Equity in science education for SWSCD through alternate content standards

Lori Andersen

Sue Bechard

Katherine Merriweather

University of Kansas

Author Note

Paper presented at the 2016 annual meeting of the National Association for Research in Science Teaching, Baltimore, Maryland. Correspondence concerning this paper should be addressed to Lori Andersen, Center for Educational Testing and Evaluation, University of Kansas, landersen@ku.edu.

Citation:
Andersen, L., Bechard, S., & Merriweather, K. (2016, April). *Equity in science education for SWSCD through alternate content standards*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Baltimore, MD.

Abstract

Students with significant cognitive disabilities (SWSCD) are the one percent of the total student population who are eligible to participate in alternate assessments based on alternate achievement standards. The publication of *A Framework for K-12 Science Education* (National Research Council, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013) have created a need for new alternate content standards and alternate assessments in science that are linked to the new general education science standards. This paper describes how a consortium of four states used Evidence-Centered Design (Mislevy, Steinberg, & Almond, 2003) and Universal Design for Learning (CAST, 2012) to develop alternate science content standards and assessments. A set of 27 alternate science content standards was created, followed by an alternate assessments at three grade spans. Evidence that supports appropriateness of the alternate standards for SWSCD and fidelity of representation of the *Framework* is presented. Two cycles of testlet/item development were conducted. Results of a pilot test (251 items; 1,606 students) and a field test (259 items; 5,663 students) are presented. Evidence for validity and accessibility of the alternate assessment is presented. Major findings include that the assessment items met accessibility, bias and sensitivity, and content requirements, and that students were able to understand and respond to assessment items. A critical issue for validity of assessment items is that students need to have been taught the content that is assessed. A survey of teachers was conducted during the field test to determine how well the enacted curriculum corresponded to the alternate content standards. Results of the survey are presented. Major findings include that many SWSCD did not experience a science curriculum that included all of the disciplinary core ideas or practices that are in the alternate science content standards. Implications of these findings for curriculum and instruction are discussed.

Introduction

Students with significant cognitive disabilities (SWSCD) comprise about nine percent of the population of students with disabilities, or about one percent of the overall student population. While this low-incidence population of students is highly heterogeneous, they exhibit several general characteristics: 1) They have a disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior; 2) They are primarily instructed using alternate content standards that are less complex than grade-level content standards; and 3) They require extensive direct individualized instruction and substantial supports to achieve measureable gains in the grade-and age-appropriate curriculum (DLM, 2013, September 24).

Historically, SWSCD have received little instruction in science across all grade levels, despite an emphasis on *science for all* in science education policy documents over the past 20 years (e.g., National Research Council, 1996; 2012). Although alternate standards and assessments that are linked to the grade-level standards are allowed for SWSCD (U.S. Department of Education, 2004), enacted science curricula remain far from supporting this option (Karvonen et al., 2011). For example, science instruction experienced by SWSCD has either lacked the rich science concepts that their general education peers were provided, or has provided content that was appropriate for much younger students (Courtade, Spooner, & Browder, 2007; Karvonen et al., 2011). The publication of *A Framework for K-12 Science Education* (National Research Council, 2012) and the *Next Generation Science Standards* (NGSS Lead States, 2013) have created a need for new science alternate assessments that are linked to the new science standards and provide equitable access to science for this population of students. This paper describes the development and validation of the alternate science content standards and assessment items to meet that need.

Literature Review

Large-scale alternate assessments for SWSCD are based on rules published in the Federal Register (December 9, 2003), which provided states, school districts, and schools with the flexibility to use scores from alternate assessments based on alternate achievement standards (AA-AAS) to determine proficiency of SWSCD for accountability purposes. State-specific alternate assessments in English language arts (ELA) and mathematics were used beginning in 2000 and science assessments followed in 2007. With the advent of college and career ready standards and the formation of multi-state assessment consortia, alternate content standards and multi-state common assessments began to be developed for SWSCD in 2010 for ELA and mathematics.

Alternate assessments in science continue to be largely state-specific. Rogers, Thurlow, and Lazarus (2015) analyzed states' science AA-AAS for 2014-2015, and found inconsistencies in the science content that was assessed across states. Rogers et al. (2015) concluded that some SWSCD might not have access to the "same rigorous grade-appropriate content" (p. 8). Providing access to such rigorous grade-appropriate content for SWSCD is a goal of the four-state consortium and the focus of the present study.

Evidence-Centered Design

Alternate assessments provide opportunities for students with significant cognitive disabilities to demonstrate understanding of academic content (Browder et al., 2007). Browder et al. (2007) argued there are two primary challenges when developing alternate content standards that link alternate assessment to general education content standards. The first challenge is defining the appropriate breadth and depth of curriculum for students with significant cognitive disabilities and the second challenge is to determine the degree of fidelity between the skill chosen for assessment and the original content (Browder et al., 2007). Standards for educational testing (AERA, APA, NCME, 2014) inform the identification of relevant evidence for AA-AAS, including evidence for the quality of the items for this unique population, the appropriateness of the assessment content for each student, and the accessibility of the assessment for each student. The challenges presented by the accessibility needs of SWSCD and differences in the new science framework (e.g., three-dimensional science learning) require new approaches to assessment design (National Research Council, 2014). Evidence-centered design (ECD) enables such challenges to be addressed; the collection and analysis of evidence such as reviews of proposed alternate content standards and assessment items by content and special education experts allows improvements in accessibility and fidelity of content representation. Evidence-centered design is a conceptual model for designing, producing, and delivering educational assessments that has been recommended for developing assessments based on the new *Framework* (National Research Council, 2014). The use of ECD supports creating tests that are valid for their intended purposes by establishing an evidentiary argument explicating the relationships among the inferences about the student, the observations that provide evidence for those inferences, and the features of situations that evoke that evidence (Mislevy, Steinberg, & Almond, 2003). In the present study, some components of ECD were used in the assessment development process and were tailored for SWSCD to incorporate considerations of cognitive complexity, accessibility, and computer-based adaptive test delivery for alternate assessments in science.

Evidence-Centered Design has been used to design large-scale assessments for K-12 general education students (Cameto & Haertel, 2011; Huff, Steinberg, & Matts, 2010), and has recently been used in the development of alternate assessments in mathematics (DeBarger et al., 2011; Flowers et al., 2015). DeBarger et al. (2011) used ECD principles to develop guidelines for design patterns, test development specifications, and exemplar task templates for an alternate assessment in mathematics, with the purpose of “developing assessments that adequately and reliably show what these students know and can do” (p. 2). The test pattern guidelines considered the enormous variability in the target population, assumptions about measuring their achievement, and the variability of design implementation procedures that made traditional assessment design approaches inapplicable without some reformulation in order to support the validity of inferences from alternate assessment scores. Especially critical to the design of alternate assessments is an understanding of the unique cognitive characteristics and access needs of SWSCD (Karvonen, Bechard, & Wells Moreaux, 2015). The wide range of communication skills within the SWSCD population presents a challenge when developing, implementing, and evaluating alternate assessments (Karvonen et al., 2011). It is particularly challenging to provide access for the students who have the lowest symbolic communication abilities. Symbolic communication abilities have been noted as the factor with the strongest association with students' access to academic content (Karvonen, Flowers, & Wakeman, 2013). To address this challenge, principles of Universal Design for Learning (UDL; Cast, Inc., 2012) were integrated into the ECD framework to promote accessibility of items through consideration of student needs and abilities at the outset of the design process.

The science consortium used an adapted version of the DeBarger ECD template to provide appropriate and sufficient evidence at many points in the test development process. The adapted templates were originally adopted by 16 states in a large alternate assessment consortium and utilized in the development of assessments for English language arts and mathematics (Bechard & Sheinker, 2012). The templates were specifically designed for clarity and ease of use, as the project engaged non-professional item writers who needed to create a large number of items in a constricted timeframe. One key difference between the two versions is that the science design template reflects a narrower focus of three levels of cognitive complexity compared to five levels of cognitive complexity used in English language arts and mathematics.

Universal Design for Learning

Universal Design for Learning is a model for creating instructional goals, assessments, methods, and materials that are accessible for all students. The UDL model uses three factors to increase accessibility when adapting materials to learner characteristics: multiple means of engagement, multiple means of representation, and multiple means of action and expression. The UDL framework is critical to enabling successful interactions between students with the most significant cognitive disabilities and academic content (Wells-Moreaux, Bechard, & Karvonen, 2015). A prior census of SWSCD (n=44,000; Dynamic Learning Maps, 2013) informed accessibility needs. The wide range of cognitive abilities, communication skills, and disabilities in this population mean that students access the content in different ways.

All students require an adequate introduction to an activity or assessment task. The alternate science assessments provide introductory activities that are designed to engage the student in the assessment task by eliciting prior knowledge, introducing the content, and stimulating interest. The majority of students' reading levels are at or below the second grade level. This led to decisions concerning multiple means of representation, particularly the accessibility of text and the use of core vocabulary. Thus, high school science content standards for SWSCD address high school science concepts, which are made accessible through text that is written at a much lower grade level. In addition, assessment materials are presented via multiple means, both via computer and the test administrator (e.g., images, video, text - including read aloud and Braille, and manipulatives). Finally, multiple means of action and expression are provided, due to students’ diverse methods of communication. While 76% of the students use expressive speech, 24% do not use speech. These students rely on other means of expression, including sign language (8%), augmentative and alternative communication devices (AAC; 19%, including eye gaze, communication boards, and voice output devices), and symbols (10%). Therefore knowing the characteristics of the students is critical to making pedagogical decisions that make science instructional goals and assessments accessible to SWSCD.

In the present study, components of ECD and UDL were used synergistically to develop alternate content standards and assessment items. Three research questions were addressed:

(1) How can the science disciplinary core ideas and science and engineering practices in the Framework be made accessible to SWSCD?;

(2) How can we know if our attempt to make the *Framework* accessible was successful?; and

(3) How well does the enacted science curriculum for SWSCD reflect the new *Framework*?

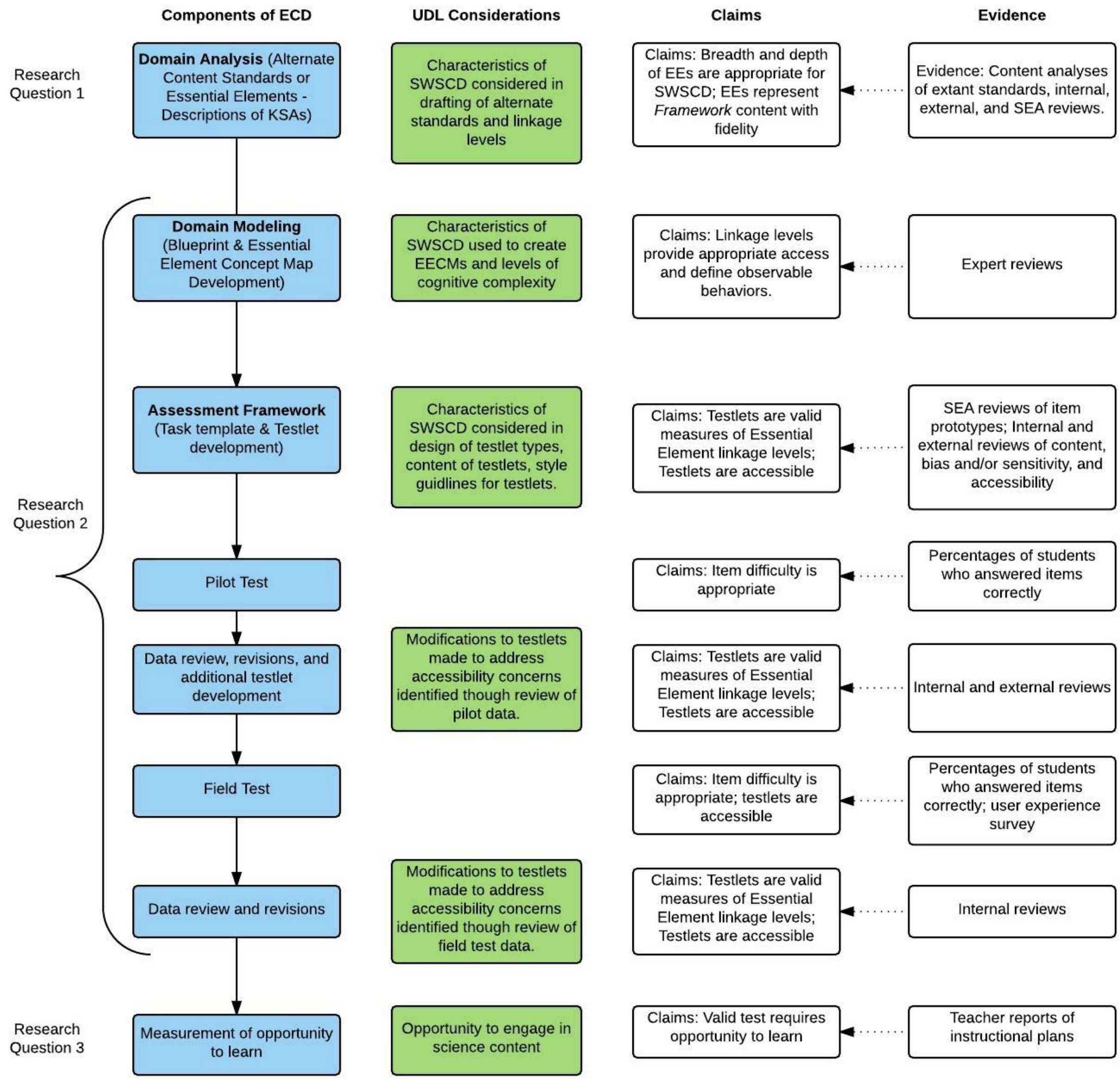
Method

Study Context

Four U.S. states, which comprised the Dynamic Learning Maps Science Consortium in the 2014-15 academic year, participated in the development of the alternate science content standards (DLM Science Essential Elements) and assessment blueprints. A sample of 1,606 students from the four states participated in a May 2015 pilot test. The participants were in grades 3 through 11, with 36% in elementary school, 35% in middle school, and 29% in high school. Data from the pilot were used as evidence to make improvements to assessment items. New or revised items were created based the interpretation of this evidence, which were field tested in November 2015. The field test included 5,663 students from five states. In November 2015, a survey was administered to 2,770 teachers of SWSCD using a stratified random sample based on student’s state, grade, and school; 872 teachers of students responded. Results from both the pilot test, field test, and survey are presented.

Figure 1

Overview of Study



The method for the development of both the alternate science content standards and the assessments used components of ECD (Mislevy et. al, 1999). An overview of the study method is presented in Figure 1. To answer the research questions, multiple components of ECD were used, which involved determining the specific content to be included in the assessment and creating descriptions of the focal knowledge, skills, and abilities, as well as potential work products. Principles of UDL were integrated into the ECD framework and promote accessibility through consideration of student needs and abilities throughout the design process. Throughout the development process, a large amount of data were collected as evidence for the claims, as indicated in Figure 1. These data include expert and teacher reviews of alternate standards, blueprints, and assessment items, as well as data from the administration of assessment items and teacher surveys.

Method for Research Question 1: *How can the science disciplinary core ideas and science and engineering practices in the Framework be made accessible to SWSCD?*

The UDL model was applied while creating alternate science content standards for SWSCD to provide access to science content at the appropriate levels of cognitive complexity, as described below. The DLM alternate science standards have three levels of complexity for each standard, called linkage levels. Each of the three linkage levels (initial, precursor, and target) represent a significant milestone en route to the alternate standard (Wells-Moreaux et. al, 2015). The target level is the grade-span alternate content standard and is the most closely linked to the general education standards, while the precursor and initial linkage levels represent decreased levels of complexity. The initial linkage level is the least complex of the three linkage levels and is designed for students who require test administrators to deliver the tasks and record their responses. During test administration, students are assigned to a linkage level based on a teacher-completed survey of the students' expressive communication abilities (Wells-Moreaux, et al, 2015).

The DLM science state partners engaged in a multi-step process of creating the alternate science content standards, called Essential Elements (EEs). An initial consideration was the breadth and depth of science content that should be included in the EEs. The DLM science states chose to develop EEs in topics that were already included in their extant alternate science assessments. Seven states were initially interested in developing the new science assessment. Therefore, the first step was to identify common topics through content analysis of the extant alternate science standards.

Seventeen common science topics in the three science disciplines (Physical Science, Life Science, and Earth and Space Science) across state's alternate standards were identified, which allowed reduction of the breadth and depth of the general education science standards content to a subset of topics that were viewed as most relevant to SWSCD by consortium stakeholders. These topics served as the starting points for developing the EEs. To strengthen the link between the EES and the general education standards, the resulting topics were mapped onto the organizational structure described in *A Framework for K-12 Science Education* (National Research Council, 2012), which provided the structure to organize science content. The *Framework* organizes topics within a corresponding, overarching disciplinary core idea, with 10 disciplinary core ideas that each contain three to five topics (National Research Council, 2012). Grade-level content within these topics was identified from the grade-span content progressions in the *Framework*.

The *Framework* provides content progressions across four grade spans, which were used to identify grade-level content. For the purpose of this project, elementary (grades 3-5), middle school (grades 6-8), and high school (grades 9-12), were selected as the grade spans at which students would be assessed. The *Framework* conceptualizes science learning as organized within three dimensions; disciplinary core ideas, science and/or engineering practices, and crosscutting concepts. The DLM science EEs uses two of the three dimensions: disciplinary core ideas and science and/or engineering practices. The disciplinary core ideas for the new alternate content standards were determined by the previously described content analysis. For this pilot project, the crosscutting concepts were not formally targeted as learning goals. Each Essential Element integrated a disciplinary core idea with a science and/or engineering practice. Appropriate grade-level content for the disciplinary core ideas and the science and/or engineering practices was identified using the *Framework* as the link to the general education standards.

The Essential Elements were drafted and revised through the application of UDL principles and feedback from content and special education experts and teachers. The resulting Essential Elements were less complex and more universally accessible than the grade-level, general education expectations in terms of both the disciplinary core idea and the science or engineering practice. These reductions in cognitive complexity were intended to make the *Framework* accessible to students with significant cognitive disabilities. In this manner, an initial set of Essential Elements, each with three linkage levels, was drafted. The concept of linkage levels is consistent with the need for appropriate levels of cognitive complexity for SWSCD and the requirement for adherence to alignment with grade level content standards. In addition, the EEs were crafted to allow for multiple means of representation and expression.

A panel of internal experts who specialized in teaching SWSCD and/or science reviewed the first draft of proposed Essential Elements (EEs). The second draft was presented to representatives from each state education agency and the educators and content specialists who they selected. Sixteen experts in science, as well as 17 individuals with expertise in instruction for students with significant cognitive disabilities reviewed the second draft. A third draft was then reviewed internally by each state. Questions that guided the review focused on investigations of the challenges inherent in AA-AAS described above and included specific inquiries related to the following questions: (1) Do the EEs fit within the topics and core ideas that are the framework for the DLM system?; (2) Do the EEs in each topic support student learning over time?; (3) Are the EEs and linkage level learning targets clearly defined?; and (4)

Do the linkage levels represent the learning target content at appropriately reduced levels of breadth and complexity?. The final EEs resulted from a final discussion and consensus vote.

Method for Research Question 2: *How can we know if our attempt to make the framework accessible was successful?*

Evidence of the accessibility of the EEs was obtained via student performance on assessment items that corresponded to each EE. Assessment items are administered in testlets, which are packages of three to four assessment items centered on a learning target. At the initial level, testlets are test administrator observations, while computer-administered testlets are used at the precursor and target level. All assessments are administered in a one-on-one, test administrator/student format.

Teacher-administered testlets consist of three to four questions that each consist of a script of statements about two picture cards followed by a question. The students selects the picture card that corresponds to the answer and the response is recorded by the test administrator. The basic structure of computer-administered testlets consisted of an engagement activity followed by three to four multiple choice items that were related to the engagement activity. The engagement activity is designed to activate prior knowledge, motivate the student, and provide a context for the items in the testlet. Engagement activities have a variety of forms, which include stories about students engaging in science experiments, descriptions of science phenomena, and informational text.

The review process of test items follows an existing 26-step procedure developed for the English language arts and mathematics assessments. This rigorous process includes multiple stages of review for content, alignment, bias, sensitivity, and accessibility to insure items assess what they are intended to measure. The reviews are conducted by the science content team, editors, internal content and special education reviewers, as well as external content and special education reviewers. Data were collected at each step of the review process. The criteria used for review are presented in Appendix A.

The content and special education internal review process for science was conducted by teachers who either possess expertise in science content or who teach students with significant cognitive disabilities. The reviewers complete trainings on the DLM assessment program and the review criteria. The internal content and special education review focuses on: adherence to DLM style guidelines, quality of science content, accessibility issues, and bias concerns. Testlet content is reviewed for clear alignment with the linkage level in terms of science concept and science or engineering practice, appropriateness of the depth of knowledge classification and the complexity of the task, quality of answer options, and correctness of science content. Testlets are reviewed for compliance with accessibility criteria, which included appropriateness of: cognitive load, text complexity, images, and alternate text for images. Bias considerations included item dependence on prior knowledge or experiences. Reviewers entered evaluative information into an online survey and/or recommended revisions to testlets. Items and testlets that did not meet criteria were revised.

The external review process for science was conducted by teachers who either possess expertise in science content or who teach students with significant cognitive disabilities. Reviewers completed applications and were selected based on expertise and experience criteria. Reviewers completed online training on the DLM program, student population, and test design criteria. Reviews were completed by a panel. Each reviewer was assigned to evaluate one specific category, either accessibility, content, or bias. External reviewers entered evaluative information through the content builder system. Testlets and items that were flagged by external reviewers were examined by the content team for revision or rejection. Revisions were made as needed to address reviewer concerns.

After testlet and item quality check requirements were met, a pilot test was administered in May 2015. A total of 1,606 students from four states participated. Each student in the pilot was administered a fixed-form test of nine testlets at the same linkage level, which was assigned based on the student’s expressive communication abilities.

Data from the pilot were reviewed by the content team to evaluate items. The percentage of students who answered each item correctly (p-value) was determined. The format of the Items was multiple choice with three answer options. If a minimum threshold of 35% of students answering correctly was not met, the items were flagged for review. Flagged items were discussed and possible causes for the flag were considered. In addition, a DIF analysis was conducted to check for unusual patterns of responses. Group consensus was used to make item-level decisions.

A field test was administered in November 2015 that included 5,663 students from four states. Each student in the field test was administered a randomly assigned, fixed-form test containing three testlets. Two testlets were at the student's assigned linkage level and one testlet was at an adjacent linkage level that was either higher or lower than the assigned level. Only the data from testlets that were at linkage levels matching the student's assigned communication band linkage level were included in the analyses. The process used to evaluate field test items was the same as the process used for the pilot test items.

Method for Research Question 3: *How well does the enacted science curriculum for SWSCD reflect the new Framework?*

Student performance on the items was one measure of validity and accessibility. However, low performance on assessment items can stem from several possible causes, including students' lack of opportunity to learn the content that was assessed. If students were not provided instruction in the content that was assessed, item performance cannot be interpreted in a meaningful way. Therefore, data from assessment items must be considered in light of students' opportunities to learn the content and skills that were assessed. This concern was addressed in research question three.

To approximate the alignment between the currently enacted alternate science curriculum and the Essential Elements, a teacher survey was designed and administered in November 2015 to a stratified random sample of teachers. A set of 18 survey questions was designed to measure students' opportunity to learn the science disciplinary core ideas and apply the science and engineering practices that are used in the Essential Elements during science instruction. For each of the disciplinary core ideas from the *Framework* that are used in the Essential Elements, teachers were asked how many hours of instruction were planned for the school year for that student in that core idea. Respondents selected from five choices: none, 1-10 hours, 11-20 hours, 21-30 hours, or more than 30 hours. Teachers were also asked about student skill use during science instruction for eight science skills that were used most frequently in the Essential Elements. Respondents selected which of the eight skills the student had used from a multiple-select list.

Results

Research Question 1: *How can the Framework be made accessible to SWSCD?*

A content analysis of existing alternate science standards from seven states was conducted. Content analysis and mapping of the common topics to the *Framework* resulted in a list of 17 DCI *Framework* topics (Table 1).

Table 1

Common Disciplinary Core Ideas (DCI) Topics in State Alternate Science Standards

Science Area	Common Disciplinary Core Idea Topics
Earth and Space Science	ESS.1B Earth and the Solar System ESS.2A Earth materials and systems ESS.2D Weather and climate ESS.3A Natural resources ESS.3C Human impacts on Earth systems
Life Science	LS.1A Structure and function LS.1B Growth and development of organisms LS.2A Interdependent relationships in ecosystems LS.2B Cycles of matter and energy transfer in organisms LS.3A Inheritance of traits LS.3B Variation of traits LS.4C Adaptation
Physical Science	PS.1A Structure and properties of matter PS.2A Forces & motion, PS.2B Types of interactions PS.3D Energy and chemical processes in everyday life PS.4A Wave properties

Note: 7 states' alternate content standards were analyzed

Forty-three alternate content standards, or Essential Elements (EEs) were developed that aligned to the disciplinary core idea topics. These EEs represent a breadth of content across the three domains of Life Science, Physical Science, and Earth and Space Science. Each EE has three linkage levels (initial, precursor, and target). Table 2 shows an example of one NGSS performance expectation and the corresponding three Essential Element linkage levels. The set of 43 EEs was finalized in December 2014 (Dynamic Learning Maps, 2014) after multiple reviews by educators and state education agencies. An example of an EE and the corresponding linkage levels is shown (Table 2).

Table 2

Essential Element and Linkage Levels for one NGSS Performance Expectation

NGSS Performance Expectation	5-ESS3-1: Obtain and combine information about ways individual communities use science ideas to protect the Earth's resources and environment.
Disciplinary Core Idea	ESS3.C: Human Impacts on Earth Systems
Science and Engineering Practice	Obtaining, Evaluating, and Communicating Information
Essential Element Code	EE.5.ESS3-1
Target Linkage Level	Use information to describe how people can help protect the Earth's resources and how that affects the environment.
Precursor Linkage Level	Compare two methods people can use to help protect the Earth's resources.
Initial Linkage Level	Identify one way to protect a resource of Earth (e.g., put paper in the recycling bin).

(Dynamic Learning Maps, 2014)

Research Question 2: *How can we know if our attempt to make the Framework accessible was successful?*

As described in the methods section, all items were evaluated by an internal team of content and special education experts for alignment to linkage level specifications and for meeting accessibility guidelines. Overall, 89% of 296 testlets were found to be content-aligned in their current form and the remaining 11% were revised. Accessibility reviews of testlets found that 49% met vocabulary guidelines, 82% met text guidelines, and 87% met image guidelines. As far as the needs of specific students in the SWSCD population, 97% of testlets were found to be accessible to students with limited working memory and 97% were found to be accessible to students with limited implicit understanding of others' emotions and intentions. Before external review, revisions were made to items. During external review, 2,446 item evaluations were conducted by panels that focused on accessibility, bias, and content specifications, with an average item rejection rates of 1.7% and 1.6% for the pilot and field test, respectively (Tables 3 and 4).

Table 3

External Review of spring 2015 Field Test Items

	Accessibility		Bias/Sensitivity		Content		Overall	
	Number	%	Number	%	Number	%	Number	%
Accept	545	84%	614	87%	566	80%	1725	84%
Revise	94	15%	85	12%	127	18%	306	15%
Reject	8	1.2%	8	1.1%	19	2.7%	35	1.7%
Total	647		707		712		2066	

Table 4

External Review of fall 2015 Field Test Items

	Accessibility		Bias/Sensitivity		Content		Overall	
	Number	%	Number	%	Number	%	Number	%
Accept	102	89%	144	96%	101	88%	347	91%
Revise	11	10%	6	4.0%	10	8.7%	27	7.1%
Reject	2	1.7%	0	0%	4	3.5%	6	1.6%
Total	115		150		115		380	

During spring 2015, a set of testlets were administered to 1,606 students. One testlet for each Essential Element linkage level, consisting of three to four items, was administered. The percentage of students who correctly answered assessment items was evidence of accessibility. Items that were answered correctly by fewer than 35% of students were flagged for review. Of 251 items administered during the spring 2015 pilot test, 38 (15%) were flagged for review. Of the 38 flagged items, nine (26%) were rejected and 28 (74%) items were rewritten. A pattern was noted in the rejected testlets. Five of the six rejected testlets (83%) were at the precursor level and one (17%) was at the initial level.

Testlets or items that were revised as well as newly created testlets or items were evaluated in the November 2015 field test. Of 259 items administered during the November 2015 field test, 74 (27%) items were flagged for review. Of the 74 flagged items, 15 (20%) were accepted, 50 (68%) were revised, and 8 (11%) were rejected. No testlets were rejected. Summaries of the average p-values by science domain and grade span are shown in Tables 5 and 6. The pattern of item flags across linkage levels was different than the pattern noted during the pilot test.

Table 5

Summary of Average P-values in Each Science Domain for the spring 2015 Pilot

	Earth & Space Science	Life Science	Physical Science
Elementary	0.43-0.75	0.45-0.61	0.38-0.71
Middle School	0.43-0.69	0.44-0.84	0.44-0.64
High School	0.32-0.82	0.41-0.77	0.41-0.67

Table 6

Summary of Average P-values in Each Science Domain for the fall 2015 Field Test

	Earth & Space Science	Life Science	Physical Science
Elementary	0.18-0.70	0.26-0.65	0.23-0.77
Middle School	0.24-0.88	0.17-0.88	0.15-0.78
High School	0.26-0.76	0.19-0.81	0.13-0.84

Table 7

Summary of Average P-values by Linkage Level and Domain for the fall 2015 Field Test

	Earth & Space Science	Life Science	Physical Science
Initial	0.18-0.56	0.18-0.58	0.13-0.58
Precursor	0.24-0.77	0.17-0.76	0.22-0.78
Target	0.25-0.88	0.26-0.88	0.15-0.84

Five precursor level testlets were rejected based on pilot data review. Changes were made to the design of these testlets to address accessibility. New testlets were administered during the field test. Average p-values were compared to evaluate changes made to precursor testlet design after pilot data review (Table 7).

Table 8

	Pilot			Field Test		
Linkage Level	n	p-value	flags	n	p-value	flags
EE.5.ESS1-2.P	126	.2989	3	50	.4733	0
EE.5.ESS3-1.P	126	.2493	3	47	.4468	0
EE.MS.PS3-3.P	110	.2888	3	57	.3391	2
EE.HS.PS1-2.P	65	.2654	3	68	.4902	0
EE.HS.ESS3-2.P	66	.2980	2	48	.5417	0

The five precursor level testlets that were rejected after pilot data review had p-values that ranged from .2493 to .2989, with two or three flags per testlet. The revised versions had field test p-values that ranged from .3391 to .5417 and only one of the revised testlets had any item flags.

Research Question 3: *How well does the enacted science curriculum for SWSCD reflect the new Framework?*

A survey was administered to 2,770 teachers of SWSCD using a stratified random sample based on student's state, grade, and district. The survey response rate was approximately 31% (n = 871). The respondents taught at a variety of grade spans; the percentages of high school and middle school teachers were the same (36% each) and the rest were elementary school teachers (29%). The majority of the teachers' students were assigned testlets at the target linkage level (45%), while 27% were at the initial level and 19% were at the precursor level, based on the student's' expressive communication abilities. Teachers were asked how many hours of science instruction were planned, for the student who was tested, in each disciplinary core idea that is included on a DLM Science blueprint. Frequency distributions of responses for each core idea are provided in Table 8. Low cell counts led to the decision to collapse the last three categories into one called "11 or more hours".

Table 9

Academic-Year Instruction Plans for Students by Disciplinary Core Idea (n = 871)

Disciplinary Core Idea	None	1 to 10 hours	11 or more hours
PS1: Matter and its interactions	171 (20%)	498 (57%)	187 (23%)
PS2: Motion and stability: Forces and interactions	207 (24%)	492 (57%)	158 (19%)
PS3: Energy	167 (19%)	512 (59%)	177 (22%)
LS1: From molecules to organisms: Structures and processes	248 (29%)	448 (51%)	159 (20%)
LS2: Ecosystems: Interactions, energy, and dynamics	218 (25%)	437 (50%)	201 (25%)
ESS1: Earth's place in the universe	172 (20%)	473 (54%)	211 (26%)
ESS2: Earth's systems	108 (12%)	490 (56%)	257 (32%)
ESS3: Earth and human activity	164 (19%)	496 (57%)	195 (22%)

Overall, the percentage of teachers who reported planning one to ten hours of instruction for that student for this academic year ranged from 50 to 59% across core ideas. Across core ideas, 12 to 29% of teachers reported planning no hours of science instruction for that student for this academic year. With the exception of ESS1 (Earth's place in the universe), each core idea in Table 8 is represented on each of the elementary, middle, and high school blueprints.

To examine the variations in instructional plans by linkage level and core idea, cross tabulations were conducted and the chi-square difference test was computed. Selected results are presented (for core idea PS1: Matter and its interactions) that are representative of observed trends across core ideas. Overall, teachers of students who took initial level testlets were more likely to report no plans for instruction in a core idea than teachers of students who took precursor or target linkage level testlets. Teachers of students who took target level testlets were more likely to report higher numbers of hours of instruction planned for a core idea than teachers of students who took precursor or initial level testlets. ($\chi^2(4) = 27.2$, $p = .000$)

Table 10

Instructional Plans for PS1 by Linkage Level

	None	1 - 10 hours	11 or more hours	Total
Initial	88 (28%)	170 (55%)	53 (17%)	311
Precursor	33 (20%)	96 (59%)	35 (21%)	164
Target	50 (13%)	232 (61%)	99 (26%)	381

The data were examined to look for trends by grade band. Overall, teachers of high school students were more likely to report no plans for science instruction for their student for that academic year than teachers of elementary or middle school students. Teachers of middle school students reported more hours of planned science instruction than teachers of elementary or high school students. ($\chi^2(4) = 20.9$, $p = .000$)

Table 11

Instructional Plans for PS1 by Grade Band

	None	1 - 10 hours	11 or more hours	Total
Elementary	44 (18%)	151 (62%)	47 (19%)	242
Middle	43 (14%)	180 (59%)	82 (27%)	305
High	84 (27%)	167 (54%)	58 (19%)	309

Teachers were asked about which of the science and engineering practices (from the *Framework)* their student would use during science instruction during the academic year. Teachers responded yes or no to each of eight practices.

Table 12

Teachers Reporting Student Use of Science and Engineering Practices (n = 871)

Practice	All students (n = 871)	Initial (n=317)	Precursor (n=164)	Target (n=390)
Asking questions and Defining Problems	81%	66%	90%	89%
Planning and Conducting Investigations	58%	50%	59%	65%
Obtaining, Evaluating, and Communicating Information	57%	50%	57%	63%
Analyzing Data	57%	40%	57%	71%
Developing and Using Models	56%	48%	53%	63%
Using Math and Computational Thinking	42%	26%	39%	63%
Constructing Explanations and Designing Solutions	29%	17%	26%	39%
Engaging in Argument from Evidence	19%	9%	17%	29%

Asking questions was the most often reported science practice (81%), followed by planning and conducting investigations (58%). The least often reported practices were evidence-based argumentation (19%) and constructing explanations (29%). With the exception of the practice of asking questions, a pattern of increasing percentages from the initial to the target linkage level was noted.

Conclusions

A set of accessible alternate science content standards was created as well as assessment items for each standard. The set of steps that were taken to create the alternate standards allowed the challenges of appropriate breadth and depth for SWSCD and fidelity of science content representation to be met, as documented by evidence gathered during internal, external, and SEA reviews. Reviews of the alternate content standards by science content experts and special education experts support that the standards represent the science content and are appropriate for SWSCD.

Accessible science alternate assessment testlets and items were created. The use of an established test development process allowed the creation of pedagogically relevant and accessible content, as documented by evidence gathered during internal and external reviews. Testlets were written by teachers who had experience with science content and/or SWSCD based on guidelines in Essential Element Concept Maps, which increased pedagogical relevance of testlets. These testlets were reviewed and revised until accessibility, content, and bias and accessibility guidelines were met. The testlets were reviewed by a panel of external experts, with a rejection rate for items of 1.7% (Tables 3 and 4). The low rate of item rejection indicated the claim that the reviewers felt that the assessment items met guidelines. Student performance data provided further evidence of accessibility. Average p-values showed that 85% of pilot test items and 73% of field test items met the threshold of .35, indicating that students selected the correct answers at a rate greater than would be expected by chance. These p-values support the claim that students were able to understand and respond to test items (e.g., items are accessible).

Some patterns in the data led to further investigation and implementation of changes to testlet design that improved accessibility and student performance. For example, a pattern was noted in the distribution of item flags by linkage level in the pilot data; 71% of pilot item flags were at the precursor linkage level. The precursor items were examined to determine the reason for the low p-values. Item-level data was considered in combination with cognitive characteristics in the population of SWSCD and the linkage level content to determine possible causes for differences in item difficulty. Linkage level descriptors at the precursor level asked students to use complex skills, such as developing models and making claims that are supported by evidence, in combination with facts retrieved from long-term memory, and reductions in long-term memory are a concern for SWSCD. The team hypothesized that these precursor level testlets could be made more accessible, while still assessing the skills that are described by the linkage level if more context was provided to students. The team believed that the provision of additional context could compensate for limitations of long-term memory. To provide this context, testlets were revised to include science stories that provided background information and activated students' prior knowledge. Data from the field test showed that all of the revised testlets had higher p-values than the previous versions. Of the five rejected precursor level testlets that were rewritten to improve accessibility, all five had higher p-values in the field test than in the pilot. The field test data showed that four of the five testlets had no items flags, while one testlet had 2 item flags. The increase in p-value occurred despite lower average p-values across domains and linkage levels for the field test (compared to the pilot). This finding supports that the iterative process built on the Evidence-Centered Design and Universal Design for Learning frameworks enabled modifications to test design that increased accessibility.

Some patterns in the data revealed differences between the groups of students who took the pilot and the field test. For example, differences were noted in the patterns of item flags for pilot and field test data. Overall, the percentage of flagged items increased from 15% for the pilot to 27% for the field test. The higher percentage of items flagged for the field test is believed to be related to differences in the characteristics of students who were administered initial-level testlets. Analysis of the item response data for initial level items showed higher selection rates of the “no response” and “attends to other stimuli” responses, rather than responses that were related to the science content tested, during the field test than during the pilot, meaning that more of the students with the lowest expressive communication abilities were not able to attend to or respond to test items. This difference in response rates between the field test and the pilot indicates that the characteristics of the field test participants may have been considerably different than the pilot participants and the initial-level items may have been too challenging for these students. Another difference between the field test and the pilot was that 58% of field test item flags were at the initial linkage level, compared to 8% of pilot item flags.

The higher rate of field test item flags supports the conclusion that initial level testlets may have been too difficult for the students with the lowest expressive communication abilities who participated in the field test.

Possible causes for this difference could include increased difficulty of field test initial level testlets compared to the pilot initial level testlets or differences in the abilities of the students who were assessed. The formats and content of initial level testlets were consistent from the pilot to the field test. Therefore, differences in actual testlet difficulty were likely to be small and differences in the abilities of the students are a more likely explanation. Further, the sample size was more than three times larger for the field test and this sample may better represent the population of SWSCD, particularly the students who have the lowest expressive communication abilities. To examine the group of initial level testlets, we looked at the average p-values by linkage level (Table 7). In each science domain, the ranges of average p-values for initial level testlets had lower minima and lower maxima than the ranges for precursor and target level testlets. Furthermore, the maximum p-values for the initial linkage level were more than 0.2 lower than the maxima for other linkage levels in all three content domains. After the field test, initial level items were examined for possible causes of low p-values (e.g., technical inaccuracies, ambiguity of wording, or text complexity) and revisions to items were made. More research is needed on initial linkage level test takers, who are the students with the lowest expressive communication abilities.

This finding of high difficulty for the students with the lowest expressive communication abilities was not entirely unexpected because of the design differences in the DLM Essential Elements for science and the Essential Elements for English language arts and mathematics. The DLM alternate assessments in English language arts and mathematics each have five linkage levels to provide maximum accessibility for the wide range of characteristics of SWSCD, while the DLM Science Essential Elements have three linkage levels. This difference in the number of linkage levels occurred because the DLM ELA and math assessments are based on a learning map that was developed prior to the Essential Elements. The learning map enabled the identification of the milestones en route to academic targets based on research about how students learn. However, a learning map for science has not yet been developed, which constrained the development of linkage levels for the science pilot (Dynamic Learning Maps, 2015). Learning map development is planned for the next phase of the DLM Science Alternate Assessment project to address this concern. Another difference between the pilot and field test was the size and nature of the samples. As both the pilot and field test samples consisted of volunteer participants, the characteristics of the sample may not match the characteristics of the population of SWSCD. However, the much larger sample size (5,663 compared to 1,606) for the field test supports the idea that the field test sample is likely to be a better representation of the population. More will be learned about the students with the lowest expressive communication abilities from the results of the operational assessment because all eligible students will be included.

The results of the teacher survey show that many students who participated in the field test were not taught science content that aligned to the new alternate content standards. Across disciplinary core ideas, 12 to 29% of teachers reported planning no hours of science instruction for that student for this academic year. However, variations in the combinations of core ideas that are included in the three blueprints (elementary, middle, and high school) may explain some teachers' plans to not teach a particular core idea. For example, ESS1 is not included on the middle school blueprint, which may explain the higher percentage of teachers who selected “none” for ESS1. However, the other seven core ideas in Table 8 are represented on each of the three blueprints (elementary, middle, and high school). Therefore, the data support that a significant portion of students may be experiencing science instruction that omits required core ideas. This is evidence of weak alignment between enacted science curricula and science assessments. The strength of this claim is limited by a relatively low response rate to the survey (31%) and the large grain size of the core ideas and the response options (e.g., none, 1 to 10 hours, 11 or more hours). On the other hand, the data show that most teachers are providing at least some science instruction (1 to 10 hours) for most students across disciplinary core ideas.

The number of hours of instruction planned varied by students’ assigned linkage level, which implies that the degree of alignment between the enacted curriculum and the alternate standards is less for certain students. Teachers of students who took initial level testlets were more likely to report no plans for instruction in a core idea than teachers of students who took precursor or target level testlets. Assignment to the linkage level was based on the student’s expressive communication skills; students with the lowest expressive communication abilities were more likely to receive no instruction in a science core idea than other students. This finding is similar to prior research by Karvonen, Flowers, and Wakeman (2013), who reported that students’ symbolic communication level had the strongest association with access to academic curricula in English language arts and mathematics.

Teachers’ reports of student use of the science and engineering practice show that most students use the skill of asking questions (81%) and the skill of planning and conducting investigations (58%). Given the long-standing emphasis in science education reform documents on science as inquiry (e.g., National Research Council, 1996), these findings are not surprising. The next set of skills that were used by most students were related to skills that were used in

English language arts (e.g., obtaining, evaluating, and communicating information) or mathematics (e.g., analyzing data, developing and using models). It is encouraging that teachers see SWSCD transferring these skills from English language arts or mathematics contexts to science. The two science and engineering practices that were used the least by students, constructing explanations and engaging in argument from evidence, were not surprising as these practices involve relatively high cognitive complexity. Expectations for these kinds of practices have historically been low for SWSCD; research on science instruction with SWSCD has focused on inquiry and is based on the prior standards framework (e.g., National Research Council, 1996) that treated *science as inquiry* as a separate content strand (Spooner et al., 2011). However, these two higher complexity practices are included in many of the DLM Science Essential Elements. Therefore, the low rates of students' uses of these practices may negatively impact student performance on the science assessment and need to be addressed through instruction.

Teachers' reports of student use of science and engineering practice varied by student linkage level. In general, the students who tested at the higher linkage levels were more likely to have had experiences with specific science and/or engineering practices. As linkage level is based on expressive communication skills, this finding is similar to the finding regarding the number of hours of instruction planned by science core idea. The students with the lowest expressive communication abilities have fewer opportunities to use the science and engineering practices. As the new alternate science standards integrate the science and engineering practices into each standard, this difference indicates a lesser degree of alignment (and access) for these students. This finding parallels the findings of Karvonen et al. (2013).

Implications

This study is important to the fields of science education and special education because studies documenting the development of next generation alternate science standards have not yet been published. We have presented the results of initial efforts to develop rigorous alternate science standards based on new general education science standards. Data from the pilot and field test provided preliminary evidence that the new alternate standards and assessments are accessible. This claim is tentative because the pilot and field test assessment data included only a voluntary portion of the population. Additional evidence from the operational assessment in spring 2016 will be examined to further substantiate this claim. Implications of this study include that the results of the alternate assessments and development of Essential Elements for science will help increase expectations for the science education of SWSCD as the enacted curricula moves toward better alignment with the Essential Elements.

The present study showed that principles of UDL, knowledge of the characteristics of SWSCD, and extant alternate standards could be combined with expert knowledge to make the science disciplinary core ideas and science and engineering practices accessible to SWSCD. In this study, a set of alternate science standards was created that was linked to the new general education science framework. Using ECD, the disciplinary core ideas and science and engineering practices were made accessible for assessment development. This study adds to the body of evidence regarding the potential for SWSCD to learn science content that is based on the *Framework*. This study has implications for equity issues related to NARST's goal of helping all learners achieve science literacy. In particular, teacher survey data indicated a greater need for professional development in science for teachers of the students with the most severe communication disabilities. SWSCD are a population that has been marginalized in science education. This study documents first efforts to make on-grade-level science content accessible to SWSCD.

References

American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (2014). *Standards for Educational and Psychological Testing*. Washington DC: AERA.

Bechard, S. (2005). Developing alternate assessments using expanded benchmarks from a nine state consensus framework in reading, writing, mathematics, and science. Paper presented at the 2005 National Council on Measurement in Education (NCME), Annual Conference, Montreal, Canada.

Bechard, S. & Sheinker, A. (2012). *Basic framework for item writers using Evidence-Centered Design (ECD)*. University of Kansas: Lawrence, KS.

Browder, D. M., Wakeman, S. Y., Flowers, C., Rickelman, R. J., Pugalee, D., & Karvonen, M. (2007). Creating access to the general curriculum with links to grade-level content for students with significant cognitive disabilities: An explication of the concept. *The Journal of Special Education*, *41*(1), 2–16. http://doi.org/10.1177/00224669070410010101

Cameto, R., & Haertel, G. (2011, March 24). Webinar on alternate assessment and evidence-centered design. SRI International: Washington D.C.

CAST, Inc. (2012). Universal Design for Learning. Retrieved from http://www.udlcenter.org/aboutudl/whatisudl

Courtade, G. R., Spooner, F., & Browder, D. M. (2007). Review of studies with students with significant cognitive disabilities which link to science standards. *Research and Practice for Persons with Severe Disabilities*, *32*(1), 43–49. http://doi.org/10.2511/rpsd.32.1.43

DeBarger, A. H., Seeratan, K., Cameto, R., Haertel, G., Knokey, A-M., & Morrison, K. (2011). Alternate assessment design–mathematics. (Technical Report 9: Implementing evidence-centered design to develop assessments for students with significant cognitive disabilities: Guidelines for creating design patterns and development specifications and exemplar task templates for mathematics). SRI International: Washington D.C.

Dynamic Learning Maps (2013, June 4). *The First Contact census student characteristics*. Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

Dynamic Learning Maps (2013, September 24). *Dynamic Learning Maps participation guidelines.* http://dynamiclearningmaps.org/sites/default/files/documents/dlm_participation_guidelines.pdf

Dynamic Learning Maps. (2014, December). *Dynamic Learning Maps essential elements for science*. Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

Dynamic Learning Maps. (2015, December). *Summary of the science Dynamic Learning Maps Alternate Assessment development process*. Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

Federal Register (December 9, 2003). Title I—Improving the academic achievement of the disadvantaged; final rule. Vol. 68, (236). U. S. Department of Education. Washington D.C. Retrieved from: https://www2.ed.gov/legislation/FedRegister/finrule/2003-4/120903a.pdf

Flowers, C., Turner, C., Herrera, B., Towles-Reeves, L., Davidson, A., & Hagge, S. (2015). Developing a large-scale assessment using components of evidence-centered design: did it work? In NCME Proceedings (pp. 1–29).

Huff, K., Steinberg, L., & Matts, T. (2010). The promises and challenges of implementing evidence-centered design in large-scale assessment. *Applied Measurement in Education*, *23*, 310–324. http://doi.org/10.1080/08957347.2010.510956

Karvonen, M., Wakeman, S. Y., Browder, D. M., Rogers, M. A. S., & Flowers, C. (2011). *Academic curriculum for students with significant cognitive disabilities*: *Special Education Teacher Perspectives a Decade After IDEA 1997*. Retrieved from ERIC database. (ED521407)

Karvonen, M., Bechard, S., & Wells-Moreaux, S. (2015). Accessibility considerations for students with significant cognitive disabilities who take computer-based alternate assessments. Paper presented at the 2015 annual meeting of the American Educational Research Association, Chicago, Illinois.

Karvonen, M., Flowers, C., & Wakeman, S. Y. (2013). Factors associated with access to the general curriculum for students with intellectual disability. *Current Issues in Education*, *16*(3), 1–19.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999, December 4). Evidence-Centered assessment design. Educational testing service. Retrieved from: http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1: 3-67. doi**:** 10.1207/S15366359MEA0101_02

National Research Council. (1996). National science education standards. Washington, DC: The National Academies Press.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

National Research Council. (2014). *Developing assessments for the next generation science standards*. Committee on developing assessments of science proficiency in K-12. Board on testing and assessment and board on science education, J. W. Pellegrino, M. R. Wilson, J. A. Koenig, J. A., & A. S. Beatty, Editors. Division of behavioral and social sciences and education. Washington, DC: The National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states.* Washington, DC: The National Academies Press.

Rogers, C. M., Thurlow, M. L., & Lazarus, S. S. (2015). *Science alternate assessments based on alternate achievement standards (AA-AAS) during school year 2014-2015.* (Synthesis Report 99). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Spooner, F., Knight, V., Browder, D. M., Jimenez, B. A., & DiBiase, W. (2011). Evaluating evidence-based practice in teaching science content to students with severe developmental disabilities. *Research and Practice for Persons with Severe Disabilities*, *36*(1-2), 62–75. http://doi.org/10.2511/rpsd.36.1-2.62

Stage, E. K., Asturias, H., Cheuk, T., Daro, P. A., & Hampton, S. B. (2013). Opportunities and challenges in Next Generation Standards. *Science*, 340, 276–277. http://doi.org/10.4324/9780203873465

U.S. Department of Education. (2004). *Standards and assessments peer review guidance: Information and examples for meeting requirements of the No Child Left Behind Act of 2001*. Washington DC: Author.

Wells-Moreaux, S., Bechard, S., & Karvonen, M. (2015). *Accessibility manual for the Dynamic Learning Maps*TM *alternate assessment, 2015-2016*. Lawrence, KS: The University of Kansas Center for Educational Testing and Evaluation.

Appendix A

DLM External Review Criteria

Content

Item

1. Alignment: The item assesses the content of the targeted node.
2. Depth of Knowledge: The level of cognitive complexity required in the item matches the DOK identified for the item.
3. Accuracy: The content of the item is technically correct (wording and graphics).
4. Correctness: The key (correct answer) is correct and distractors are incorrect. Nothing in the item cues the correct response. The distractors are not intentionally confusing.
5. Item Design: The item type is logical and appropriate for the content being assessed. Graphics support the item and do not contribute additional information to the item.

Testlet

6. Instructional Relevance: The testlet is instructionally relevant to students for whom it was written. The content of the testlet is grade level appropriate.
7. Testlet Design: For ELA, embedded items are placed within the story text at logical places and conclusion items are placed at the end.

Special Education/Accessibility

Item

1. Accessible Text: Text within item provides an appropriate level of challenge and maintains a link to grade-level content without introducing unnecessary, confusing, or distracting verbiage.
2. Accessible Graphics: Alt text is appropriate for the item. A tactile representation of the graphic would support the item in a way that is parallel to the visual graphic.
3. Accessible Item Design: The item displays correctly on the screen (all elements are visible, no scrolling required).

Testlet

4. Instructional Relevance: The testlet is instructionally relevant to students for whom it was written.
5. Intended accessibility: The testlet is accessible to students with sensory or motor impairments, as intended in the EECM.
6. Barrier-free for SWSCD: The testlet does not introduce barriers for students with (a) limited working memory, (b) communication disorders or language differences dependent on spoken English grammatical structures, or (c) limited implicit understandings of others' emotions and intentions.

Bias/Sensitivity

Item – all bias criteria

1. Fair Construct: Item does not require background knowledge or experiences outside the bounds of the targeted content.
2. Diversity: There is a fair representation of diversity in ethnicity, gender, disability, and family composition.
3. People Positive: Stereotypes are avoided. Appropriate labels are used for groups of people. People-first language is used for individuals with disabilities.
4. Bias Impact: Language used does not prevent nor advantage any group from demonstrating what they know about the measurement target.

Testlet – all sensitivity criteria

5. Sensitive Items: Testlet items are free of content that is controversial, disturbing, or emotionally charged due to issues of culture, region, gender, religion, ethnicity, socio-economic status, occupation, or current events.

6. Sensitive Text: Text is free of content that is controversial, disturbing, or emotionally charged due to issues of culture, region, gender, religion, ethnicity, socio-economic status, occupation, or current events.