



**A Proposed Typology for
Characterizing States' Alternate
Assessments Based on Alternate
Achievement Standards:
Developing a Common Vocabulary to
Describe These Assessments**

*Mari Quenemoen
Rachel Quenemoen
Jacqui Kearns
Sarah Kennedy*

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: U.S. Department of Education Office of Special Education Programs Grant No. H3244040001. Opinions expressed in this document do not necessarily reflect those of these funders.

Executive Summary

States currently administer a variety of alternate assessments based on alternate achievement standards (AA-AAS) to students with the most significant disabilities, or approximately 1% of the student population. As the field of AA-AAS has changed rapidly since federal guidelines were set in 2003, there has been no shared language for describing and understanding the approaches that states use to test students with the most significant disabilities. The purpose of this report is to propose a typology of AA-AAS approaches using vocabulary that accurately describes test and item characteristics observed in states' AA-AAS. The suggested terms are derived from commonly-used definitions in general and special education assessment literature. Based on observations of current test approaches, this report suggests that in addition to test approach, we must be more specific about the item formats that item-based assessments contain. While this report touches on some scoring practices, a second report (Quenemoen, Perie, & Kearns, 2010) will evaluate in greater depth the interactions between assessment approaches and scoring methods.¹ The typology proposed in this report represents a starting point for building a shared vocabulary for stakeholders to describe and evaluate AA-AAS approaches.

The AA-AAS approach types proposed in this report are portfolios, rating scales, and item-based tests.

1. *Portfolios* are a collection of student work (including, but not limited to, worksheets, student-produced products, videos, pictures, or data sheets) that measure a limited number of benchmarks or objectives (usually two to six per content area). Tasks and activities are teacher-designed or modified. Portfolios may be *unstructured*, or relatively *structured* with fixed protocols for the collection of evidence.
2. On *rating scales*, teachers rate student performance on a relatively long pre-scripted list of skills based on classroom observation. Evidence may or may not be required.
3. *Item-based tests* consist of pre-scripted test items that students respond to in a one-on-one test administration setting. Items may include one or a combination of the following: *performance tasks*, *writing prompts*, *constructed-response items*, or *multiple-choice items*.

Twenty-one states currently use a portfolio for their AA-AAS, while 23 use an item-based test consisting of some combination of different item types. Five states use a rating scale approach, and one state uses both a portfolio and an item-based test.

¹ Information on scoring methods for each AA-AAS can be found in Appendices B and C.

Assessment Approach Types	Portfolio	Rating Scale	Item-based Test
Sub-types and item formats	Portfolio sub-types: <ul style="list-style-type: none"> - Structured portfolio - Unstructured portfolio 		Item formats: <ul style="list-style-type: none"> - Performance task - Writing prompt - Constructed response - Multiple choice

The typology represents an effort to develop a shared vocabulary, based on common usage in current assessment literature, for researchers, states, test vendors, policymakers, and practitioners to describe and evaluate AA-AAS approaches. Researchers will continue to test and validate this typology with special and general education assessment experts and policymakers to build consensus. The suggested typology of test approaches (portfolio, rating scale, and item-based) and item formats (performance task, writing prompt, constructed response, and multiple choice) should help clarify the often murky language used to describe AA-AAS for a variety of stakeholders:

1. Researchers who conduct surveys of state practices should consider using this typology as a framework for gathering information about AA-AAS. States would have the opportunity to choose descriptors that are defined consistently throughout the field to characterize their AA-AAS approaches, and researchers would have a shared basis for comparison and analysis.
2. States should consider using this typology to describe desired characteristics of AA-AAS in any request for bidders to develop or improve their AA-AAS.
3. Test vendors should carefully reflect on the terminology used in marketing or selling AA-AAS, and should consider conforming to the definitions in this typology or specifically identifying where they differ.
4. Policymakers and practitioners should consider using the typology as a frame for understanding how approaches differ and how the choice of approach may interact with guiding philosophy and beliefs about student characteristics and student learning.

Acknowledgements

The authors would like to thank Claudia Flowers, Harold Kleinert, Martha Thurlow, and Marianne Perie for their meticulous review of this report and their thoughtful comments. Leslie Moore, Ashleigh Gray, and Ashley Perkins provided research support. Thank you also to the many state staff who provided helpful feedback and additional information. Any remaining factual or interpretive errors are the responsibility of the authors only.

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Quenemoen, M., Quenemoen, R., Kearns, J., & Kennedy, S. (2010). *A Proposed Typology for Characterizing States' Alternate Assessments Based on Alternate Achievement Standards: Developing a Common Vocabulary to Describe These Assessments*. Lexington, KY: National Alternate Assessment Center.

Table of Contents

A Proposed Typology for Characterizing States' Alternate Assessments Based on Alternate Achievement Standards:	7
Introduction.....	7
Surveys/categorization procedures to date	7
Purpose	8
Methods.....	9
AA-AAS Approaches: A Proposed Typology	10
Test Approaches.....	12
Portfolios	12
Rating Scales	14
Item-based Tests.....	16
Performance Tasks	16
Writing Prompts	18
Constructed-Response Items	18
Multiple-Choice Items.....	19
Item Combinations	20
Gray Areas	21
Multiple Levels	22
Multiple Alternate Assessments.....	23
Recommendations.....	24
References.....	25

List of Tables

Table 1. Key features of each test approach	10
Table 2. States that use a portfolios	12
Table 3. States that use a rating scale	12
Table 4. States that use an item-based test.....	12
Table 5. Examples of structured portfolios.....	13
Table 6. Examples of unstructured portfolios.....	14
Table 7. “Number sense” performance thread, Indiana’s 3-5 grade mathematics AA-AAS	15
Table 8. Examples of rating scales	15
Table 9. Key features of item formats	16
Table 10. Example of a performance task	17
Table 11. Example of a writing prompt.....	18
Table 12. Examples of constructed-response items.....	19
Table 13. Examples of multiple-choice items.....	20
Table 14. Gray area: performance tasks vs. constructed-response items	22
Table 15. States that use multiple AA-AAS for students at varying levels within each grade band	23
Table 16. Proposed AA-AAS approaches typology	24

List of Figures

Figure 1. Number of states using each test approach.....	11
Figure 2. Item combinations on item-based tests.....	21

A Proposed Typology for Characterizing States' Alternate Assessments Based on Alternate Achievement Standards:

Developing a Common Vocabulary to Describe These Assessments

Mari Quenemoen, Rachel Quenemoen, Jacqui Kearns, Sarah Kennedy

Introduction

The Individuals with Disabilities Education Act Amendments (IDEA) of 1997 included the first federal requirement on alternate assessment, building on the inclusive assessment provisions of the Improving America's Schools Act (IASA) of 1994. In the years that followed, states were charged with developing a new form of large-scale assessment that would include students with the most significant cognitive disabilities in state assessment systems used for Title I reporting purposes. In 2003, federal regulations (U.S. Department of Education, 2003) further defined the requirements of alternate assessments based on alternate achievement standards (AA-AAS) for use in accountability testing required by the No Child Left Behind Act (NCLB) of 2001.

Initially, states tended to use one or a combination of assessment formats, including what one researcher described as portfolio, checklist, performance assessment, observation in structured and unstructured settings, or sample of student work (Roeber, 2002). Over the following years, other researchers found that assessments named "portfolio" or "checklist" or "performance assessment" varied greatly, and that many AA-AAS actually combined elements of more than one approach (Elliott & Roach, 2007; Gong & Marion, 2006; Quenemoen, 2009; Quenemoen, Thompson, & Thurlow, 2003).

Surveys/categorization procedures to date

The National Center on Educational Outcomes (NCEO) has published an annual or bi-annual survey of all 50 states' assessment practices for students with disabilities since 1991, and began reporting on states' alternate assessment activities in the 1997 survey (Erickson & Thurlow, 1997). Since 2001, the survey has used the categories "portfolio/body of evidence," "checklist," "IEP analysis," and later, "performance tasks" to survey states about their AA-AAS assessment approaches. Each report, however, cautioned that the variability across AA-AAS approaches was so great as to make these categories of questionable utility. The 2005 survey stated that, "It may be that the traditional way of describing alternate assessment approaches is no longer the best because there is considerable overlap across approaches that states take" (Thompson, Johnstone, Thurlow, & Altman, 2005, p. 11). The 2009 survey dropped the term "checklist" and added the term "multiple-choice test," stating that "practices varied widely and defied easy categorization" (Altman et al., 2010, p. 20). In an effort to be more specific about assessment practices, the 2005 and 2009 surveys added sub-descriptors for test approach choices such as "does / does not

require the submission of evidence” for performance tasks, or “does / does not include a set of structured performance tasks” for portfolios.

In 2008, researchers from SRI International and Policy Studies Associates compiled the National Profile on Alternate Assessments based on Alternate Achievement Standards (NSAA) based on data collected in the 2006-07 school year (Cameto et al., 2009). This study asked if states used one or more of four approaches: rating scale/checklist, portfolio/body of evidence, performance tasks/events, or multiple choice/constructed response. The NSAA used these categories without defining the characteristics of each approach.

Survey methodology requires that respondents understand the language on the questionnaire in the same way that the survey designer intends. Without explicit definitions or a systematic discussion of vocabulary, survey results may in fact misrepresent actual practices. An overview of naming practices in Appendix A shows inconsistencies across the two recent national surveys and online state materials. To describe South Dakota’s AA-AAS, NSAA used the terms “rating scale/checklist” and “portfolio/body of evidence,” while NCEO called it “performance tasks with evidence.” Most categories were similar across the NSAA and NCEO surveys, but not identical. For instance, the NSAA called Oklahoma’s AA-AAS a “portfolio/body of evidence,” and NCEO called it a “body of evidence/portfolio without standardized performance tasks.” This report’s main contribution is not to suggest the “right name” for each state’s test, but to break down and define the major characteristics of tests and test items in an effort to systematize the vocabulary the field uses to talk about those assessments.

Naming practices for AA-AAS are made more complicated by perceptions among states that some approaches are more likely to pass federal and public scrutiny than others. A state that believes that the term “multiple choice” will be perceived by stakeholders or policymakers as too limiting for these students, for instance, may choose instead to use the term “performance task.” A state that hesitates to call its assessment a “portfolio” due to perceived stakeholder criticism may instead call it a “body of evidence.” This report represents an effort to standardize the way states, researchers, and policymakers talk about test approaches and item formats in a way that illuminates, rather than obscures, actual practices.

Purpose

The purpose of this report is to propose a typology of AA-AAS approaches using vocabulary that accurately describes test and item characteristics observed in states’ AA-AAS. The suggested terms are derived from commonly-used definitions in general and special education assessment literature. Based on observations of current test approaches, this report suggests that in addition to test approach, we must be more specific about the item formats that item-based assessments contain. While this report touches on some scoring practices, a second report (Quenemoen, Perie, & Kearns, 2010) will evaluate in greater depth the interactions between assessment approaches

and scoring methods.² The typology proposed in this report represents a starting point for building a shared vocabulary for stakeholders to describe and evaluate AA-AAS approaches.

Methods

Two researchers from the National Alternate Assessment Center (NAAC), a federally funded research center, identified and reviewed states' online AA-AAS materials, including administration and technical manuals, training materials, and sample test items using both a manual search on states' Department of Education websites and keyword searches on Google. The researchers compared results, and, when inconsistent, re-analyzed the data until interrater agreement reached 100%.

They then verified observed characteristics of each assessment with state AA-AAS specialists via email or telephone using a descriptive narrative based on a set of draft definitions. For test approaches, states were asked to verify that their AA-AAS consisted of one of the following:

- (1) evidence of student work collected throughout the year;
- (2) ratings of student performance on a list of skills; or
- (3) scripted items to which the student must respond.

If the state used (3), scripted items, researchers further asked states to verify whether items could be characterized in one or more of the following ways:

- (a) the student engages in an activity that involves two or more steps, or the student answers a series of related questions;
- (b) the student answers an open-ended question that may have one or more correct answers; or
- (c) the student answers a question by choosing the correct answer from a number of possible choices.

As the project progressed, researchers modified item option (a), “the student engages in an activity that involves two or more steps, or the student answers a series of related questions,” which was meant to capture characteristics of performance tasks, clarifying that in order to be called a performance task, questions had to follow logically from one to the next and fit together in an activity. Unrelated questions on the same reading passage, for instance, could not be considered a performance task. Researchers also found that the use of concrete manipulatives to solve a problem within an activity indicated that the item was a performance task (as opposed to using manipulatives to communicate an answer choice, e.g., on multiple-choice items). These

² Information on scoring methods for each AA-AAS can be found in Appendices B and C.

descriptors were derived from definitions in literature on special and general education assessment.

Based on states' answers, researchers determined whether the state used a (1) portfolio; (2) rating scale; or (3) item-based test, and whether the items on an item-based test were one or more of a combination of (a) performance tasks; (b) constructed-response items; or (c) multiple-choice items. For states that provided clear descriptions of test approach in their online materials, or provided sample items, researchers wrote simple descriptive narratives using the appropriate elements above, and states then verified or modified those narratives. Researchers later added "writing prompts" as a fourth but less common item format designed to assess written expression. All but three states verified the descriptions or provided additional information.³ Data collected is current up to March, 2010.

The goal of gathering data from 50 states' AA-AAS practices was not to produce a comprehensive survey of practices. This has been done already (Altman et al., 2010; Cameto et al., 2009). The main purpose of this research was to produce a typology of test approaches and item formats based on current practices using standardized definitions. Especially in cases where no sample items were available for analysis, the approach or format listed for any given state does not authoritatively represent current practice. The process of assigning a type to each state was used only to determine whether the proposed typology could be applied to all current practices. The primary goal of this exercise was to develop the typology, which should help states, researchers, and policymakers talk about these assessments with more precise, shared language.

AA-AAS Approaches: A Proposed Typology

The AA-AAS approach types proposed in this report draw from common usage in general and special education assessment literature. This report distinguishes between *test approaches*, which include portfolios, rating scale assessments, and item-based assessments, and *item formats*, which include performance tasks, writing prompts, constructed-response items, and multiple-choice items. Some item-based assessments are comprised of only one type of item, and others are comprised of multiple item types. Table 1 describes the key features of each test approach.

Table 1. Key features of each test approach

<p>Portfolios</p> <ul style="list-style-type: none">- Collection of student work (including, but not limited to, worksheets, student-produced products, videos, pictures, or data sheets)- Measure a limited number of benchmarks or objectives (usually two to six per content area)- Tasks and activities are teacher-designed

³ The states that did not verify their profiles were Montana, Nebraska, and New Mexico.

Rating Scales

- Teachers rate student performance on a relatively long pre-scripted list of skills based on classroom observation
- Evidence may or may not be required

Item-based Tests

- Consist of pre-scripted test items that students respond to in a one-on-one test administration setting
- Items may include one or a combination of the following: performance tasks, writing prompts, constructed-response items, or multiple-choice items

According to this typology system, 21 states use a portfolio, 23 states use an item-based test, five states use a rating scale, and one uses both a portfolio and an item-based test. These numbers represent AA-AAS used to meet accountability requirements under the Elementary and Secondary Education Act (ESEA) in the fifty states.

Figure 1. Number of states using each test approach

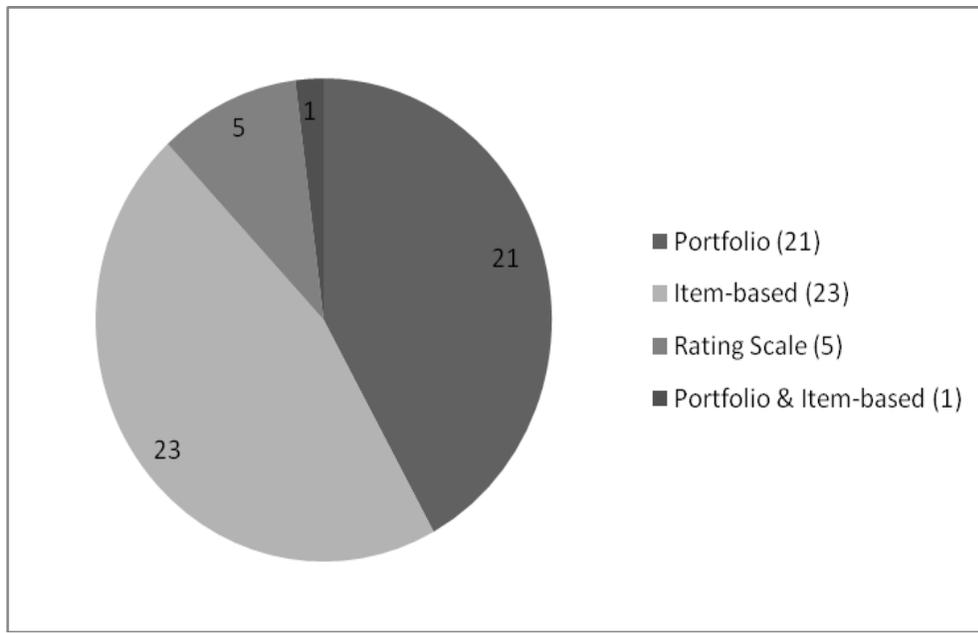


Table 2. States that use a portfolio

Alabama	Arkansas	Delaware
Georgia	Idaho	Kansas
Kentucky	Maryland	Massachusetts
Mississippi	Missouri	New Hampshire
New Jersey	New York	Ohio
Oklahoma	Rhode Island	Tennessee
Vermont	Virginia	Washington
Wyoming ⁴		

Table 3. States that use a rating scale

Connecticut	Hawaii ⁵	Indiana
Iowa	South Dakota	

Table 4. States that use an item-based test

Alaska	Arizona	California
Colorado	Florida	Illinois
Louisiana	Maine	Michigan
Minnesota	Montana	Nebraska
Nevada	New Mexico	North Carolina
North Dakota	Oregon	Pennsylvania
South Carolina	Texas	Utah
West Virginia	Wisconsin	Wyoming

Test Approaches

Portfolios

Most assessment textbooks state that a portfolio is a “purposeful collection of student work” that demonstrates student learning and achievement (McMillan, 2006; Mertler, 2003). AA-AAS portfolios are always a collection of evidence of student performance related to a relatively small number of learning objectives linked to state academic content or extended/expanded content standards. A portfolio usually addresses two to six “benchmarks” or “objectives” per content area. States provide broad guidance for what the student should be able to know and do for each benchmark or objective, and the teacher nearly always designs activities or tasks for the student to practice and perform. There are some exceptions to this rule: New York and Maryland give

⁴ Wyoming uses both a portfolio and an item-based test.

⁵ Hawaii will implement a new assessment in 2010-11.

teachers the option to select tasks from a task-bank, but also allow them to modify existing tasks or create new tasks.

Portfolios for the general population of students are usually student-driven, and are used to help students reflect on their own learning (Seitz & Bartholomew, 2008). For a general assessment portfolio, student-driven means that students often choose what they feel is their best or most representative work, and engage in self-evaluation. On an AA-AAS portfolio, students may or may not play an active role in selecting work, but elements on some scoring rubrics (such as participation, choice, interactions, and self-monitoring) award points for evidence that the student was somewhat engaged in the portfolio process. These practices do not always earn points, but are noted as best practice indicators; one state encourages students to personalize portfolios by adding personal artwork and producing written reflections on the portfolio, but does not award points for these elements toward the student's final achievement score.

Portfolios range in degree of structure, from loose requirements for the collection of evidence to a required series of structured task trials. Most portfolios allow teachers a significant degree of control over the design of tasks and activities that the student will perform to demonstrate skill, but many portfolios also provide a structured protocol for the collection of evidence. Of the 22⁶ states that use a portfolio approach, 16 provide guidelines for gathering evidence or assessing student performance at structured intervals. Some require evidence of performance at baseline, endpoint, and any number of points in between. Others require evidence of performance in fixed collection periods. Protocols for the collection of evidence over fixed intervals may allow for greater comparability of student achievement and progress across student portfolios. Six states do not require evidence to be collected at fixed intervals, leaving that choice up to each teacher, which makes them relatively unstructured. See Table 5 for examples of structured portfolios and Table 6 for examples of unstructured portfolios. See Appendix B for more detail on each state's portfolio. Other commonly used names for the portfolio approach are "datafolio" or "body of evidence."

Table 5. Examples of structured portfolios

Georgia: Teachers design tasks for the student to perform during two different collection periods, at least three weeks apart. Teachers collect primary evidence of the student engaging in the task, and secondary evidence of student achievement, along with an interview and observation form.

Kentucky: Teachers design tasks that the student must perform with less than 50% accuracy at baseline. Students can perform a task up to four times for each of the five assessment targets (for a student identified with symbolic communication/dimension A) or three assessment targets (for pre-symbolic communication/dimension B) in each content area.

⁶ This number includes Wyoming, which also uses an item-based assessment.

Maryland: Teachers select "mastery objectives" from a task bank, or submit teacher-designed tasks for approval, and collect evidence of student performance on each task. Students must perform with under 50% accuracy at baseline, and must perform with 80-100% accuracy for an objective to be scored "mastered."

Massachusetts: Teachers create tasks and fill out data charts to show student performance on skills related to selected standard strands on at least eight different dates, as well as two pieces of evidence for each skill. The student should perform the skill with below 80% accuracy and/or independence at baseline.

Oklahoma: Teachers design activities for each standard, and record accuracy and supports pre-instruction, mid-instruction, and post-instruction. Five pieces of evidence per core subject area are required.

Table 6. Examples of unstructured portfolios

Arkansas: Teachers generate tasks, which students perform throughout the year. Each entry includes a description of the teacher generated task and three pieces of evidence.

Ohio: Teachers create two or three tasks per standard application (a broad description of an activity), and collect evidence of student performance on each task.

Rating Scales

A rating scale is a less common approach to AA-AAS (10%, or $n=5$). This approach requires the test administrator, usually the teacher who knows the student best, to rate the student against a relatively long pre-scripted list of discrete skills. Teachers use a rating scale to indicate the degree to which a student exhibits a characteristic, or to rate the student's level of performance based on classroom observation throughout the year (McMillan, 2006; Mertler, 2003).

Many surveys of AA-AAS practices have used the term "checklist" or the combined category "rating scale/checklist" to describe a test approach, and Connecticut describes its assessment in its online materials as a "checklist." However, no state currently uses a checklist for its AA-AAS as the term is strictly defined in the assessment literature. A rating scale provides a scale which teachers use to rate a student's level of performance on each skill. By contrast, a checklist only allows teachers to check whether or not specific skills are present or have been observed, but not the degree to which the student has achieved the skill (Mertler, 2003; Nitko, 2006).

Indiana calls its assessment a "matrix assessment," and it is unique in its use of "performance threads." Rather than providing a mastery or proficiency scale of 1–4 on each discrete skill, for instance, Indiana's assessment breaks down each standard into very specific descriptions for

each level of performance. For instance, on the grades 3–5 math assessment, the standard “number sense” is delineated from least to most complex, as demonstrated in Table 7.

Table 7. “Number sense” performance thread, Indiana’s 3-5 grade mathematics AA-AAS⁷

Least complex
→ demonstrates awareness of the presence of objects
→ identifies more
→ uses numbers to compare
→ names and orders quantities
→ describes relationships between numbers and quantity
→ identifies numbers and quantity to 100
→ identifies numbers and quantity to 1000
→ compares numbers on a number line
→ compares parts and whole
Most complex

Teachers indicate where on each skill continuum the student lies, and evidence is not required. This approach is similar to other rating scales in that teachers rate student performance on a list of skills, but the performance thread approach is unique.

Of the five states that use rating scales, three also require the teacher to create structured tasks and submit evidence of student performance on selected skills, much like a portfolio system. Although the teacher rates observed student performance on a long list of skills, the teacher also picks a limited number of skills from each content area that the student practices over a number of structured trials. Another state requires teachers to keep evidence of performance as part of its monitoring system. Table 8 provides a general description of selected states’ rating scale approaches. See Appendix B for more details about each state.

Table 8. Examples of rating scales

Connecticut: Teachers rate students on downward extensions of each essence statement for each performance standard. Ratings are based on trained teacher interactions and observations throughout the year. Each skill is rated as “mastery/independent,” “developing/supported,” or “does not demonstrate.” Mastery indicates accuracy and independence on at least 80% of trials. Evidence is collected only for monitoring requirements.

⁷ Retrieved February 5, 2010 from <https://ican.doe.state.in.us/beta/tm.htm>.

South Dakota: All skills on the rating form are rated. Additionally, teachers select one item from each indicator in reading and each content strand in math and science, and collect evidence of student performance on those skills, documenting at least three trials of each skill. A five point rating scale combines accuracy and level of assistance for each skill.

Iowa: Teachers instruct students on selected items from a list of pre-scripted skills. For the assessment, the student performs four trials of selected skills, and the teacher submits evidence of student performance on those trials. The teacher scores for percent accuracy on the most recent trial for each skill, and marks which skills were “already mastered,” “not taught,” or “fully prompted.” At least one skill per content area must have been taught or the student will count as an “exclusion.” Teachers are encouraged to teach as many skills as possible.

Item-based Tests

While portfolios and rating scales can be described as assessment approaches, other tests are best described as comprised of items. The four major types of items found on AA-AAS are performance tasks, writing prompts, constructed-response items, and multiple-choice items (see Table 9).

Table 9. Key features of item formats

Performance Tasks

- Both process- and outcome-oriented
- May contain a series of related questions/activities
- May involve solving problems using manipulatives

Writing Prompts

- Students produce a written response to a prompt
- Intended to assess written expression

Constructed Response

- Students produce an independent response to a question or prompt

Multiple Choice

- Students select the correct answer from two or more possible answer choices, which could include a combination of pictures, words, or tactile objects

Performance Tasks

Many states use the term “performance task” to describe a test with scripted items, but the items on these tests actually vary widely in their structure and how they require a student to respond. Performance tasks have one or more of the following characteristics: (a) They are both process-

and outcome-oriented; (b) They consist of a series of related questions or activities; or (c) They require students to solve a problem using manipulatives.

When completing a performance task, a student produces an “extended written or spoken answer, by engaging in group or individual activities or by creating a specific product” (Nitko, 2006, p. 242). In NAEP science assessments, performance tasks involve students solving a problem using concrete objects, and scores are based on not only the solution but on the process of arriving at the solution (Winick, Avallone, Smith, & Crovo, 2009). See Table 10 for an example of an AA-AAS science performance task.

Table 10. Example of a performance task

<p><u>California Level IV Science Sample Item</u>⁸</p> <p>Manipulatives: 1 flashlight 1 small ball</p> <p>Task Preparation: Place on surface flashlight and small ball. Tell student that these objects will be used as a model for the Sun and Earth. Turn on flashlight and position it approximately 2 feet from ball so that light is easily seen striking one side of ball. Give first cue. While pointing to flashlight, give second cue. While pointing directly down at top of the ball, give third cue.</p> <p>Cue/Direction: <i>Which object represents the Sun?</i> <i>How is the flashlight like the Sun?</i> <i>Which is the warm side of Earth?</i></p> <p>Correct answers: flashlight represents the Sun; gives off light; lit side of ball.</p>
--

As the definition suggests, a performance task emphasizes the process as well as the outcome, and requires the student to engage in an activity. For this report, if a state answered that the items on its test require a student to “engage in an activity that involves two or more steps” or “answer a series of related questions,” researchers classified the items as performance task items. For a series of related questions to be categorized as a performance task, the questions should fit together in a meaningful way in the context of an activity. A series of unrelated questions that require open-ended responses on the same reading passage, for example, would be categorized as constructed-response items. Questions on a performance task should follow logically from one to the next.

⁸ Example retrieved August 12, 2010 from: <http://www.startest.org/pdfs/CAPA.task-science.level-iv.2008.pdf>

Writing Prompts

At least four states' AA-AAS include one or more writing prompts that the student must respond to with a written narrative. Writing prompts are similar to, but distinct from, performance tasks.⁹ These items are also process-oriented, but they are aimed at assessing the student's performance of written expression. The use of the word "prompt" here is consistent with that term in general assessment; it does not mean level of support, such as a "system of prompts," common in special education terminology. See Table 11 for an example of an AA-AAS writing prompt.

Table 11. Example of a writing prompt

North Carolina Grade 10 Writing Sample Item¹⁰

Stacey and Kelly are in the tenth grade. They are in the same class. Their class is having a fundraiser. They want to raise money for new school supplies. The class is selling candy. Stacey sells a lot of candy. Stacey earns money for the class fundraiser. Kelly does not sell any candy. Kelly does not earn any money for the class fundraiser.

SAY: "Write at least four facts about Stacey and Kelly's class fundraiser."

SAY: "Make sure that you write complete sentences. Remember to use capital letters and end punctuation marks correctly."

SAY: "Begin writing your sentences now."

Constructed-Response Items

Constructed-response items are not process-oriented activities, but require students to produce an independent answer to a prompt or a question. These items can take the form of simple question/response formats, cloze items, or simple mathematical calculations. Constructed-response questions are open-ended and may allow for several ways to arrive at a correct answer (Mertler, 2003). While multiple-choice questions provide a set of predetermined possible answers, constructed-response questions require students to create their own answer or response (McMillan, 2006). In the absence of sample items, researchers classified test items as constructed response if states answered that their test items require a student to "answer an open-ended question that may have one or more correct answers." See Table 12 for examples of constructed-response items.

¹⁰ *NCEXTEND1* Writing Grade 10 Sample Prompt, retrieved August 12, 2010 from: <http://dpi.state.nc.us/docs/accountability/testing/ncextend1/ncextend1writinggrade10samplesjanuary2009.pdf>

Table 12. Examples of constructed-response items

Wyoming 3rd Grade Reading Sample Item¹¹

SAY: We are going to read a story about “Joe’s Day at School” and then I will ask you some questions about the story. You can follow along as I read.

DO: Present the student with the story “Joe’s Day at School”

Joe is in 4th grade. Every morning his mom gives him a ride to school. At school, Joe sits next to his friend Billy. Joe’s favorite subject is P.E. He likes to play basketball in P.E. class with Billy. Joe has P.E. class on Monday, Wednesday, and Friday. Joe has reading and writing every day, and music class on Thursday’s. After school he rides the bus home. He does his homework and gets ready for another day.

SAY: Who was this story mostly about?

Pennsylvania 11th Grade Reading Sample Item¹²

The sun wasn’t up yet when Mr. Santana started work. He poured the muffin batter into pans before sliding the pans into the oven. He wiped the white powder off of the counter after rolling the bread dough. He sprinkled sugar onto to the chocolate cookies inside the glass case.

Keep the paragraph from the previous skill in front of the student.

Say: What time of day did Mr. Santana start working?

Response: “before dawn”

Alternate responses: “very early morning”; “morning”

Multiple-Choice Items

Few states define their AA-AAS as a multiple-choice test, but in fact many do use multiple-choice items. Multiple-choice items require a student to choose the correct answer from two or more given answer choices. These answer choices may be textual or pictorial, and in some cases they may be a set of concrete manipulatives. Rather than engage in a process-oriented task or produce an independent response, students choose from a fixed set of possible responses. In general assessment language, multiple-choice questions use a “stem” question or incomplete phrase, with a set of answer options that the student can choose from (Mertler, 2003). True/false and yes/no questions can be considered a form of multiple-choice item. Some states call this item format “selected response.” See Table 13 for examples of multiple-choice items.

¹¹ Example sent as attachment through e-mail communication with Charlene Turner, February 25, 2010

¹² Example sent as attachment through e-mail communication with Lynda Balmer-Lupp, February 17, 2010.

Table 13. Examples of multiple-choice items

Illinois, Grade 5, Mathematics Sample Item¹³

Present the primary task to the student. Say: “Here is a pencil, a paper clip, and a ruler.” (The ruler should be 12 inches long.) “Which one is the shortest?”

Point to each option and say: “The pencil,” “the paper clip,” “the ruler.”

Minnesota, Grade 8, Reading Sample Item¹⁴

Say: We are going to read a story about Franklin D. Roosevelt.

Present “The Man on the Dime” or have the student read the passage.

Say: What job did Franklin D. Roosevelt do?

Present the answer options in order. *Point to each option as you say it.*

- A. He was a doctor.
- B. He was a president.
- C. He was a scientist.

Item Combinations

Eleven AA-AAS use only one type of item, and 14 tests use a combination of different item types.¹⁵ Out of 25 item-based tests, 18 use multiple-choice items, either alone or in combination with other items. See Figure 2 for number of item-based AA-AAS using each item combination.

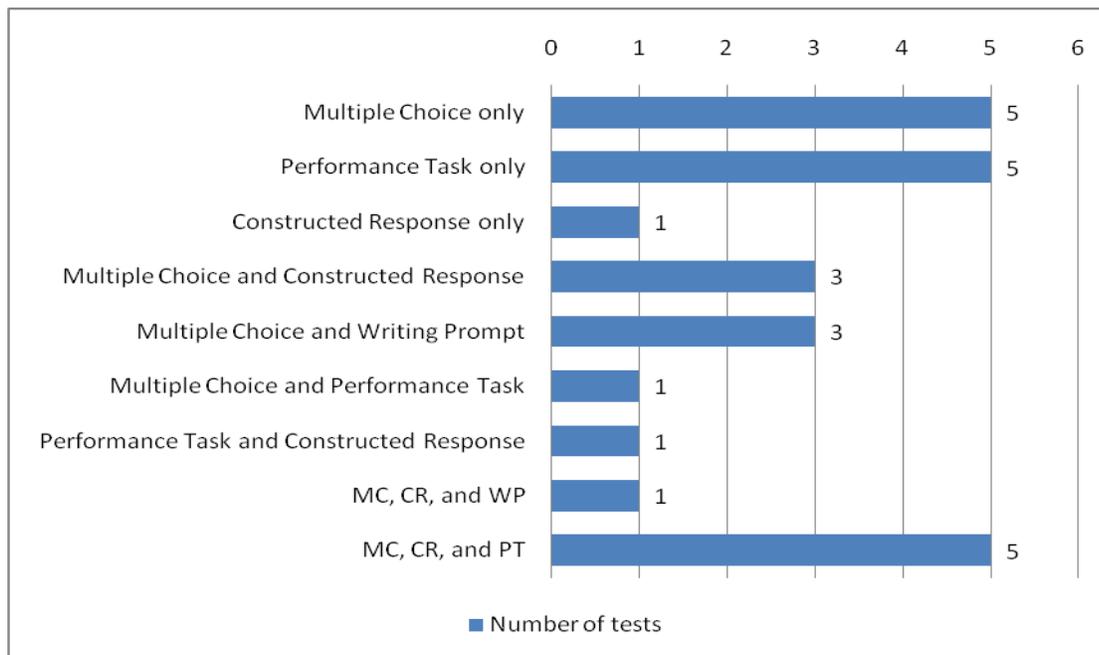
¹³ Example retrieved August 12, 2010 from: http://www.isbe.net/assessment/pdfs/2010/iaa/Math_gr5.pdf

¹⁴ Example retrieved August 12, 2010 from:

http://www.education.state.mn.us/MDE/Accountability_Programs/Assessment_and_Testing/Assessments/Alternate/Alternate_Item_Samplers/index.html

¹⁵ Michigan’s highest level test uses different item types from its two lower level tests, and they are represented as two different item combinations on this table.

Figure 2. Item combinations on item-based tests



Gray Areas

Assessment approaches still exhibit some degree of variability within each typology. New York and Maryland’s portfolio assessments allow teachers to choose activities from an item bank, modify existing activities, or create their own. In Maryland, teacher-created tasks must be reviewed and approved before administering. They still share several common characteristics with other portfolio assessments, including a high degree of teacher control over test design on a limited number of objectives within each content area. Maine characterizes its assessment as a portfolio, but teachers do not create their own tasks; instead, they always select tasks from a task bank. Each “task” resembles a worksheet with performance tasks, constructed-response items, or multiple-choice items. Thus, this report calls Maine’s AA-AAS an item-based assessment.

The line between constructed-response items and performance tasks can be blurry too. For instance, according to its 2009 test administration manual, Arizona’s assessment is comprised of what it calls multiple-choice items, performance tasks, and rater items. Based on examples in the administration manual, the multiple-choice items clearly conform to the definition laid out in this report. The sample “performance tasks” are a series of questions related to a common story which require simple and one-dimensional student-generated answers, resembling constructed-response items. Several examples of “rater items” look more like performance tasks in their activity-based format and use of manipulatives (see Table 14, rater item example 1). On the other hand, another “rater item” looks more like a constructed-response item (see Table 14, rater item example 2), requiring the student to produce a response that is not process-oriented and does not involve manipulating concrete items. This report characterizes Arizona’s test as comprised of

performance tasks, constructed-response items, and multiple-choice items, but it is not clear how the performance tasks and constructed-response items are distributed between the two sections.

Table 14. Gray area: performance tasks vs. constructed-response items

<p>Arizona Item Examples from 2009 Test Administration Directions</p> <p><u>“Performance Task” examples:</u></p> <p>Read the book <i>The Zoo</i> to the students.</p> <p>Question: Ask the student to describe the setting for the story.</p> <p>Correct Answer: The student responds that the story takes place in a zoo without assistance or with a single repetition of instruction or redirection.</p> <p>Question: Ask the student to identify the main character in the story.</p> <p>Correct Answer: The student responds “Tim” or “the Tiger” without assistance or with a single repetition of instruction or redirection.</p> <p><u>“Rater Item” examples:</u></p> <p>Example 1: When presented with the number 3, the student picks up or points to three objects or manipulatives.</p> <p>Example 2: When presented with a problem, student will select correct operation (i.e., fill in the blank: $4 \div 1 = 5$).</p>
--

Multiple Levels

Four states use multiple AA-AAS for students at varying “levels” of symbolic communication or perceived cognitive functioning. As indicated in Table 15, different test levels within the same assessment may be comprised of different items, or items in different proportions. Some states’ item-based assessments¹⁶ use one test for students in each grade-band, but allow teachers to choose each item from multiple levels of complexity. Other states’ portfolios¹⁷ provide different scoring rubrics for different students, depending on level of symbolic communication or educational setting. Scoring details and rubric dimensions for each state are included in Appendices B and C, and will be addressed in a separate report (Quenemoen et al., 2010).

¹⁶ Florida, Texas, and New Mexico

¹⁷ Mississippi and Tennessee

Table 15. States that use multiple AA-AAS for students at varying levels within each grade band

CA	Students take test II – V based on grade-level, but students “with the most significant disabilities” take test I.
KY	Dimension A assessment targets are for symbolic language users, and Dimension B assessment targets are for pre- and emerging-symbolic language users.
MI	MI uses three test levels: the "Participation" test is for students with severe cognitive impairments; the "Supported Independence" test is for students with moderate cognitive impairments; the "Functional Independence" test is for students with mild cognitive impairments.
PA	The AA-AAS is divided into three levels of complexity within each grade band: Level A, B, and C.

Multiple Alternate Assessments

Wyoming is the only state that requires all students to submit two test formats for the AA-AAS: a portfolio and an item-based test. Kentucky and Montana use different “alternate assessments” for different purposes, such as state-level accountability mandated under IDEA but not part of ESEA accountability requirements. In Kentucky, students in the general population are all required to take the ACT before leaving high school, while students with significant cognitive disabilities take another test called the “Transition Attainment Record,” used to track post-school outcomes over time. In addition, students with significant cognitive disabilities in Kentucky take an alternate “Attainment Task” test in lieu of statewide assessments in social studies and writing. Montana provides an alternate test for the Iowa Test of Basic Skills for students with significant cognitive disabilities. This report does not address alternate assessments other than the AA-AAS, or for purposes other than federal accountability under ESEA, but it is important to be aware that some states have developed “alternates” for other purposes.

Recommendations

In summary, this report proposes a typology consisting of three AA-AAS approaches, with sub-types of two approaches, as shown in Table 16.

Table 16. Proposed AA-AAS approaches typology

Assessment Approach Types	Portfolio	Rating Scale	Item-based Test
Sub-types and item formats	Portfolio sub-types: <ul style="list-style-type: none"> - Structured portfolio - Unstructured portfolio 		Item formats: <ul style="list-style-type: none"> - Performance task - Writing prompt - Constructed response - Multiple choice

The typology represents an effort to develop a shared vocabulary, based on common usage in current assessment literature, for researchers, states, test vendors, policymakers, and practitioners to describe and evaluate AA-AAS approaches. Researchers will continue to test and validate this typology with special and general education assessment experts and policymakers to build consensus. The suggested typology of test approaches (portfolio, rating scale, and item-based) and item formats (performance task, writing prompt, constructed response, and multiple choice) should help clarify the often murky language used to describe AA-AAS for a variety of stakeholders:

1. Researchers who conduct surveys of state practices should consider using this typology as a framework for gathering information about AA-AAS. States would have the opportunity to choose descriptors that are defined consistently throughout the field to characterize their AA-AAS approaches, and researchers would have a shared basis for comparison and analysis.
2. States should consider using this typology to describe desired characteristics of AA-AAS in any request for bidders to develop or improve their AA-AAS.

3. Test vendors should carefully reflect on the terminology used in marketing or selling AA-AAS, and should consider conforming to the definitions in this typology or specifically identifying where they differ.
4. Policymakers and practitioners should consider using the typology as a frame for understanding how approaches differ and how the choice of approach may interact with guiding philosophy and beliefs about student characteristics and student learning.

References

- Altman, J., Lazarus, S. S., Quenemoen, R. F., Kearns, J., Quenemoen, M. & Thurlow, M. L. (2010). *2009 Survey of the states: Accomplishments and new issues at the end of a decade of change*. Minneapolis MN: University of Minnesota, National Center on Educational Outcomes.
- Cameto, R., Knokey, A.-M., Nagle, K., Sanford, C., Blackorby, J., Sinclair, B., & Riley, D. (2009). *National profile on alternate assessments based on alternate achievement standards. A report from the national study on alternate assessments* (NCSE 2009-3014). Menlo Park, CA: SRI International.
- Elliott, S. N., & Roach, A. T. (2007). Alternate assessments of students with significant disabilities: Alternative approaches, common technical challenges. *Applied Measurement in Education*, 20(3), 301-333.
- Erickson, R., & Thurlow, M. (1997). 1997 State special education outcomes: A report on state activities during educational reform. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Gong, B., & Marion, S. (2006). *Dealing with flexibility in assessments for students with significant cognitive disabilities* (Synthesis Report 60). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Improving America's Schools Act of 1994, PL 103-382, 20 U.S.C. §§ 630 *et seq.*
- Individuals with Disabilities Education Act Amendments (IDEA) of 1997, PL 105-17, 20 U.S.C. §§ 1400 *et seq.*
- McMillan, J. H. (2006). *Classroom assessment: Principles and practice for effective standards-based instruction* (4th Edition). Boston, MA: Allyn & Bacon.
- Mertler, C. A. (2003). *Classroom assessment: A practical guide for educators*. Glendale, CA: Pycszak Publishing.
- Nitko, A. J. (2006). *Educational assessment of students* (5th Edition). Upper Saddle River, NJ: Prentice Hall.
- No Child Left Behind Act of 2001, PL 107-110, 115 Stat. 1425, 20 U.S.C. §§ 6301 *et seq.*
- Quenemoen, M., Perie, M., & Kearns, J. (2010). *Scoring alternate assessments based on alternate achievement standards: A proposed typology of AA-AAS scoring practices*. Lexington, KY: National Alternate Assessment Center.
- Quenemoen, R. F. (2009). The long and winding road of alternate assessments. In W. Schafer & R. Lissitz (Eds.), *Alternate assessments based on alternate achievement standards: Policy, practice, and potential* (pp. 127-156). Baltimore: Paul Brookes.
- Quenemoen, R., Thompson, S., & Thurlow, M. (2003). *Measuring academic achievement of*

- students with significant cognitive disabilities: Building understanding of alternate assessment scoring criteria* (Synthesis Report No. 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report No. 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Seitz, H., & Bartholomew, C. (2008). Powerful portfolios for young children. *Journal of Early Childhood Education*, 36, 63-68.
- Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). *2005 State special education outcomes: Steps forward in a decade of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thompson, S., & Thurlow, M. (2003). *2003 State special education outcomes: Marching on*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2003, December). *Title I—Improving the academic achievement of the disadvantaged, final regulations*. Washington, DC: Author.
- Winick, D. M., Avallone, A. P., Smith, C. E., & Crovo, M. (2009). *Science framework for the 2009 National Assessment of Educational Progress*. Retrieved from <http://nagb.org/search.aspx?q=NAEP%20Science>.

APPENDIX A: A comparison of test naming/categorization

State	State materials	2009 NSAA Report	2009 NCEO Survey	This report
AL	portfolio	portfolio/body of evidence	portfolio/body of evidence with standardized performance tasks	unstructured portfolio
AK	performance task	performance task/events	performance tasks with evidence	item-based (performance tasks, constructed response, and multiple choice)
AZ	multiple choice, performance task, rater items	rating scale/checklist, performance task/events, and multiple choice/constructed response	performance tasks with evidence and multiple choice	item-based (performance tasks, constructed response, and multiple choice)
AR	portfolio	portfolio/body of evidence	portfolio/body of evidence with standardized performance tasks	unstructured portfolio
CA	performance assessment	performance task/events	performance tasks without evidence	performance tasks
CO	standards-based assessment	performance task/events and multiple choice/constructed response	multiple choice	item-based (constructed response and multiple choice)
CT	rating scale	rating scale/checklist	other (checklist)	rating scale
DE	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio

FL ¹⁸	performance-based assessment	rating scale/checklist	performance tasks with evidence and multiple choice	item-based (multiple choice)
GA	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
HI ¹⁹	[no indication]	rating scale/checklist and portfolio/body of evidence	developing/revising	rating scale
ID	portfolio	rating scale/checklist and portfolio/body of evidence	developing/revising	unstructured portfolio
IL ²⁰	performance-based assessment	portfolio/body of evidence	multiple choice	item-based (multiple choice)
IN	matrix assessment	rating scale/checklist	performance tasks without evidence	rating scale
IA	rating scales	rating scale/checklist and portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	rating scale
KS	datafolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
KY	portfolio	rating scale/checklist, portfolio/body of evidence, and performance task/events	portfolio/body of evidence with standardized performance tasks	structured portfolio

¹⁸ Florida implemented a new test format in 2007-08.

¹⁹ Hawaii will implement a new test format in 2010-11.

²⁰ Illinois implemented a new test format in 2007-08.

LA	performance tasks	performance task/events	performance tasks without evidence	item-based (performance tasks and multiple choice)
ME ²¹	portfolio	portfolio/body of evidence and multiple choice/constructed response	performance tasks with evidence	item-based (performance tasks, constructed response, and multiple choice)
MD	portfolio	portfolio/body of evidence and performance task/events	portfolio/body of evidence with standardized performance tasks	structured portfolio
MA	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
MI	activity-based observation, selected response (also "accessing print" and "expressing ideas")	performance task/events and multiple choice/constructed response	performance tasks without evidence and multiple choice	item-based (performance tasks, writing prompt, constructed response, and multiple choice) ²²
MN	performance tasks	performance task/events	performance tasks without evidence	item-based (multiple choice)
MS ²³	portfolio	rating scale/checklist and portfolio/body of evidence	developing/revising	structured portfolio
MO	portfolio-based assessment	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio

²¹ Maine implemented a Task Bank in 2007-08, and eliminated teacher-designed tasks.

²² Michigan's highest level test is a combination of multiple choice and writing prompts, and its two lower level tests are a combination of performance tasks and constructed response.

²³ Mississippi began using a new test format in 2009-10.

MT	evidence-based test, consists of "tasklets"	performance task/events	performance tasks with evidence	item-based (performance tasks, constructed response, and multiple choice)
NE ²⁴	selected response items	performance task/events	developing/revising	item-based (multiple choice)
NV	performance-based assessment	performance task/events	developing/revising	item-based (multiple choice and writing prompt)
NH	portfolio	portfolio/body of evidence	developing/revising	structured portfolio
NJ	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
NM	task - collection of items and materials organized around a theme	performance task/events	performance tasks with evidence	item-based (performance tasks)
NY	datafolio	portfolio/body of evidence	portfolio/body of evidence with standardized performance tasks	structured portfolio
NC	performance items	performance task/events	performance tasks without evidence	item-based (multiple choice and writing prompt)
ND	activities	rating scale/checklist and performance task/events	performance tasks without evidence	item-based (constructed response and multiple choice)
OH	collection of evidence	portfolio/body of evidence	portfolio/body of evidence with standardized performance tasks	unstructured portfolio

²⁴ Nebraska began field-testing new assessments in 2009-10.

OK	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
OR	[no indication]	performance task/events	performance tasks without evidence	item-based (performance tasks)
PA	[no indication]	performance task/events and multiple choice/constructed response	other (performance tasks with multiple choice and constructed response)	item-based (performance tasks, constructed response, and multiple choice)
RI	datafolio with structured performance tasks	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
SC	performance tasks	performance task/events	performance tasks without evidence	item-based (multiple choice)
SD	rating form and supporting evidence	rating scale/checklist and portfolio/body of evidence	performance tasks with evidence	rating scale
TN	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
TX	assessment tasks	portfolio/body of evidence and performance task/events	performance tasks with evidence	item-based (performance tasks)
UT	assessment tasks	performance task/events	developing/revising	item-based (performance tasks)
VT	portfolio	portfolio/body of evidence	portfolio/body of evidence with standardized performance tasks	structured portfolio

VA	collection of evidence	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	unstructured portfolio
WA	portfolio	portfolio/body of evidence	portfolio/body of evidence without standardized performance tasks	structured portfolio
WV	performance tasks	multiple choice/constructed response	multiple choice	item-based (multiple choice and constructed response)
WI ²⁵	multiple choice, short answer, writing prompt	rating scale/checklist and portfolio/body of evidence	multiple choice	item-based (multiple choice, constructed response, and writing prompt)
WY	portfolio, performance event	rating scale/checklist, portfolio/body of evidence, and performance task/events	portfolio/body of evidence with standardized performance tasks	structured portfolio and item-based (constructed response)

²⁵ Wisconsin implemented a new test format in 2007-08.

APPENDIX B: Portfolios²⁶

1. Unstructured Portfolios

State	Test Details	Scoring Details	Scoring Elements															
			Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice	
AL	Three pieces of evidence per content standard collected during a three week window in March. For each extended standard, there are four levels of complexity with a broad description of how to assess the standard.	Alignment to the extended content standard (non-alignment makes the evidence unscorable), complexity (4), level of assistance (3), and mastery of content (3).	3	3		pre-req	4											
AR	Teachers generate tasks, which students perform throughout the year. Each entry includes a description of the teacher generated task and three pieces of evidence.	Each dimension can be scored up to 4 points, but is weighted as follows: performance x 4, context x 2, level of assistance x 1. Settings is scored only once for each content area.	4 (x4)	4 (x1)					4 (x2)		4 (scored once)							
ID	Teachers collect two pieces of evidence for each extended content objective, plus a data chart. For each extended standard, there are broad descriptions at multiple levels of complexity.	Accuracy (4), level of independence (4), and complexity/alignment (4).	4	4			4											

²⁶ Numbers in each cell represent the number of points possible for each rubric element per scored objective/strand, mirroring the scoring details text for each portfolio.

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
OH	Teachers create two or three tasks per standard application (a broad description of an activity), and collect evidence of student performance on each task.	Instructional context (4) and performance/accuracy (3). The student achievement score is calculated by multiplying performance by instructional context. The rubric also measures level of independence (4), and settings and interactions (4), but these do not count toward the achievement score.	3						4								
VA	Teachers select evidence of student achievement on stem/bullet statements (broad descriptions of task/goals). Teachers may choose an alternate standard from a higher or lower grade, if appropriate.	Performance on each standard is scored 0 – 4 points according to evidence of student skill and knowledge.															
WY	Teachers design tasks aligned to standard benchmarks, and collect two pieces of evidence for each. WY also requires an item-based test.	Performance (4), level of independence (4), and generalization across contexts (4). Level of complexity is scored 1 - 4, but is then weighted (1 = 1, 2 = 3, 3 = 5, and 4 = 8).	4	4			4 (weighted up to 8)			4							

2. Structured Portfolios

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self-Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
DE	Teachers create activities, which students perform throughout the year. In addition to submitting three pieces of evidence, the teacher must fill out a progress report showing data collected on at least six occasions.	The "activity" element (5) combines age appropriateness, using a schedule, using supports, and the inclusion of a progress update. "Self-determination" (5) combines choice, planning, self-monitoring, and feedback. Other elements are settings (5) and interactions (5).			Activity (5)			Activity (5)			5	5	Self-Determination (5)	Self-Determination (5)	Activity (5)		Self-Determination (5)
GA	Teachers design tasks for the student to perform during two different collection periods, at least three weeks apart. Teachers collect primary evidence of the student engaging in the task, and secondary evidence of student achievement, along with an interview and observation form.	Context (4), achievement/progress (4), generalization (4), and fidelity to standard (3). "Generalization" includes settings and interactions.			4	3			4	Generalization (4)	Generalization (4)	Generalization (4)					
KS	Teachers design tasks and select three pieces of evidence for five indicators. The student has at least five opportunities to demonstrate skill.	1-5 scale according to degree of accuracy across the five trials.															

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
KY	<p>Kentucky: Teachers design tasks that the student must perform with less than 50% accuracy at baseline. Students can perform a task up to four times for each of the five assessment targets (for a student identified with symbolic communication/dimension A) or three assessment targets (for pre-symbolic communication/dimension B) in each content area.</p>	<p>Complexity (4), supports (4), and performance accuracy (4) or progress (3). Students who communicate at the pre-symbolic level can score up to 3 points for a gain of at least 40 points over baseline performance. Dimension A students can score up to 4 points for 90 – 100% accuracy (taken from the highest scoring probe submitted for each assessment target along with one student work sample).</p>	4 (Dimension A)		3 (Dimension B)		4								4		
MD	<p>Teachers select "mastery objectives" from a task bank, or submit teacher-designed tasks for approval. Teachers collect evidence for each task, which are each scored as "mastered" or "not mastered."</p>	<p>Students must perform under 50% at baseline, and must perform with 80-100% accuracy for an objective to be scored "mastered."</p>															

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
MA	Teachers create tasks and fill out data charts to show student performance on skills related to selected standard strands on at least eight different dates, as well as two pieces of evidence for each skill. The student should perform the skill at below 80% accuracy and/or independence at baseline.	Level of complexity (5), accuracy (4), independence (4), self-evaluation (4), and generalization (3).	4	4			5			3				4			
MS	Teachers create tasks and collect evidence from the student's first and last performance of the skill. Secondary evidence (1-3 pieces per objective) should be on a different activity aligned to the same objective as the primary evidence.	Students who are symbolic language learners use the "attainment" scoring guide, which measures accuracy on a 5 point scale. Pre-symbolic language users use the "progress" scoring guide, which measures progress on a 5 point scale. Each item is also assigned up to 5 points for level of complexity. All students must score below 40% at baseline.	5 (Attainment)		5 (Progress)		5 (both)										

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
MO	Teachers create activities for students to perform six times over the course of two collection periods. Teachers may or may not provide actual student work as evidence.	Skill performance (4), level of independence (4), and connection to standards (3). Performance and independence are scored 6 times throughout two collection periods for each indicator, resulting in an average score for each.	4	4		3											
NH	Teachers create activities for each standard strand and complete a progress chart with three data points in each collection period. Two work samples per targeted skill are also required.	Evidence of progress (4), connection to general curriculum (4), supports (4), generalization (4), and self-determination (4). "Generalization" includes settings, and "self determination" includes evidence of choice, self-monitoring, planning, and self-evaluation.			4	4				generalization (4)	generalization (4)		self-determination (4)	self-determination (4)	4		self-determination (4)
NJ	Teachers create activities for each standard strand. Students perform the skill a minimum of five times within two different activities for each content strand. Two pieces of evidence for each strand are required.	Complexity/link to indicator (4), level of independence (4), and accuracy (4). The student must perform each skill with less than 40% accuracy at baseline, and final accuracy score is based on the final activity.	4	4			4										

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
NY	For each content area indicator, teachers can choose from sample tasks, modify a sample task, or create a new task. Teachers submit data sheets for trials on three different dates and submit supporting evidence.	Accuracy (4) and level of independence (4).	4	4													
OK	Teachers design activities for each standard, and record accuracy and supports pre-instruction, mid-instruction, and post-instruction. Five pieces of evidence per core subject area are required.	Level of independence (8), progress (5), accuracy (4), participation (1), connection to standard (1), and age-appropriateness (1).	4	8	5	1		1								1	
RI	Teachers design tasks, record accuracy, and collect samples of student work during three collection periods across the school year.	Connection to the content-strand (8), evidence of progress (8), accuracy (4), and level of independence (4).	4	4	8	8											

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
TN	Teachers create tasks for each alternate performance indicator (broadly defined task or goal). In addition to collecting evidence, the teacher fills out a graph to measure progress on activities.	Connection to content standards (50), documentation of progress (50), evidence of choice (20), supports (10), settings (10), and interactions (10). For students with excessive absences, another rubric applies: connection to content standards (30), documentation of progress (30), choice (12), supports (6), settings (6), and interactions (6). A third rubric applies to homebound students: connection to standards (30), documentation of progress (30), and choice (12).			50/30/30	50/30/30					10/6/-	10/6/-			10/6/-		20/12/12
VT	Teachers develop tasks for each standard strand (must be <50% independent performance at baseline), and submit endline products and data charts for each in the spring.	After demonstrating an appropriate baseline and evidence of instruction, products are rated "strong," "sufficient," or "insufficient" for alignment depth/ breadth and accuracy.	strong/sufficient/insufficient			strong/sufficient/insufficient											

State	Test Details	Scoring Details	Accuracy / Performance	Level of Independence	Progress	Alignment to Standards	Complexity	Appropriateness	Context	Generalization	Settings	Interactions	Self Determination	Self-Evaluation	Supports/Accommodations	Participation	Choice
WA	Teachers set a performance goal for each standard strand, and design activities to measure each goal. Teachers collect evidence and record student performance in percentage accuracy at baseline, midline, and endpoint. Baseline percentage cannot exceed the stated goal.	Alignment with grade level expectations (4), alignment to targeted skill (4), level of performance (4), and generalization/contexts (4).	4			4 + 4				4							

APPENDIX C: Rating Scales

State	Test details	Scoring details
CT	Teachers rate students on skills that are downward extensions of each essence statement for each performance standard. Ratings are based on trained teacher interactions and observations throughout the year. Evidence is collected only for monitoring requirements.	Each skill is rated as mastery/independent, developing/supported, or does not demonstrate. Mastery indicates accuracy and independence at least 80% of the time.
HI ²⁷	The teacher chooses standards that are aligned to the student's IEP goals, and collects two pieces of evidence of student performance related to each. All items on the list must be rated, even if they are not IEP-aligned.	Each standard is rated by the teacher and a second rater as "non-existent," "emerging," "progressing," or "mastered."
IA	Teachers instruct students on selected items from a list of pre-scripted skills. For the assessment, the student performs four trials of selected skills, and the teacher submits evidence of student performance on those trials.	The teacher scores for percent accuracy on the most recent trial for each skill, or marks which skills were "already mastered," "not taught," or "fully prompted." At least one skill per content area must have been taught or the student will count as an "exclusion." Teachers are encouraged to teach as many skills as possible.
IN	The teacher rates the student on a series of skill progressions based on observations and evidence collected over the course of the year. Evidence is not submitted.	Each student is rated through the use of a matrix of advancing approximations (a performance thread) to the content standard of the grade level band in which the student is enrolled.
SD	All skills on the rating form are rated. Additionally, teachers select one item from each indicator in reading and each content strand in math and science, and collect evidence of student performance on those items, documenting at least 3 trials of each skill.	A five point rating scale combines accuracy and level of assistance for each item.

²⁷ Hawaii will use a new assessment format in 2010-11.

APPENDIX D: Item-based tests

1. Performance tasks only

State	Test details	Scoring details
CA	Students take test II – V based on grade-level, but students “with the most significant disabilities” take test I.	For level II – V tests, tasks are scored up to 4 points for mastery and level of task completion (partially completes task, minimally completes task). For level I tests, tasks are scored up to 5 for a combination of accuracy and level of prompting/assistance.
NM	Twelve to sixteen tasks are grouped according to difficulty, comprised of a mix of low, moderate, and high tasks. Each task may have multiple items. The teacher uses a Student Placement Questionnaire to determine the task level at which to start, and uses a scripted system of assistance.	Each task is scored up to 2 or 3, according to a scripted scaffolding system combining accuracy and level of assistance.
OR	A set of pre-requisite skills determine the level of support allowed.	2 (correct), a 1 (partially correct), a 0 (incorrect), a D (teacher did not administer item because it was deemed too difficult for the student), or I (inappropriate).
TX	For each standard “essence statement,” teachers can choose tasks from three levels of complexity.	Accuracy (2), but level three tasks are weighted by 1.5, level two tasks are weighted by 1.2, and level one tasks are weighted by 1. Level of support (2) according to standardized levels of prompting. For a level 2 or 3 complexity level task, the student can earn an additional generalization point for each predetermined criterium that is performed without prompting.
UT	The student performs each task three times across multiple settings.	Each task can be scored up to 4 levels: level 1 is minimal (no correct trials), level 2 is partial (1 correct trial), level 3 is sufficient (2 correct trials), and level 4 is substantial (3 correct trials, but also 3 activities/objects, 3 people, and 3 settings).

2. Multiple-choice items only

State	Test details	Scoring details
FL	Each question is written at three levels of complexity.	Students progress through three levels of complexity per item in a grade level content based assessment (starting at Participatory). Possible item scores are 0, 1, 2, 3, 6, or 9 based on the highest level of complexity (3, 6, or 9) or level of support (2, 1, or 0) at which a student provides an accurate response.
IL	Three answers choices per item.	Scaffolded scoring 1–4 combining accuracy and level of assistance. A correct answer after the administrator provides the answer is scored 2, and an incorrect answer is scored 1.
MN	Three answer choices per item.	Scaffolded scoring 0–3. An incorrect answer after two levels of assistance is scored a 1. No answer is scored 0.
NE	Three answer choices per item.	Scored for accuracy.
SC	Related multiple choice items are grouped together in a set to form a "task."	Maximum points vary for each item, and points are scaffolded, combining accuracy and level of assistance. A small number of tasks are rated only for level of engagement.

3. Constructed-response items only

WY	Also requires a portfolio.	Items use a scaffolded scoring system, combining accuracy and level of assistance, up to 4 points. Hand-over-hand assistance results in score of one point. Zero points are earned if the student refuses to complete the task.
----	----------------------------	---

4. Multiple-choice items and performance tasks

State	Test details	Scoring details
LA	Activities may require students to choose an appropriate answer option, perform sorting and sequencing tasks, work with simple manipulatives, or identify something included within a simple scenario.	Each performance task is scored on a 0 to 2 point or a 0 to 1 point scale, according to an item-specific rubric. Two-point tasks allow the possibility of a partially correct response.

5. Multiple-choice items and constructed-response items

State	Test details	Scoring details
CO	Students choose from four possible answers or produce an independent response.	Constructed response items are scored up to 6 points, combining accuracy and level of independence, and allowing for partially correct answers. Multiple-choice items are scored 1-3 for a correct response according to level of independence.
ND	Teachers choose one scripted task per standard for the student to perform on four separate trials.	Items are scored primarily for accuracy, but a set of “secondary indicators” can add additional points, including for settings, choice, planning, supports, and self-monitoring.
WV	Four choices per multiple-choice item.	CR items are scaffolded up to 6 points, and MC items are scaffolded up to 3 points, both combining accuracy and level of assistance.

6. Multiple-choice items and writing prompts

State	Test details	Scoring details
MI	Functional Independence tests are multiple choice with writing prompts.	On the Functional Independence test, multiple-choice items are scored for accuracy only.
NV	Three answer choices per item, plus writing prompts.	Scored for accuracy, or flagged as "guided response," which indicates that the student could not answer without teacher intervention (which renders a 0 score).

NC	Three answer choices (in the form of cards) per item, plus writing prompts.	Scored for accuracy. Writing prompts scored for factual content.
----	---	--

7. Performance tasks and constructed-response items

State	Test details	Scoring details
MI	Participation and Supported Independence tests are a combination of performance tasks and constructed response items.	For Participation and Supported Independence tests, items are scored up to 3 (P) or 2(SI) for a combination of accuracy and independence.

8. Multiple-choice items, constructed-response items, and writing prompts

State	Test details	Scoring details
WI	Items include multiple choice, open-ended short answer, and a writing prompt.	Multiple-choice items are scored for accuracy, and constructed-response items may be awarded partial credit for a partially correct answer.

9. Multiple-choice items, performance tasks, and constructed-response items

State	Test details	Scoring details
AK	Activities may include selecting the appropriate card, responding to a question, or sequencing cards.	Regular test items are scored for performance and can include partial credit. If a student scores a zero on three consecutive items in three consecutive tasks for a content area, the assessor administers the Expanded Level of Support (ELOS) items, which represent prerequisite skills. ELOS items are scored 1 – 4 according to level of assistance.
AZ	Comprised of multiple-items along with “Rater Items” and “Performance Tasks,” which are each a combination of constructed-response items and performance tasks.	“Performance Tasks” (both PT and CR) are scored on a 0-2 scale of no response, modeled responses to independent responses. “Rater Items” (both PT and CR) are scored up to 4 points for a combination of accuracy and level of independence. Multiple-choice items are correct/incorrect.
ME	Teachers select tasks from a task bank according to age-appropriate level of complexity (with some flexibility within grade band).	A task that is not age appropriate is unscorable. Level of complexity (8), accuracy (3), and level of assistance (3).

MT	Items are a combination of multiple-choice questions, yes/no questions, open-ended questions, and scripted tasks. Evidence is required for some items.	Scaffolded scoring, combining accuracy and level of independence, up to 4 points.
PA	Tests are designed at three levels of complexity (A, B and C), and item types are in different proportions depending on grade and level of complexity. The levels are pre-assigned to students based on past test performance or, for initial test takers, on a checklist completed by their teachers.	Scaffolded scoring, accuracy and level of independence, up to 5 points.

QuickTime™ and a decompressor are needed to see this picture.

