SUMMARY OF THE SCIENCE DYNAMIC LEARNING MAPS® ALTERNATE ASSESSMENT DEVELOPMENT PROCESS

Technical Report #16-02

10/21/2016

1. CONTENTS

1. EXECUTIVE SUMMARY 4

2. ESSENTIAL ELEMENTS FOR SCIENCE 5

3. BLUEPRINT DEVELOPMENT 7

4. CONTENT DEVELOPMENT 9

4.A. Test and Item Design 9

4.B. Item Writer Workshops 10

4.C. Content Review Process 11

5. 2015 SPRING SCIENCE PILOT TEST 12

5.A. Purpose 12

5.B. Eligibility 12

5.C. Design 12

5.D. Participation 13

5.E. Data Review 15

6. 2015 FALL FIELD TEST 17

6.A. Purpose 17

6.B. Eligibility 17

6.C. Design 17

6.D. Participation 20

6.E. Data Review 22

6.F. Field Test Survey 24

7. CONCLUSIONS 30

Nash, B., & Bechard, S. (2016). *Summary of the Science Dynamic Learning Maps® Alternate Assessment Development Process* (Technical Report No. 16-02). Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

1. Executive Summary

The Dynamic Learning Maps® (DLM®) Alternate Assessment for science is a two-phased project designed to support students with the most significant cognitive disabilities who are learning science content standards. The first phase of the project focused on the creation of an end-of-year assessment in one elementary, middle, and high school grade, as well as an end-of-course assessment in high school biology. The parameters for defining the second phase of the DLM science project are currently under development in collaboration with DLM science state partners. This report outlines and describes the development process for Phase I of the DLM science project.

The learning standards known as the Essential Elements (EEs) for science were developed in a four-stage process from August to December of 2014 in collaboration with the science state partners. The science standards that states were using at the time were leveraged as the starting point for EE development. Several important factors influenced the development process, including the incorporation of scientific inquiry practices and the lack of a fine-grained learning map model as the starting point, the latter of which led to the decision to use three linkage levels for EE development.

In each of the middle school and high school grade spans, four collections of EEs were suggested to serve as the Phase I DLM blueprints for science. These suggested collections were based on survey ratings from science and special education experts who reviewed the EEs selected for the 2014–2016 standards framework. Because of the smaller numbers of EEs at the elementary school grade span and in high school biology, the standards framework served as the blueprint. Science state partners voted on the final collection of EEs to serve as the blueprint in December 2014.

The structure of DLM assessments is designed to enhance accessibility for students with the most significant cognitive disabilities. Testlet and item design for science followed closely that of the DLM English language arts and mathematics assessments. Two item writer workshops were conducted in January and July of 2015 for the purpose of developing content for the operational administration of science in spring 2016. A pilot test and field test were conducted during the 2015 calendar year. Results from the pilot test and field test provided useful insight for improving science testlet content and design in preparation for the 2016 operational assessment. In addition, a survey was administered during the field test to collect data on students' science academic skills, opportunity to learn science content, and overall experience with the science assessment.

The development process for the DLM science assessment was intentionally ambitious to meet the needs of the science state partners. The result is a science assessment that is accessible to students with the most significant cognitive disabilities and is based on content and standards intended to improve teaching and learning science curriculum within this population of students.

2. Essential Elements for Science

The development process for the science Dynamic Learning Maps® (DLM®) assessment system took a different approach than the English language arts and mathematics DLM assessment programs in that the science project began with the creation of the Essential Elements (EEs) for science without a contiguous map development process. The DLM EEs for science are specific statements of knowledge and skills linked to the grade-span expectations identified in the Next Generation Science Standards (NGSS) and the Framework for K-12 Science Education (National Research Council, 2012) and represent the most frequently assessed alternate standards in the initial group of seven states interested in DLM science. As such, this set of EEs addresses a small number of science standards, representing a breadth, but not depth, of coverage across the entire standards framework. (Note: NGSS codes are used to provide a general education link to the DLM EEs.) The purpose of the DLM EEs is to build a bridge from the content in the Framework for K-12 Science Education to academic expectations for students with the most significant cognitive disabilities. This version of the EEs provides content for science assessments for the first phase of the project. In the second phase of the project, the DLM Science Consortium intends to develop a learning map model based on research about how students learn science content and engage in scientific and engineering practices. Revisions will be made when the science map project is complete, at which time we anticipate the EEs will be aligned to the map, with revisions and additions as appropriate. DLM science member states will be given 1–2 years' notice of revised EEs, so educators can adjust their instruction before new assessments are delivered.

As displayed in Table 1, the DLM EEs for science were developed in a four-stage process from August to December of 2014. The first draft began with guidance from the DLM science states to develop EEs for three grade spans: elementary school (represented by grade 5 standards), middle school, and high school (including EEs appropriate for end-of-course high school biology). The DLM state partners identified cross-grade topics common among the DLM states' science standards. These topics included the domains of physical science, life science (from which the high school biology topics were identified), and earth/space science. Most states' science standards included scientific inquiry practices, typically as a separate inquiry strand that was not integrated into the core content areas. The DLM state partners selected core content for EE development that was common across states, showed strong progressions across grades, and was most important for students with the most significant cognitive disabilities to be prepared for college, career, or community life. To identify the number of standards to address by grade and domain (Table 2), a DLM science standards framework for EE development was approved.

Table 1. Timeline for the Development of the DLM Science Essential Elements.

Draft	Development	2014 Timeline
1	Essential Elements created by DLM staff and invited science experts and reviewed internally	Aug 28 – Aug 29
2	DLM state partner science and special education experts conduct in-person educator review	Oct 14 – Oct 15
3	States conduct internal reviews	Oct 27 – Nov 7
4	Final state review	Nov 18 – Dec 3

Table 2. Count of State Standards to Address in Essential Element Development.

Grade Span	Physical Science	Life Science	Earth & Space Science
Elementary School	4	2	3
Middle School	4	4	6
High School	4	5	6
High School Biology	NA	10	NA

Note: NA, not applicable.

Another important consideration for EE development concerned the number of linkage levels to include in the EEs. The DLM EEs for English language arts and mathematics contain five linkage levels. The absence of a fine-grained learning map model for science led to the decision to use fewer linkage levels. Three linkage levels were recommended for science EEs to allow for development of a science map model to support additional linkage levels that will be based on research and evidence. Table 3 shows the science linkage levels compared with the English language arts and mathematics linkage levels.

Table 3. Linkage-Level Comparison.

Content Areas	Linkage Levels				
ELA and Math	Initial	Distal Precursor	Proximal Precursor	Target	Successor
Science	Initial		Precursor	Target	NA

Note: ELA, English language arts; NA, not applicable.

The first draft of the EEs was compiled by DLM staff and then reviewed internally by an expert panel of science and special education consultants, which resulted in a second draft. The second draft was presented to representatives from each state education agency and the educators and content specialists they selected. Sixteen experts in science, as well as 17 individuals with expertise in instruction for students with the most significant cognitive disabilities, from across five states reviewed the draft documents. This review resulted in significant changes that:

- clarified the science concepts that are the essential targets for measurement.
- revised verbs to convey clear statement of what the student should do related to scientific and engineering practices.
- focused on universal access issues.
- revised the EEs to be more measurable.
- aligned the linkage levels with the Target-level EEs and across the grade span and refined Initial and Precursor levels.
- provided examples within the EE statements.

A third draft was then reviewed internally by each state, considering these guiding 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?
4. Do the linkage levels represent the learning-target content at appropriately reduced levels of breadth and depth?

A final discussion and consensus vote by participating states in December 2014 resulted in the final EEs.

3. Blueprint Development

In both the middle school and high school grade spans, four collections of Essential Elements (EEs) were suggested to serve as the Phase I science Dynamic Learning Maps® (DLM®) blueprints. These suggested collections were based on initial feedback from science and special education experts who reviewed the EEs selected for the 2014–2016 standards framework. Due to the smaller numbers of EEs at the elementary school grade span and in high school biology, the standards framework in those grades served as the blueprint.

This year-end blueprint assumes an assessment in which students take a 25- to 30-item assessment in the form of testlets containing three to four items written to assess a single EE. Despite a commitment to a blueprint that would maximize the breadth of content coverage, given the number of EEs within the framework at each grade level, it was necessary to select and weigh the EEs to meet the assessment length requirement of approximately 10 EEs per grade level. Because the framework for the elementary school level contained nine EEs and high school biology had 10 EEs, the focus of the blueprint decisions were on the middle school and high school levels. The blueprint options presented to states covered content in all three science domains, but with different emphases.

The following principles guided the development of the blueprint options.

- Use the feedback from the panel of science and special education educators to prioritize content that has the potential to maximize student growth in academic skills across grades.
- Use knowledge of academic content and instructional methods to prioritize content that is considered important by stakeholders and central to the construct.
- Prioritize content that can be applied to real-world or workplace problems.
- Maximize the breadth of coverage of EEs, given the time needed to administer an assessment to students taking alternate assessments.

The following steps outline the development process used for the Phase I science DLM blueprints.

1. Ten of the 31 educators (32%) who attended the EE review meeting in October 2014 rated the EEs via electronic survey. The educators had a range of experiences in science education ($n = 4$) or special education ($n = 6$) and represented all five science partner states:

    - Iowa, one rater
    - Missouri, three raters
    - Mississippi, two raters
    - Kansas, two raters
    - Oklahoma, two raters

    Ratings were based on three criteria using a 4-point agreeability scale (agree, somewhat agree, somewhat disagree, disagree).

    - The EE reflects a high but reasonable expectation for students with the most significant cognitive disabilities at this grade level.
    - The EE is important for learning what students will need in life after secondary education.
    - The EE is relevant to current science instruction in the classroom.

2. These ratings were compiled and used to develop the four different blueprint options for the middle school and high school grade spans. The results from the educator rating survey suggested two different blueprint options with respect to the aforementioned criteria. Two additional blueprint options also consider the EEs from a breadth-of-content perspective. These blueprint options were reviewed by states between November 24 and December 9 of 2014.

3. On December 9, 2014, a final vote by states was conducted during an in-person governance meeting. Final science DLM blueprints can be found on the DLM website (http://dynamiclearningmaps.org/).

The science EEs are arranged into the three domains, 10 core ideas, and 14 topics shown in Table 4.

Table 4. Domains, Core Ideas, and Topics in Science.

Domain	Core Idea	Topic
Physical	PS1: Matter and Its Interactions	PS1.A: Structure and Properties of Matter
	PS2: Motion and Stability: Forces and Interactions	PS2.A: Forces and Motion
		PS2.B: Types of Interactions
	PS3: Energy	PS3.B: Conservation of Energy and Energy Transfer
		PS3.D: Energy in Chemical Processes and Everyday Life
Life	LS1: From Molecules to Organisms: Structure and Processes	LS1.A: Structure and Function
		LS1.B: Growth and Development of Organisms
		LS1.C: Organization for Matter and Energy Flow in Organisms
	LS2: Ecosystems: Interactions, Energy, and Dynamics	LS2.A: Interdependent Relationships in Ecosystems
	LS3: Heredity: Inheritance and Variation of Traits	LS3.B: Variation of Traits
	LS4: Biological Evolution: Unity and Diversity	LS4.C: Adaptation
Earth & Space	ESS1: Earth's Place in the Universe	ESS1.B: Earth and the Solar System
	ESS2: Earth's Systems	ESS2.A: Earth Materials and Systems
		ESS2.D: Weather and Climate
	ESS3: Earth and Human Activity	ESS3.A: Natural Resources
		ESS3.C: Human Impacts on Earth Systems

4. Content Development

4.A. Test and Item Design

The structure of Dynamic Learning Maps® (DLM®) assessments is designed to enhance accessibility for students with the most significant cognitive disabilities. Assessment items are grouped into instructionally relevant testlets that model classroom instructional activities and thus provide a familiar context to the student. The testlet is a package of three to four assessment items centered on a learning target that begins with an engagement activity.

Each science testlet assesses a single Essential Element at a single linkage level. Similar to English language arts and mathematics assessment design, science item specifications are described in the Essential Element Concept Maps (EECMs). EECMs use principles of Evidence-Centered Design and Universal Design for Learning to define science content specifications for assessment. They provide a guide to the item writer on how to develop accessible testlets and related testlet sets. Each EECM defines the content framework of a particular Target-level Essential Element with three levels of complexity and identifies key vocabulary at each level. The EECM also describes and defines common misconceptions, common questions to ask, and prerequisite and requisite skills. It also identifies accessibility issues related to particular concepts and tasks.

There are two types of testlets in science—teacher-administered and computer-administered. The teacher-administered testlet type is used at the Initial level in science. In teacher-administered testlets, the teacher engages the student and familiarizes the student with picture response cards or objects. The teacher uses a script to present the pictures or objects and asks the student a question that requires the student to choose one of the pictures or objects. The picture response cards are provided in PDF format for the teacher to print before administering the testlet. The student responds by selecting the appropriate picture response card or object. The teacher enters the student's responses into the KITE™ system.

Computer-administered testlets are used at the Precursor and Target levels in science. The student completes computer-administered testlets independently. Computer-administered testlets begin with an engagement activity. The engagement activity is designed to activate prior knowledge, motivate the student, and provide a context for the items in the testlet.

Science stories are often used as engagement activities. Science stories are thought to help students by minimizing the dependence on long-term memory of facts. Factual information or observations are provided, and the items ask the student to engage in a science or engineering practice that involves that content. Guidelines from the DLM Style Guide for English language arts and mathematics have been applied to science stories. Examples of science stories include: descriptions of a student engaging in a science experiment and descriptions of science processes or activities. Science stories do not teach science content.

After the engagement activity, students answer three to four items. Any data needed for the student to answer the question are presented on the item screen to remove the need to go back and forth between screens.

4.B. Item Writer Workshops

Educators completed applications and were selected based on a rubric that considered special education experience and content experience. Forty-nine educators were selected to participate in the first item writer workshop (IWW) held in January 2015. Before the IWW, educators completed an online training course with modules and quizzes. The course provided content about the population taking DLM assessments, the DLM assessment program, principles of item writing, and the test-development process. As a culminating assignment, the educators drafted a testlet.

Educators then participated in a day and a half of in-person training that reviewed what was learned in the online training and provided feedback on their draft testlets along with more details on item writing. Three days were spent writing items. Item writers were assigned linkage levels to write for based on their experience and expertise. Item writers were grouped by grade band (elementary school, middle school, high school, and high school biology) for the purpose of collaboration in their work.

In July 2015, a second IWW was held. Nineteen educators completed another online training course with modules and quizzes that built on what was learned in the previous IWW. The additional training focused on information about science and engineering practices. Four days were spent writing items. Item writers were assigned linkage levels to write for based on their experience and expertise. Item writers were again grouped by grade band (elementary school, middle school, and high school) for the purpose of collaboration in their work.

4.C. Content Review Process

The review process for the science items and testlets followed the existing 26-step workflow developed from the English language arts and mathematics projects. Descriptions of two of these steps, the content and special education (CSPED) review and the external review processes, are provided in the following paragraphs.

The CSPED review process for science was conducted by educators who either possess expertise in science content or who teach students with the most significant cognitive disabilities. The reviewers completed training on the DLM assessment program and the review criteria. The CSPED review consists of several types of review: adherence to DLM style guidelines, quality of science content, accessibility issues, and bias concerns. Testlet content was 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 were 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. CSPED reviewers entered evaluative information into an online survey or recommended revisions to testlets. Testlets that did not meet criteria were revised.

The external review process for science was conducted by educators who either possess expertise in science content or who teach students with the most 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 assessment 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.

5. 2015 Spring Science Pilot Test

5.A. *Purpose*

The spring 2015 science pilot testing window was from April 22 through June 5, 2015, and included Iowa, Kansas, Missouri, and Oklahoma. States were able to select their own windows within the consortium-wide window if needed. Results from the spring pilot tests were used for research and development purposes only and were not reported this year.

The purposes of the pilot test were to:

- evaluate the new science testlet content
- develop a method for using student information to determine which linkage level of testlet to deliver
- gather feedback from educators about testlet content with respect to students' opportunity to learn, perceived difficulty, and general testlet structure

5.B. *Eligibility*

To be eligible for the Dynamic Learning Maps® (DLM®) science pilot test, students needed to be in grades 3-12, have the most significant cognitive disabilities, and be eligible for their state's current alternate assessment based on alternate achievement standards. Students were enrolled based on their typical science grade within the appropriate grade band. States were encouraged to implement the same eligibility guidelines used for alternate assessment participation in English language arts and mathematics for the science pilot test. All computer-administered testlets included read-aloud capability; however, the pilot test was not specifically designed for students who are blind or have visual impairments.

5.C. *Design*

The linkage level was chosen for each student based on information from the student's First Contact survey. The First Contact survey is a survey of learner characteristics that goes beyond basic demographics. This survey covers a variety of areas, including communication, academic skills, and attention. All questions must be completed because the system assigns each student to a specific testlet linkage level based on educator responses. For the spring pilot test, only the expressive communication questions were used for testlet linkage-level assignment. This assignment was the same for all administered testlets. That is, students received all testlets at only one linkage level.

All students were assigned testlets that covered the entire blueprint. During the spring pilot test, students received a fixed-form assessment that contained either nine or 10 testlets (i.e., 10 for high school biology and nine for all other grade spans) at the same linkage level. Each testlet included three to four items related to one Essential Element in the blueprint.

5.D. PARTICIPATION

The spring 2015 science pilot testing was conducted from April 22 through June 5, 2015. A total of 1,605 students from four states—Iowa, Kansas, Missouri, and Oklahoma—completed assessments. The total number of participants by grade span are presented in Table 5 and indicate that 36% of students were in elementary school (grades 3-5), 35% were in middle school (grades 6-8), and 29% were in high school (grades 9-12).

Table 5. Number of Participants in the Spring 2015 Science Pilot Test by Grade Span.

Grade Span	Students
Elementary	577
Middle School	562
High School	448
Biology	20
Total	**1,607**

Note: Oklahoma administers an end-of-course biology assessment at the high school grade span.

Table 6. Demographic Summary of Students Participating in the Spring 2015 Science Pilot Test.

Demographic	Number	Percentage
Gender		
Female	568	35.3
Male	1,033	64.3
Missing	6	0.4
Primary Disability		
Autism	254	15.8
Deaf-Blindness	1	0.1
Developmental Delay	11	0.7
Documented Disability	0	0.0
Emotional Disturbance	7	0.4
Hearing Impairment	9	0.6
Intellectual Disability	435	27.1
Multiple Disabilities	90	5.6
No Disability	0	0.0
Orthopedic Impairment	4	0.2
Other Health Impairment	98	6.1
Specific Learning Disability	62	3.9
Speech or Language Impairment	6	0.4
Traumatic Brain Injury	11	0.7
Visual Impairment	5	0.3
Missing	614	38.2
Race		
White	1,182	73.6
African American	169	10.5
Asian	55	3.4
American Indian	95	5.9
Two or More Races	29	1.8
Native Hawaiian or Pacific Islander	3	0.2
Missing	74	4.6
Hispanic Ethnicity		
No	875	54.4
Yes	156	9.7
Missing	576	35.8
ESOL Participation		
Not ESOL eligible/monitored student	1,528	95.1
ESOL eligible/monitored student	79	4.9

Note: ESOL, English Speakers of Other Languages program.

In the 2015 spring science pilot testing, each Essential Element and linkage level was assessed by a single testlet. In total, 102 testlets were piloted, each consisting of three or four items. The number of testlets by grade span is presented in Table 7.

Table 7. Number of Testlets by Grade Span.

Linkage Level	Elementary School	Middle School	High School	High School Biology*	Total
Initial	9	9	9	10	37
Precursor	9	9	9	10	37
Target	9	9	9	10	37
Total	**27**	**27**	**27**	**30**	**111**

** Note: There are seven unique EEs on the high school biology blueprint; three of the EEs overlap with the high school life science blueprint.*

5.E. Data Review

Following the pilot test, item statistics were computed for all items and testlets. Specifically, the percentage correct was calculated for every item, and a z score was calculated for every item to reflect the standardized difference between the item's percentage correct and the weighted average percentage correct for items within the testlet. Using these item statistics, items were flagged for further review.

Items were flagged for review if they met either of the following statistical criteria:

- too challenging: less than 35% correct
- significantly easier or harder than other items within the same testlet (standardized difference): any percentage correct greater than 2 standard errors from the mean percentage correct

Data reviews were conducted by the content team. Flagging criteria served as one source of evidence for the content teams in evaluating item quality. Final judgments were content based. The team reviewed all items that were flagged and had a sample size of at least 20 cases. Due to low participation ($n < 20$) in high school biology, item statistics were not calculated; rather, all biology items were examined using insights gained from the review of other items.

Flagged items were discussed, and possible causes for the flag were considered. Group consensus was used to make item-level decisions. The content team's options included: (1) make no change to the item, (2) identify concerns requiring item modification that are clearly identifiable and can improve the item, (3) identify concerns requiring item modifications that are not clearly identifiable but should be considered because the item's content is worth preserving, or (4) reject item because it is not worth revising. After item-level decisions were made, testlets for items assigned to options 3 or 4 were evaluated to determine if the testlet would be retained or rejected.

Using the criteria outlined above, Table 8 reports the percentage of flagged items from the total number of eligible items for each grade span. High school biology items did not meet sample-size requirements and therefore could not be included in the data review. However, all of these items were reviewed by the content team after the completion of the data review, and decisions were made based on lessons learned from other items and testlets. Table 9 displays the decisions that were made by the content team as a result of the data review and additional review of high school biology items.

Table 8. Item Flags for Content Administered During the 2015 Science Spring Pilot Test.

Grade Span	Number of Flagged Items	Number of Eligible Items	Percentage Flagged
Elementary School	13	83	15.7
Middle School	12	83	14.5
High School	13	85	15.3
High School Biology	NA	NA	NA
Total	**38**	**251**	**15.1**

Note: Sample sizes were less than 20 for all high school biology testlets.

Table 9. Content Team Response to Item Flags for the 2015 Science Spring Pilot Test.

Grade Span	Number of Reviewed Items	Accept	Revise	Reject
Elementary School	13	0	6	7
Middle School	12	1	5	6
High School	13	1	6	6
High School Biology	27	20	6	1
Total	**65**	**22**	**23**	**20**
Percentage of Total		**33.8**	**35.4**	**30.8**

Note: Sample sizes <20 for all high school biology testlets; therefore, all items were included in the content review.

Of the 38 flagged items, 27 (71%) were at the Precursor level. This finding led the content team to examine the Precursor testlets to determine possible causes for more flags at the Precursor level. Linkage-level descriptors at the Precursor level ask students to use more complex skills, such as developing models and making claims that are supported by evidence. The content team decided that the difficulty of Precursor-level testlets could be reduced, while still assessing the skills that are described by the linkage level, if more context was provided to students. In revised testlets, science stories were used to provide context and to activate students' prior knowledge. Revisions to high school biology items generally involved accessibility of tables and graphs and consistency of format and presentation. All items and testlets that were revised were included in the fall 2015 field test.

6. 2015 Fall Field Test

6.A. Purpose

The fall 2015 science field testing window was from November 9 through December 9, 2015. Participating states included Iowa, Kansas, Missouri, Oklahoma, West Virginia, and Mississippi. Results from the fall field tests were used for research and development purposes and to contribute to the data for the spring 2016 model parameter calibrations. A science survey was also administered to a sample of field test participants and results were used for research and development purposes.

The purposes of the fall field test were to:

- evaluate new and edited science testlet content.
- develop and evaluate a method for assigning an appropriate first testlet based on students' science academic skills.
- create relationships between linkage levels for modeling purposes.
- evaluate students' opportunity to learn science content and practices, obtain information on students' science academic skills, and to receive feedback concerning students' overall experience using the Dynamic Learning Maps® (DLM®) science assessment system.

This section presents findings related to the new and edited science content and results from the field test survey. Separate reports are planned for the science First Contact evaluation and modeling.

6.B. Eligibility

The eligibility criteria for the fall field test were the same as the pilot test's eligibility criteria, with one exception: the 2015 fall field test was also designed for students who are blind or have visual impairments.

6.C. Design

The linkage level chosen for each student was based on information from the students' First Contact survey. For the fall field test, again, only the expressive-communication questions were used for testlet linkage-level assignment. This assignment placed students into one of three science linkage levels. For each linkage level, several fixed-form assessments were available for administration, and each assessment contained two testlets at the assigned linkage level and one testlet at an adjacent linkage level. That is, all students received two testlets at their assigned linkage level and one testlet at a higher or lower linkage level. For high school biology, a fixed-form assessment contained seven testlets; four were at the assigned linkage level and three were at an adjacent level.

Testlets did not cover the entire blueprint. Each testlet included three to four items related to one Essential Element (EE) in the blueprint.

A combination of new content developed at the July item writing workshop and revised content from the spring pilot served as the content field-tested in the fall. One testlet at each EE and linkage level was field-tested with the goal of developing an item bank that was two items deep for every EE and linkage level.

Table 10 displays an example of the matrix design for one grade span and science domain employed for the 2015 science fall field test.

Table 10. 2015 Science Fall Field Test Sampling Design Example.

Linkage Level Assigned by First Contact	**Form**	**Initial**			**Precursor**			**Target**		
		EE1	**EE2**	**EE3**	**EE1**	**EE2**	**EE3**	**EE1**	**EE2**	**EE3**
Initial	1	X	X		X					
	2		X	X		X				
	3	X		X			X			
Precursor	4	X			X	X				
	5		X			X	X			
	6			X	X		X			
	7				X	X		X		
	8					X	X		X	
	9				X		X			X
Target	10				X			X	X	
	11					X			X	X
	12						X	X		X

Note. EE# delineates one of the three EEs available within one science domain.

In the 2015 fall field test, each EE and linkage level was assessed by a single testlet. In total, 111 testlets were tested, each consisting of three or four items. The number of testlets by grade span is presented in Table 11.

Table 11. Number of Testlets by Grade Span.

Linkage Level	Elementary School	Middle School	High School	High School Biology*	Total
Initial	9	9	9	10	37
Precursor	9	9	9	10	37
Target	9	9	9	10	37
Total	**27**	**27**	**27**	**30**	**111**

** Note: There are seven unique EEs on the high school biology blueprint; three of the EEs overlap with the high school blueprint.*

6.D. PARTICIPATION

A total of 5,613 students participated in the 2015 field test. The total number of participants by grade span is presented in Table 12. The number of participants by grade span shows that 31% of students were in elementary school (grades 3-5), 33% were in middle school (grades 6-8), 36% were in high school or the end-of-instruction biology course (grades 9-12).

Table 12. Number of Participants in the Fall 2015 Science Field Test by Grade Span.

Grade Span	Students
Elementary	1,718
Middle School	1,869
High School	1,958
Biology	68
Total	**5,613**

Note: Oklahoma administers an end-of-instruction Biology assessment at the high school grade span.

Table 13 displays the demographic summary for the field test participants by gender, primary disability, comprehensive race, Hispanic ethnicity, and participation in English Speakers of Other Languages (ESOL) programs. Approximately 65% of participants were male students, 69% did not indicate a primary disability, 74% were white, 94% were not of Hispanic ethnicity, and 98% of students were not eligible or monitored for ESOL participation. Please note that the primary disability field is not currently a required field for educators to complete. Also note that braille and large print were not available for the field test. However, students who had a visual impairment indicated in their Personal Needs Profile were assigned testlets that were specifically designed to remove any visual barriers.

Table 13. Demographic Summary of Students Participating in the Fall 2015 Science Field Test.

Demographic	Number	Percentage
Gender		
Female	1,978	35.24
Male	3,635	64.76
Primary Disability		
Autism	372	6.63
Deaf-Blindness	3	0.05
Developmental Delay	3	0.05
Documented Disability	165	2.94
Emotional Disturbance	21	0.37
Hearing Impairment	1	0.02
Intellectual Disability	615	10.96
Multiple Disabilities	156	2.78
No Disability	2	0.04
Orthopedic Impairment	16	0.29
Other Health Impairment	86	1.53
Specific Learning Disability	20	0.36
Speech or Language Impairment	8	0.14
Traumatic Brain Injury	13	0.23
Visual Impairment	3	0.05
Missing	4,129	73.56
Race		
White	4,176	74.40
African American	1,056	18.81
Asian	114	2.03
American Indian	95	1.69
Alaska Native	19	0.34
Two or More Races	126	2.24
Native Hawaiian or Pacific Islander	16	0.29
Missing	11	0.20
Hispanic Ethnicity		
No	5,288	94.21
Yes	322	5.74
ESOL Participation		
Not ESOL eligible/monitored student	5,508	98.13
ESOL eligible/monitored student	105	1.87

Note: ESOL, English Speakers of Other Languages program.

6.E. Data Review

Following the field test, item statistics were computed for all items and testlets, and the same process and criteria for data review was followed.

Tables 14 and 15 display the results of the data review. The number of items flagged out of the number eligible indicates that approximately 26% of eligible items were flagged for further review based on item performance. Of those reviewed by the content team, 20% were not revised, 68% were revised, and almost 11% were rejected from the item pool.

Table 14. Item Flags for Content Administered During the 2015 Science Fall Field Test.

Grade Span	Number of Flagged Items	Number of Eligible Items	Percentage Flagged
Elementary School	19	81	23.50
Middle School	26	85	31.00
High School	29	90	28.90
High School Biology*	0	23	0.00
Total	**74**	**279**	**26.50**

Note: Sample sizes were <20 for all Initial- and Precursor-level high school biology testlets.

Table 15. Content Team Response to Item Flags for the 2015 Science Fall Field Test.

Grade Span	Accept	Revise	Reject
Elementary School	5	14	0
Middle School	2	19	5
High School	8	17	3
High School Biology	NA	NA	NA
Total	**15**	**50**	**8**
Percentage of Total	**20.30**	**68.0**	**10.80**

Note: NA, not applicable.

Based on the findings from the data review, the content team determined that the decision from the pilot test results to add context through science stories, particularly at the Precursor linkage level, was effective at improving student performance. Recommendations were also made to reduce the text complexity of the Initial-level testlets, particularly in the test-administrator directives to the student (e.g., "Show me the one that changes from a solid to a liquid."). These Initial-level testlets were able to be revised to be more concise and clear. In some cases, unnecessarily difficult vocabulary was removed.

6.F. Field Test Survey

As part of the field test administration, a survey was also administered to educators to obtain feedback on their students' science academic skills, opportunity to learn science content, and overall experience with the science field test. Students were randomly selected and enrolled in the survey. If a student was enrolled in the survey, the rostered educator would complete the survey questions about that student. Of the 2,037 students enrolled in the survey, 837 had completed surveys, for a response rate of approximately 41%.

Table 16 displays the demographic data for the students whose educators responded to the fall field test survey. Included in Table 16 are reported numbers and percentages of gender, disability, race, ethnicity, and ESOL participation.

There were three sections in the survey. The first section asked educators to indicate how consistently each student used specific science academic skills during science instruction. Table 17 shows the number and percentage of students who demonstrated each skill on a scale of never to consistently. Most students could sort objects by common properties, identify similarities and differences, and recognize patterns 21–50% of the time. Conversely, most students never or almost never compared initial and final conditions to determine change, used data to answer questions, identified cause-and-effect relationships, identified evidence to support a claim, or used diagrams to explain phenomena.

Table 16. Demographic Summary of Students Whose Educators Participated in the Science Field Test Survey.

Demographic	Number	Percentage
Gender		
Female	281	33.57
Male	556	66.43
Primary Disability		
Autism	26	3.11
Deaf-Blindness	0	0.00
Developmental Delay	1	0.12
Documented Disability	35	4.18
Emotional Disturbance	2	0.24
Hearing Impairment	0	0.00
Intellectual Disability	47	5.62
Multiple Disabilities	14	1.67
No Disability	0	0.00
Orthopedic Impairment	1	0.12
Other Health Impairment	3	0.36
Specific Learning Disability	1	0.12
Speech or Language Impairment	0	0.00
Traumatic Brain Injury	1	0.12
Visual Impairment	0	0.00
Missing	706	84.35
Race		
White	650	77.66
African American	121	14.46
Asian	18	2.15
American Indian	28	3.35
Two or More Races	15	1.79
Native Hawaiian or Pacific Islander	2	0.24
Missing	3	0.36
Hispanic Ethnicity		
No	777	92.83
Yes	60	7.17
ESOL Participation		
Not ESOL eligible/monitored student	812	97.01
ESOL eligible/monitored student	25	2.99

Note: ESOL, English Speakers of Other Languages program.

Table 17. Perceived Consistency of Student Skill During Science Instruction.

Skill	Never or Almost Never (0–20%)		Occasionally (21–50%)		Frequently (51–80%)		Consistently (81–100%)		Missing	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Sort objects or materials by common properties	226	27.5	240	29.2	232	28.2	124	15.1	15	1.8
Identify similarities and differences	310	37.9	306	37.4	162	19.8	41	5.0	18	2.2
Recognize patterns	319	38.9	295	35.9	154	18.8	53	6.6	16	1.9
Compare initial and final conditions to determine if something changed	462	56.1	245	29.8	99	12.0	17	2.1	14	1.7
Use data to answer questions	482	58.6	239	29.0	89	10.8	12	1.6	14	1.7
Identify cause-and-effect relationships	489	59.6	245	29.9	72	8.8	14	1.7	17	2.0
Identify evidence that supports a claim	564	68.8	198	24.2	51	6.2	7	0.9	17	2.0
Use diagrams to explain phenomena	583	71.3	175	21.4	51	6.2	9	1.1	19	2.3

The second section of the survey asked educators to indicate the average number of hours they either spent on instruction or planned for instruction of science curriculum during the 2015–2016 school year. Table 18 shows the number and percentage of educators by average number of hours spent on instruction or planned for instruction of science content within 10 topics. The number and percentage of educators who either spent time engaging their students or planned to engage their students in science practices during science instruction are displayed in Table 19. Please note that educators could select more than one science practice.

Overall, the majority of educators spent, on average, 1–10 hours of instruction on most science topics during the 2015–2016 school year. Approximately 40% of educators did not spend any instructional time on the topics of heredity or biological evolution. The science practice that educators engaged their students in most frequently was asking questions and defining problems, and the least frequently used practice was engaging in argument from evidence.

Table 18. Average Number of Hours Spent Instructing Science Topics.

Science Topic	None		1–10 hours		11–20 hours		21–30 hours		More than 30 hours		Missing	
	n	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**	***n***	**%**
Matter and Its Interactions	166	19.8	481	57.5	119	14.2	21	2.5	37	4.4	13	1.6
Motion and Stability: Forces and Interactions	202	24.1	475	56.8	106	12.7	21	2.5	21	2.5	12	1.4
Energy	162	19.4	495	59.1	116	13.9	28	3.4	23	2.8	13	1.6
From Molecules to Organisms: Structure and Processes	239	28.6	433	51.7	112	13.4	20	2.4	19	2.3	14	1.7
Ecosystems: Interactions, Energy, and Dynamics	214	25.6	423	50.5	133	15.9	40	4.8	14	1.7	13	1.6
Heredity: Inheritance and Variation of Traits	359	42.9	366	43.7	75	9.0	13	1.6	12	1.4	12	1.4
Biological Evolution: Unity and Diversity	333	39.8	387	46.2	76	9.1	11	1.3	15	1.8	15	1.8
Earth's Place in the Universe	167	20.0	460	55.0	135	16.1	42	5.0	20	2.4	13	1.6
Earth's Systems	107	12.8	475	56.8	160	19.1	50	6.0	32	3.8	13	1.6
Earth and Human Activity	160	19.1	482	57.6	126	15.1	38	4.5	18	2.2	13	1.6

Table 19. Science Practices in Which the Student Was Engaged (N = 837).

Science Practice	*n*	%
Asking questions and defining problems	680	81.2
Planning and carrying out investigations	497	59.4
Analyzing and interpreting data	480	57.4
Obtaining, evaluating, and communicating information	477	57.0
Developing and using models	465	55.6
Using mathematics and computational thinking	348	41.6
Constructing explanations and designing solutions	241	28.8
Engaging in argument from evidence	160	19.1

Note: Educators were allowed to select multiple responses.

The third section of the survey asked educators to respond to questions regarding their students' experiences using the DLM science assessment system. Specifically, educators were asked about Personal Needs Profile (PNP) features that met their students' accessibility needs and about factors that negatively and positively affected their students' experiences using the system. Tables 20–22 summarize responses to these questions.

Results indicated that almost 60% of students used the synthetic read aloud with sentence highlighting feature to meet their needs, whereas only about 8% used a switch. With respect to factors that affected students' assessment experiences, most educators thought that their students had not yet learned the topics covered by the assessments, the items did not correspond to their students' true knowledge and skills, and that the engagement activities and vocabulary were too complex, which negatively affected the experience. Conversely, the majority of educators believed that the instructions for the test administrator were clear and that this positively affected students' experiences.

Table 20. Personal Needs Profile (PNP) Features That Met Students' Accessibility Needs (N = 837).

Accessibility Features	*n*	%
Synthetic read aloud with sentence highlighting (text to speech)	495	59.1
Magnification	99	11.8
Other display changes (color contrast, reverse contrast)	97	11.6
Switch (single-switch or two-switch system)	66	7.9

Note: Educators were allowed to select multiple responses.

Table 21. Factors That Negatively Affected Students' Assessment Experience (N = 837).

Negative Experiences	*n*	%
Student has not yet learned the topics covered by the assessments	523	62.5
The items did not correspond to the student's true knowledge, skills, and abilities	447	53.4
Complexity of the engagement activity	437	52.2
The vocabulary used in the testlets was too complex	418	49.9
Student has had limited experience with a computer	141	16.9
Too many testlets	126	15.1
Use of video as the engagement activity	65	7.8
Instructions to the test administrator were not clear	54	6.5

Note: Educators were allowed to select multiple responses.

Table 22. Factors That Positively Affected Students' Assessment Experience (N = 837).

Positive Experiences	*n*	%
Clear instructions to the test administrator	440	52.6
Use of video as the engagement activity	273	32.6
Quality of the engagement activity	269	32.1
This student was instructed in the areas covered by the assessments	190	22.7
The student was familiar with the vocabulary used in the testlets	189	22.6
The items corresponded to the student's true knowledge, skills, and abilities	148	17.7
Intuitiveness of the assessment system	109	13.0

Note: Educators were allowed to select multiple responses.

7. Conclusions

The development process for the Dynamic Learning Maps® (DLM®) science assessment was intentionally ambitious to meet the needs of the science state partners. The result is a science assessment that is accessible to students with the most significant cognitive disabilities and is based on content and standards that are intended to improve teaching and learning science curriculum within this population. The DLM science program was able to leverage much of what was already built and learned from the English language arts and mathematics assessment programs in terms of administration systems, accessibility features, content development and review processes, and testlet and item design. The science pilot test and field test data provided useful information for nuances specific to assessing science content, such as providing additional context within testlets to reduce cognitive load and reducing text complexity at the lowest linkage level. Finally, engagement activities for science evolved throughout the development process into more instructionally relevant science stories that guide students through familiar science activities and experiments. These science stories are intended to draw on students' prior experiences and knowledge and to provide context for assessing relevant science skills.

Findings from the field test survey indicated that there is opportunity for improvement in providing access to science curriculum to students with the most significant cognitive disabilities. With increased opportunities to learn science content and engage in scientific practices, it is anticipated that students will be better able to demonstrate science academic skills, such as using data to answer questions and identifying cause-and-effect relationships. Finally, ongoing research and development initiatives need to focus on text and content complexity of testlet items to better align with students' knowledge and skills of science content and to improve user experience.

As the DLM Science Consortium develops a learning map model for science, the findings described here will help inform potential revisions and additions made to the EEs and linkage-level statements for science during the second phase of project development.

References

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.

NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.