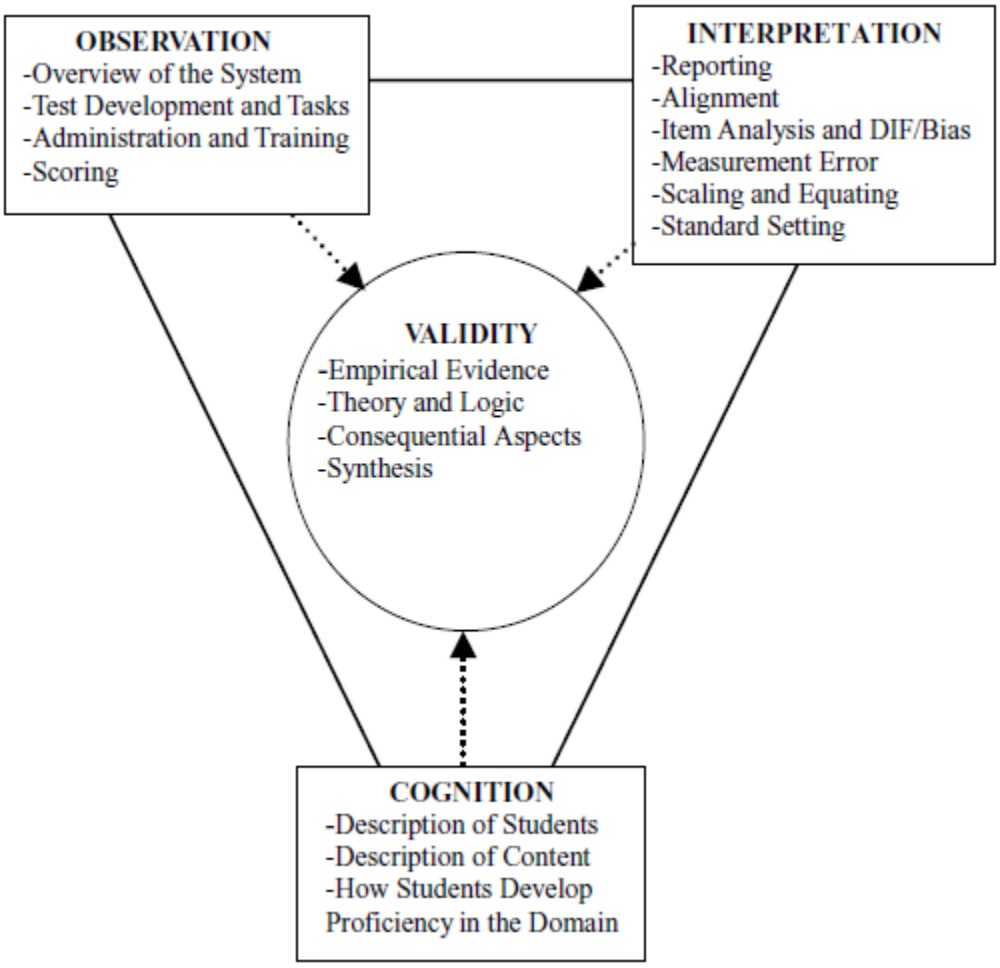
Beyond training and development, Marion and Pellegrino (2006) recommend states provide a series of technical documents to support implementation of an alternate assessment system, reproduced in Figure 3.9 below within their "assessment triangle" model of evaluating validity. They propose organizing these technical support materials "as a set of at least four documents," including:

- A somewhat familiar "nuts and bolts" volume that includes chapters that psychometricians are used to seeing in technical manuals;
- A validity evaluation of the sort that we have been discussing so far;
- A stakeholder summary that is drawn from the validity evaluation and the nuts and bolts volume; and
- A transition document that contains extensive procedural details to aid when programs transition from one contractor to another and/or there is turnover in state DOE personnel.[115]

[115] Bulleted items taken verbatim from: Marion and Pellegrino, "A Validity Framework," Op. cit., p. 50.

The authors "acknowledge that this approach to technical documentation is more labor and resource intensive than what has been done in the past." However, they continue, "it is crucial that there be a plan for systematic data collection and reporting."[116]

Figure 3.9: Technical Documentation Recommended for Alternate Assessment



Source: Marion and Pellegrino, "A Validity Framework," Op. cit., p. 52.

[116] Ibid., p. 54.

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds client expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

http://www.hanoverresearch.com/evaluation/index.php

CAVEAT

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties that extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every client. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Clients requiring such services are advised to consult an appropriate professional.

4401 Wilson Boulevard, Suite 400
Arlington, VA 22203
P 202.559.0500 **F** 866.808.6585
www.hanoverresearch.com