DYNAMIC®
LEARNING MAPS

2021–2022 Technical Manual Update

Pennsylvania Science Supplement

January 2023

DLM TECHNICAL MANUAL PENNSYLVANIA SUPPLEMENT

FOREWARD INFORMATION

The Dynamic Learning Maps® (DLM®) Consortium is made up of 21 state departments of education that use and develop the DLM Alternate Assessment System. DLM assessments are computer-based and accessible to students with significant cognitive disabilities for whom general state assessments are not appropriate, even with accommodations. Decisions regarding the assessment and reporting are made at the consortium level. As a result, some counts in the technical manual must be rounded to protect the identity of students in smaller states.

DLM serves as the statewide alternate assessment for accountability in Pennsylvania. DLM provides a technical manual annually to include data representative of all students from the consortium states, as well as a state specific supplement. The following Pennsylvania supplement to the DLM technical manual is available to better examine state-specific data.

The manual contains tables that are representative of all states in the consortium. Therefore, there may be some slight differences in the data represented in this report and Pennsylvania's final accountability reporting data and student score reporting. For example, Pennsylvania assesses students at grades 3–8 and 11 in English language arts and mathematics, and grades 4, 8, and 11 in science. If a student is assessed at a grade level outside of these testing grades, that assessment is invalidated in the state's final accountability and reporting process (Table 7.1, grade 9).

Questions on the data contained in this Pennsylvania Supplement can be directed to alternateassessment@pattankop.net.

Dynamic Learning Maps Consortium. (2022, December). *2021–2022 Technical Manual Update—Pennsylvania Science Supplement*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Acknowledgements

The publication of this technical manual update builds on the documentation presented in the *2015–2016 Technical Manual—Science* and annual technical manual updates. This document represents further contributions to a body of work in the service of supporting a meaningful assessment system designed to serve students with the most significant cognitive disabilities. Hundreds of people have contributed to this undertaking. We acknowledge them all for their contributions.

Many contributors made the writing of this technical manual update possible. Dynamic Learning Maps® (DLM®) staff who made significant writing contributions to this technical manual update are listed below with gratitude.

W. Jake Thompson, Ph.D., *Assistant Director for Psychometrics*
Amy K. Clark, Ph.D., *Associate Director for Operational Research*
Brooke Nash, Ph.D., *Associate Director for Psychometrics*

The authors also wish to acknowledge Ashley Hirt, Jeffrey Hoover, Elizabeth Kavitsky, Jennifer Kobrin, and Noelle Pablo for their role in developing, organizing, and compiling this manual. The authors also wish to acknowledge Brianna Beitling, Amber Cavasos, Alson Cole, Karen Erickson, Zachary Hopper, Sarah Koebley, Jessica Lancaster, Mari Langas, and Delaney Wilson for their contributions to this manual. Finally, the authors wish to thank Kristy Bledsoe, Lucas Cooper, Justin Dean, Aaron Gates, Whitney Lohrenz, and Sara Lundberg for their editing and project management work. For a list of project staff who supported the development of this manual through key contributions to design, development, or implementation of the Dynamic Learning Maps Alternate Assessment System, please see the *2015–2016 Technical Manual—Science*, and the subsequent annual technical manual updates.

We are also grateful for the contributions of the members of the DLM Technical Advisory Committee who graciously provided their expertise and feedback on the DLM System. Members of the Technical Advisory Committee during the 2021–2022 operational year include:

Russell Almond, Ph.D., *Florida State University*
Karla Egan, Ph.D., *EdMetric*
Claudia Flowers, Ph.D., *University of North Carolina at Charlotte*
Robert Henson, Ph.D., *University of North Carolina at Greensboro*
Joan Herman, Ed.D., *University of California, Los Angeles*
James Pellegrino, Ph.D., *University of Illinois Chicago*
Edward Roeber, Ph.D., *Michigan Assessment Consortium*
David Williamson, Ph.D., *The College Board*
Phoebe Winter, Ph.D., *Independent Consultant*

Contents

1 Overview **1**
1.1 Data Suppression 1
1.2 State-Specific Supplement Overview 1

2 Content Structures **3**

3 Assessment Design and Development **4**

4 Assessment Delivery **5**
4.1 Key Features of the Science Assessment Model 5
4.1.1 Assessment Administration Windows 5
4.2 Evidence from the DLM System 6
4.2.1 Administration Time 6
4.2.2 Device Usage 7
4.2.3 Blueprint Coverage 8
4.2.4 Adaptive Delivery 9
4.2.5 Administration Incidents 12
4.2.6 Accessibility Support Selections 12
4.3 Evidence From Monitoring Assessment Administration 13
4.3.1 Data Forensics Monitoring 13
4.4 Evidence From Test Administrators 14
4.4.1 User Experience With the DLM System 14
4.5 Conclusion 15

5 Modeling **16**

6 Standard Setting **17**

7 Reporting and Results **18**
7.1 Student Participation 18
7.2 Student Performance 19
7.2.1 Overall Performance 20
7.2.2 Subgroup Performance 20
7.3 Mastery Results 22
7.3.1 Linkage Level Mastery 22
7.4 Data Files 23
7.5 Score Reports 23
7.5.1 Individual Student Score Reports 23
7.6 Quality-Control Procedures for Data Files and Score Reports 23
7.7 Conclusion 23

8 Reliability **25**

9 Training and Professional Development **26**

10 Validity Argument **27**

11 References **28**

List of Tables

4.1	Distribution of Response Times per Testlet in Minutes	7
4.2	Essential Elements Required for Blueprint Coverage	8
4.3	Student Blueprint Coverage by Complexity Band	9
4.4	Correspondence of Complexity Bands and Linkage Levels	9
4.5	Adaptation of Linkage Levels Between First and Second Science Testlets	11
4.6	Accessibility Supports Selected for Pennsylvania Students	13
4.7	Test Administrator Responses Regarding Test Administration	15
7.1	Student Participation by Grade or Course	18
7.2	Demographic Characteristics of Participants	19
7.3	Percentage of Students by Grade and Performance Level	20
7.4	Science Performance Level Distributions by Demographic Subgroup	21

List of Figures

4.1 Distribution of Devices Used for Completed Testlets 8
7.1 Students' Highest Linkage Level Mastered Across Science Essential Elements by Grade. 22

1. Overview

During the 2021–2022 academic year, the Dynamic Learning Maps® (DLM®) Alternate Assessment System offered assessments of student achievement in mathematics, English language arts (ELA), and science for students with the most significant cognitive disabilities in grades 3 through 8 and high school.

A complete technical manual was created for the first year of operational administration for science (Dynamic Learning Maps Consortium [DLM Consortium], 2017). Additionally, the 2021–2022 update to the science technical manual provides updated information for the 2021–2022 administration, including only sections with changes (DLM Consortium, 2022). This volume provides state-specific information for two of those chapters. For a complete description of the DLM system for science, refer to the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

1.1. Data Suppression

In order to ensure that individual students cannot be identified, disaggregated counts have been randomly rounded to the nearest 10, the suppression threshold specified by Pennsylvania. Random rounding means that a single value could round up or down, with the probability equal to the distance to each rounded value (Matthews & Harel, 2011). For example, a value of 17 would have a 30% chance of rounding down to 10 and a 70% chance of rounding up to 20 (i.e., values are more likely to round to their nearest end point). This method ensures that all the data is properly deidentified, while providing the maximum amount of information. That is, when using simple data suppression, groups that are above the suppression threshold must often be complementarily suppressed in order to ensure that groups below the suppression threshold are properly deidentified. Random rounding allows for results to be reported for all groups, while preserving student confidentiality.

1.2. State-Specific Supplement Overview

Chapter 1 provides an overview of the contents of the Pennsylvania state-specific supplement.

Chapter 2 and Chapter 3 do not include data specific to a single state and are not included in the state-specific supplement.

Chapter 4 provides an update on assessment delivery for Pennsylvania during the 2021–2022 year. The chapter provides a summary of administration time and device usage, update analyses of blueprint coverage and adaptive delivery, a summary of updated Personal Needs and Preferences Profile selections, and test administrator survey results regarding user experience and.

Chapter 5 and Chapter 6 do not include data specific to a single state and are not included in the state-specific supplement.

Chapter 7 reports the 2021–2022 operational results for Pennsylvania, including student participation data. The chapter details the percentage of students at each performance level; subgroup performance by gender, race, ethnicity, and English learner status; and the percentage of students who showed mastery at each linkage level. Finally, the chapter provides descriptions of changes to score reports and data files during the 2021–2022 administration.

Chapter 8, Chapter 9, and Chapter 10 do not include data specific to a single states and are not included in

the state-specific supplement. For a complete summary, see the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022).

2. Content Structures

Essential Elements (EEs) are a key feature of the Dynamic Learning Maps® (DLM®) Alternate Assessment System, and serve as the conceptual and content basis for the DLM alternate assessment for science. For a description of the process used to develop the EEs, including the detailed work necessary to align them to the *Framework for K–12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012) and the Next Generation Science Standards (NGSS Lead States [NGSS], 2013), and to the needs of the student population, see Chapter 2 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

3. Assessment Design and Development

For a description of updates to the Dynamic Learning Maps® (DLM®) Alternate Assessment System's item and test development for the 2021–2022 academic year, including a summary of external reviews of items and testlets for content, bias, and accessibility; a description of the operational assessments; and a description of field tests, see Chapter 3 of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022).

For a complete description of item and test development, including a summary of item and testlet information; external reviews of items and testlets for content, bias, and accessibility; a description of operational assessments; and a description of field tests, see the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

4. Assessment Delivery

Chapter 4 of the Dynamic Learning Maps® (DLM®) Alternate Assessment System *2015–2016 Technical Manual—Science* (DLM Consortium, 2017) describes general test administration and monitoring procedures. This chapter describes updated procedures and data collected in 2021–2022, including a summary of administration time, adaptive routing, Personal Needs and Preferences Profile selections, and test administrator survey responses regarding user experience and accessibility.

Overall, administration features remained consistent with the 2020–2021 intended implementation, including the availability of instructionally embedded testlets, spring operational administration of testlets, the use of adaptive delivery during the spring window, and the availability of accessibility supports.

For a complete description of test administration for DLM assessments, including information on available resources and materials and information on monitoring assessment administration, see the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

4.1. Key Features of the Science Assessment Model

This section describes DLM test administration for 2021–2022. For a complete description of key administration features, including information on assessment delivery, the Kite Suite®, and linkage level assignment, see Chapter 4 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017). Additional information about changes in administration can also be found in the *Test Administration Manual* (DLM Consortium, 2021a) and the *Educator Portal User Guide* (DLM Consortium, 2021d).

4.1.1. Assessment Administration Windows

Assessments are administered in the spring assessment window for operational reporting. Optional assessments are available during the instructionally embedded assessment window for educators to administer for formative information. Additional descriptions of how Essential Elements (EEs) and linkage levels are assigned during the spring assessment window can be found in the Adaptive Delivery section later in this chapter.

4.1.1.1. Instructionally Embedded Assessment Window

During the instructionally embedded assessment window, testlets are optionally available for test administrators to assign to their students. When choosing to administer the optional testlets during the instructionally embedded assessment window, educators decide which EEs and linkage levels to assess for each student. The assessment delivery system recommends a linkage level for each EE based on the educator's responses to the student's First Contact survey, but educators can choose a different linkage level based on their own professional judgment. The dates for the instructionally embedded assessment window are determined by which assessment model each state participates in for English language arts (ELA) and mathematics (i.e., Instructionally Embedded or Year-End). States that only participate in the science assessments follow the dates for the Year-End model. In 2021–2022, the instructionally embedded assessment window occurred between September 13, 2021, and February 23, 2022, for states who participate in the Year-End model and between September 13, 2021, and December 17, 2021, for states who participate in the Instructionally Embedded model. States were given the option of using the entire window or setting their own dates within the larger window. In Pennsylvania, the instructionally

embedded assessment window occurred between September 13, 2021, and February 23, 2022.

4.1.1.2. Spring Assessment Window

During the spring assessment window, students are assessed on all of the EEs on the assessment blueprint in science. The linkage level for each EE is determined by the system. As with the instructionally embedded assessment window, dates for the spring assessment window are determined by which assessment model is used for ELA and mathematics. In 2021–2022, the spring assessment window occurred between March 14, 2022, and June 10, 2022, for states who participate in the Year-End model and between February 7, 2022, and May 20, 2022, for states who participate in the Instructionally Embedded model. States were given the option of using the entire window or setting their own dates within the larger window. In Pennsylvania, the spring assessment window occurred between March 14, 2022, and May 20, 2022.

4.2. Evidence from the DLM System

This section describes evidence collected by the DLM System during the 2021–2022 operational administration of the DLM alternate assessment. The categories of evidence include data relating to administration time, device usage, adaptive routing, and accessibility support selections.

4.2.1. Administration Time

Estimated administration time varies by student and subject. Testlets can be administered separately across multiple testing sessions as long as they are all completed within the testing window.

The published estimated total testing time per testlet is around 5–15 minutes. The estimated total testing time is 45–135 minutes per student in the spring assessment window. Published estimates are slightly longer than anticipated real testing times because of the assumption that test administrators need time for setup. Actual testing time per testlet varies depending on each student's unique characteristics.

Kite Student Portal captured start dates, end dates, and time stamps for every testlet. The difference between these start and end times was calculated for each completed testlet. Table 4.1 summarizes the distribution of test times per testlet for students in Pennsylvania. The distribution of test times in Table 4.1 is consistent with the distribution observed in prior years. Most testlets took around three minutes or less to complete. Time per testlet may have been impacted by student breaks during the assessment. Testlets with shorter than expected administration times are included in an extract made available to each state. States can use this information to monitor assessment administration and address as necessary. For a description of the administration time monitoring extract, see section 4.3.1 of this chapter.

Table 4.1

Distribution of Response Times per Testlet in Minutes

Grade	Min	Median	Mean	Max	25Q	75Q	IQR
Elementary	.133	2.12	2.86	89.30	1.37	3.33	1.97
Middle school	.117	1.78	2.48	79.83	1.12	2.87	1.75
High school	.133	2.12	2.83	89.75	1.33	3.33	2.00

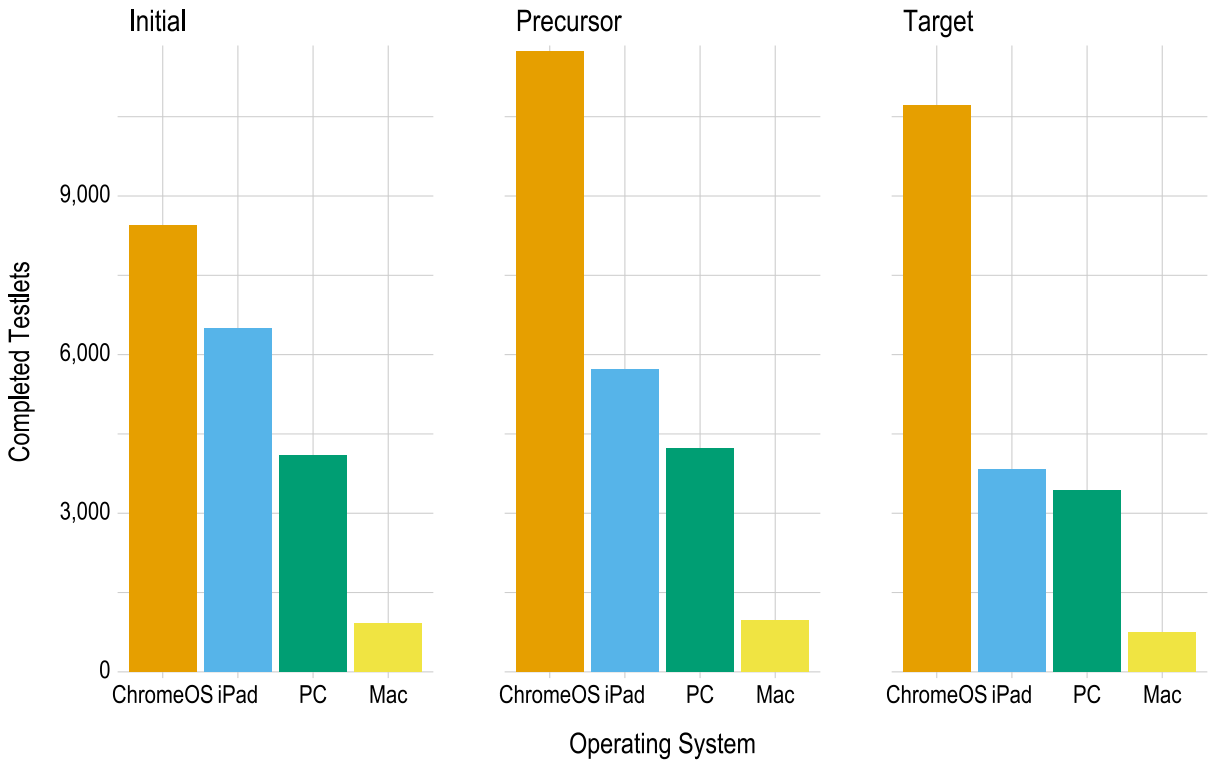
Note. Min = minimum, Max = maximum, 25Q = lower quartile, 75Q = upper quartile, IQR = interquartile range.

4.2.2. Device Usage

Testlets may be administered on a variety of devices. Kite Student Portal captured the operating system used for each testlet completed. Although these data do not capture specific devices used to complete each testlet (e.g., SMART Board, switch system, etc.), they provide high-level information about how students access assessment content. For example, we can identify how often an iPad is used relative to a Chromebook or traditional PC. Figure 4.1 shows the number of testlets completed on each operating system by subject and linkage level for 2021–2022. In Pennsylvania, 50% of testlets were completed on a Chromebook, 26% were completed on an iPad, 19% were completed on a PC, and 4% were completed on a Mac.

Figure 4.1

Distribution of Devices Used for Completed Testlets



4.2.3. Blueprint Coverage

Each student is assessed on all EEs included on the assessment blueprint.[1] Table 4.2 summarizes the number of EEs required for each grade or course.

Table 4.2

Essential Elements Required for Blueprint Coverage

Grade or Course	*n*
Elementary	9
Middle school	9
High school	9
Biology	10

Across all grades, 95% of students in Pennsylvania were assessed on all of the EEs and met blueprint requirements. Table 4.3 summarizes the total number of students and the percentage of students meeting

[1] For a description of the assessment blueprints see Chapter 3 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

blueprint requirements based on their complexity band. When comparing complexity band distributions, there was a slightly lower percentage of Foundational students not meeting requirements. However, all complexity band groups had over 90% of students meeting the coverage requirements.

Table 4.3

Student Blueprint Coverage by Complexity Band

Complexity Band	*n*	% meeting requirements
Foundational	950	90.6
Band 1	2,230	94.9
Band 2	2,440	96.3
Band 3	1,380	96.2

Note. Counts were randomly rounded to the nearest 10.

4.2.4. Adaptive Delivery

During the spring 2022 test administration, the science assessments were adaptive between testlets, following the same routing rules applied in prior years. That is, the linkage level associated with the next testlet a student received was based on the student's performance on the most recently administered testlet, with the specific goal of maximizing the match of student knowledge and skill to the appropriate linkage level content.

- The system adapted up one linkage level if the student responded correctly to at least 80% of the items measuring the previously tested EE. If the previous testlet was at the highest linkage level (i.e., Target), the student remained at that level.
- The system adapted down one linkage level if the student responded correctly to less than 35% of the items measuring the previously tested EE. If the previous testlet was at the lowest linkage level (i.e., Initial), the student remained at that level.
- Testlets remained at the same linkage level if the student responded correctly to between 35% and 80% of the items on the previously tested EE.

The linkage level of the first testlet assigned to a student was based on First Contact survey responses. The correspondence between the First Contact complexity bands and first assigned linkage levels are shown in Table 4.4.

Table 4.4

Correspondence of Complexity Bands and Linkage Levels

First Contact complexity band	Linkage level
Foundational	Initial
Band 1	Initial
Band 2	Precursor
Band 3	Target

Following the spring 2022 administration, analyses were conducted to determine the mean percentage of testlets that adapted from the first to second testlet administered for students within a grade or course and complexity band. The aggregated results can be seen in Table 4.5.

Due to small sample size, data regarding the adaptation of linkage levels in Pennsylvania was only available for grades 4, 8, and 11. For the majority of students across grades who were assigned to the Foundational Complexity Band by the First Contact survey, testlets did not adapt to a higher linkage level after the first assigned testlet (ranging from 54% to 57%). A similar pattern was seen for students assigned to Band 3, with the majority of students not adapting down to a lower linkage level after the first assigned testlet (ranging from 62% to 81%). In contrast, students assigned to Band 1 tend to adapt up to a higher linkage level after their first testlet (ranging from 51% to 75%). Consistent patterns were not as apparent for students who were assigned to Band 2. Results indicate that linkage levels of students assigned to higher complexity bands are more variable with respect to the direction in which students move between the first and second testlets. However, this finding of more variability in the higher complexity bands is consistent with prior years, which showed the same trend. Several factors may help explain these results, including more variability in student characteristics within this group and content-based differences across grades. For a description of previous findings, see Chapter 4 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017) and the subsequent technical manual updates (DLM Consortium, 2018a, 2018b, 2019, 2021b).

Table 4.5

*Adaptation of Linkage Levels Between First and Second Science Testlets (*N = *7,010)*

Grade	Foundational		Band 1		Band 2			Band 3	
	Adapted up (%)	Did not adapt (%)	Adapted up (%)	Did not adapt (%)	Adapted up (%)	Did not adapt (%)	Adapted down (%)	Did not adapt (%)	Adapted down (%)
Grade 3	*	*	*	*	*	*	*	*	*
Grade 4	43.3	56.7	75.3	24.7	26.8	48.1	25.2	62.0	38.0
Grade 5	*	*	*	*	*	*	*	*	*
Grade 6	*	*	*	*	*	*	*	*	*
Grade 7	*	*	*	*	*	*	*	*	*
Grade 8	46.4	53.6	68.3	31.7	35.0	43.3	21.7	67.4	32.6
Grade 9	*	*	*	*	*	*	*	*	*
Grade 10	*	*	*	*	*	*	*	*	*
Grade 11	46.3	53.7	50.6	49.4	41.7	40.7	17.6	80.8	19.2
Grade 12	*	*	*	*	*	*	*	*	*
Biology	*	*	*	*	*	*	*	*	*

* These data were suppressed because $n < 50$.

Note. Foundational and Band 1 correspond to the testlets at the lowest linkage level, so testlets could not adapt down a linkage level. Band 3 corresponds to testlets at the highest linkage level in science, so testlets could not adapt up a linkage level.

4.2.5. Administration Incidents

DLM staff annually evaluates testlet assignment to ensure students are correctly assigned to testlets. Administration incidents that have the potential to affect scoring are reported to state education agencies in a supplemental Incident File. No incidents were observed during the 2021–2022 operational assessment windows. Assignment of testlets will continue to be monitored in subsequent years to track any potential incidents and report them to state education agencies.

4.2.6. Accessibility Support Selections

Accessibility supports provided in 2021–2022 were the same as those available in previous years. The DLM *Accessibility Manual* (DLM Consortium, 2021c) distinguishes accessibility supports that are provided in Kite Student Portal via the Personal Needs and Preferences Profile, require additional tools or materials, or are provided by the test administrator outside the system. Table 4.6 shows selection rates for the three categories of accessibility supports. Overall, 6,123 students (78%) had at least one support selected. The most commonly selected supports in 2021–2022 were human read aloud, spoken audio, and test administrator enters responses for student. For a complete description of the available accessibility supports, see Chapter 4 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017). Additionally, educators reported in the First Contact survey that 51% of students in Pennsylvania were able to access a computer independently, with or without assistive technology.

Table 4.6

*Accessibility Supports Selected for Pennsylvania Students (*N = *7,806)*

Support	*n*	%
Supports provided in Kite Student Portal		
Spoken audio	3,390	43.4
Magnification	1,030	13.2
Color contrast	600	7.7
Overlay color	240	3.1
Invert color choice	180	2.3
Supports requiring additional tools/materials		
Calculator	2,240	28.7
Individualized manipulatives	1,580	20.2
Single-switch system	260	3.3
Alternate form - visual impairment	140	1.8
Two-switch system	60	0.8
Uncontracted braille	10	0.1
Supports provided outside the system		
Human read aloud	5,000	64.1
Test administrator enters responses for student	2,990	38.3
Partner-assisted scanning	320	4.1
Sign interpretation of text	120	1.5
Language translation of text	60	0.8

Note. Counts were randomly rounded to the nearest 10.

4.3. Evidence From Monitoring Assessment Administration

Monitoring of assessment administration was conducted using various materials and strategies. DLM project staff developed an assessment administration monitoring protocol for use by DLM staff, state education agency staff, and local education agency staff. Project staff also reviewed Service Desk contacts and hosted regular check-in calls to monitor common issues and concerns during the assessment window. This section provides an overview of all resources and supports as well as more detail regarding the assessment administration observation protocol and its use, check-in calls with states, and methods for monitoring testlet delivery.

4.3.1. Data Forensics Monitoring

Two data forensics monitoring reports are available in Educator Portal. The first report includes information about testlets completed outside of normal business hours. The second report includes information about testlets that were completed within a short period of time.

The Testing Outside of Hours report allows state education agencies to specify days and hours within a day that testlets are expected to be completed. Each state can select its own days and hours for setting expectations. For example, a state could elect to flag any testlet completed outside of Monday through

Friday from 6:00 a.m. to 5:00 p.m. local time. The Testing Outside of Hours report then identifies students who completed assessments outside of the defined expected hours. Overall, 347 (1%) science testlets were completed outside of the expected hours by 308 (4%) students in Pennsylvania.

The Testing Completed in a Short Period of Time report identifies students who completed a testlet within an unexpectedly short period of time. The threshold for inclusion in the report was testlet completion time of less than 30 seconds. The report is intended for state users to identify potentially aberrant response patterns; however there are many legitimate reasons a testlet may be submitted in a short time period. Overall, 1,005 (2%) testlets were completed in a short period of time by 593 (8%) students in Pennsylvania.

4.4. Evidence From Test Administrators

This section first describes evidence collected from the spring 2022 test administrator survey. Data on user experience with the DLM System as well as student opportunity to learn is evaluated annually through a survey that test administrators are invited to complete after administration of the spring assessment. Test administrators receive one survey per rostered DLM student, which collects information about that student's assessment experience. As in previous years, the survey was distributed to test administrators in Kite Student Portal, where students completed assessments. The survey consisted of four blocks. Blocks 1 and 4 were administered in every survey. Block 1 included questions about the test administrator's perceptions of the assessments and the student's interaction with the content, and Block 4 included questions about the test administrator's background. Block 2 was spiraled, so test administrators received one randomly assigned section. In these sections, test administrators were asked about one of the following topics per survey: relationship to ELA instruction, relationship to mathematics instruction, or relationship to science instruction. Block 3 was added in 2021 and remained in the survey in 2022 to gather information about educational experiences during the COVID-19 pandemic.

4.4.1. User Experience With the DLM System

A total of 2,642 test administrators from Pennsylvania responded to the survey (81%) about 5,661 students' experiences. Test administrators are instructed to respond to the survey separately for each of their students. Participating Pennsylvania test administrators responded to surveys for a median of two students. Pennsylvania test administrators reported having an average of 9 years of experience in science and 11 years of experience with students with significant cognitive disabilities.

The following sections summarize responses regarding both educator and student experience with the system.

4.4.1.1. Educator Experience

Test administrators were asked to reflect on their own experience with the assessments as well as their comfort level and knowledge administering them. Most of the questions required test administrators to respond on a 4-point scale: *strongly disagree, disagree, agree,* or *strongly agree*. Responses are summarized in Table 4.7.

Nearly all Pennsylvania test administrators (95%) agreed or strongly agreed that they were confident administering DLM testlets. Most respondents (89%) agreed or strongly agreed that the required test administrator training prepared them for their responsibilities as test administrators. Most test

administrators also responded that they had access to curriculum aligned with the content that was measured by the assessments (80%) and that they used the manuals and the Educator Resources page (88%).

Table 4.7

Test Administrator Responses Regarding Test Administration

	SD		D		A		SA		A+SA	
Statement	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
I was confident in my ability to deliver DLM testlets.	25	1.6	58	3.8	737	47.8	722	46.8	1,459	94.6
Required test administrator training prepared me for the responsibilities of a test administrator.	45	2.9	128	8.3	856	55.7	509	33.1	1,365	88.8
I have access to curriculum aligned with the content measured by DLM assessments.	58	3.8	245	16.0	815	53.2	415	27.1	1,230	80.3
I used manuals and/or the DLM Educator Resource Page materials.	36	2.3	156	10.1	905	58.7	445	28.9	1,350	87.6

Note. SD = strongly disagree; D = disagree; A = agree; SA = strongly agree; A+SA = agree and strongly agree.

4.5. Conclusion

Delivery of the DLM System was designed to align with instructional practice and be responsive to individual student needs. Assessment delivery options allow for necessary flexibility to reflect student needs while also including constraints to maximize comparability and support valid interpretation of results. The dynamic nature of DLM assessment administration is reflected in the initial input through the First Contact survey, as well as adaptive routing between testlets. Evidence collected from the DLM System, test administration monitoring, and test administrators indicates that students are able to successfully interact with the system to demonstrate their knowledge, skills, and understandings.

5. Modeling

The Dynamic Learning Maps® (DLM®) Alternate Assessment System draws upon a well-established research base in cognition and learning theory but relatively uncommon operational psychometric methods to provide feedback about student performance. The approach uses innovative operational psychometric methods to provide feedback about student mastery of skills. For modeling evidence from 2021–2022, including a complete description of the psychometric model used to calibrate and score the DLM assessments, the psychometric background, the structure of the assessment system suitability for diagnostic modeling, and a detailed summary of the procedures used to calibrate and score DLM assessments, see Chapter 5 of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022).

6. Standard Setting

The standard setting process for the Dynamic Learning Maps® (DLM®) Alternate Assessment System in science derived cut points for assigning students to four performance levels based on results from the 2015–2016 DLM alternate assessments. For a description of the process, including the development of policy performance level descriptors, the 3-day standard setting meeting, follow-up evaluation of impact data and cut points, and specification of content-specific performance level descriptors, see Chapter 6 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

7. Reporting and Results

Chapter 7 of the Dynamic Learning Maps® (DLM®) Alternate Assessment System *2015–2016 Technical Manual—Science* (DLM Consortium, 2017) describes assessment results for the 2015–2016 academic year, including student participation and performance summaries, and an overview of data files and score reports delivered to state education agencies. Technical Manual updates provide a description of data files, score reports, and results for each corresponding academic year.

This chapter presents Pennsylvania-specific spring 2022 student participation data; the percentage of students achieving at each performance level; and subgroup performance by gender, race, ethnicity, and English learner status. This chapter also reports the distribution of students by the highest linkage level mastered during spring 2022. Finally, this chapter describes updates made to score reports during the 2021–2022 operational year. For a complete description of score reports and interpretive guides, see Chapter 7 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

7.1. Student Participation

During spring 2022, assessments were administered to 7,010 students in Pennsylvania. The assessments were administered by 3,047 educators in 1,656 schools and 659 school districts. A total of 61,318 test sessions were administered during the spring assessment window. One test session is one testlet taken by one student. Only test sessions that were complete at the close of the spring assessment window counted toward the total sessions.

Table 7.1 summarizes the number of students assessed in each grade and course. More than 2,370 students participated in each of the elementary and the middle school grade bands.[2] In high school, almost 2,200 students participated.

Table 7.1

*Student Participation by Grade or Course (*N = *7,010)*

Grade	Students (*n*)
3	30
4	2,450
5	20
6	20
7	30
8	2,320
11	2,140

Note. Counts were randomly rounded to the nearest 10.

Table 7.2 summarizes the demographic characteristics of the students who participated in the spring 2022 administration. The distribution of students across the different subgroups was fairly consistent with prior

[2] In an effort to increase science instruction beyond the tested grades, several states promoted participation in the science assessment at all grade levels (i.e., did not restrict participation to the grade levels required for accountability purposes).

years' distributions. The majority of participants were male (69%) and white (55%). About 6% of students were monitored or eligible for English learning services.

Table 7.2

*Demographic Characteristics of Participants (*N = *7,010)*

Subgroup	*n*	%
Gender		
Male	4,820	68.8
Female	2,190	31.2
Race		
White	3,850	54.9
African American	1,520	21.7
Two or more races	1,330	19.0
Asian	270	3.9
American Indian	20	0.3
Native Hawaiian or Pacific Islander	10	0.1
Alaska Native	10	0.1
Hispanic ethnicity		
Non-Hispanic	5,920	84.5
Hispanic	1,090	15.5
English learning (EL) participation		
Not EL eligible or monitored	6,570	93.7
EL eligible or monitored	440	6.3

Note. Counts were randomly rounded to the nearest 10.

In addition to the spring assessment window, instructionally embedded science assessments are also made available for educators to optionally administer to students during the year. Results from the instructionally embedded assessments do not contribute to final summative scoring but can be used to guide instructional decision-making. No students in Pennsylvania took an instructionally embedded testlet during 2021–2022.

7.2. Student Performance

Student performance on DLM assessments is interpreted using cut points,[3] which describe student achievement using four performance levels. A student's performance level is determined based on the total number of linkage levels mastered across the assessed EEs.

For the spring 2022 administration, student performance was reported using the same four performance levels approved by the DLM Governance Board for prior years:

[3] For a description of the standard setting process used to determine the cut points, see Chapter 6 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

- The student demonstrates *Emerging* understanding of and ability to apply content knowledge and skills represented by the EEs.
- The student's understanding of and ability to apply targeted content knowledge and skills represented by the EEs is *Approaching the Target*.
- The student's understanding of and ability to apply content knowledge and skills represented by the EEs is *At Target*. This performance level is considered to be meeting achievement expectations.
- The student demonstrates *Advanced* understanding of and ability to apply targeted content knowledge and skills represented by the EEs.

7.2.1. Overall Performance

Table 7.3 reports the percentage of Pennsylvania students achieving at each performance level from the spring 2022 administration for science. At the elementary level, the percentage of students who achieved at the At Target or Advanced levels (i.e., proficient) was approximately 15%; in middle school the percentage of students meeting or exceeding At Target expectations was approximately 26%; in high school the percentage was approximately 18%.

Table 7.3

Percentage of Students by Grade and Performance Level

Grade	Emerging (%)	Approaching (%)	At Target (%)	Advanced (%)	At Target + Advanced (%)
3 (*n* = 30)	50.0	25.0	25.0	0.0	25.0
4 (*n* = 2,450)	61.0	24.0	12.6	2.4	15.0
5 (*n* = 20)	33.3	33.3	33.3	0.0	33.3
6 (*n* = 20)	33.3	33.3	33.3	0.0	33.3
7 (*n* = 30)	40.0	20.0	20.0	20.0	40.0
8 (*n* = 2,320)	52.4	22.3	21.9	3.4	25.3
11 (*n* = 2,140)	50.7	31.2	15.8	2.3	18.1

Note. Counts were randomly rounded to the nearest 10.

7.2.2. Subgroup Performance

Data collection for DLM assessments includes demographic data on gender, race, ethnicity, and English learning status. Table 7.4 summarizes the Pennsylvania disaggregated frequency distributions for science, collapsed across all assessed grade levels.

Table 7.4

*Science Performance Level Distributions by Demographic Subgroup (*N = *7,010)*

Subgroup	Emerging		Approaching		At Target		Advanced		At Target + Advanced	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender										
Male	2,640	54.8	1,200	24.9	830	17.2	150	3.1	980	20.3
Female	1,220	55.7	600	27.4	330	15.1	40	1.8	370	16.9
Race										
White	2,010	52.2	1,010	26.2	710	18.4	120	3.1	830	21.6
African American	880	57.9	390	25.7	220	14.5	30	2.0	250	16.4
Two or more races	770	57.9	330	24.8	200	15.0	30	2.3	230	17.3
Asian	180	66.7	50	18.5	30	11.1	10	3.7	40	14.8
American Indian	10	25.0	10	25.0	10	25.0	10	25.0	20	50.0
Native Hawaiian or Pacific Islander	10	33.3	10	33.3	10	33.3	0	0.0	10	33.3
Alaska Native	10	>99.9	0	0.0	0	0.0	0	0.0	0	0.0
Hispanic ethnicity										
Non-Hispanic	3,200	54.1	1,530	25.9	1,020	17.3	160	2.7	1,180	20.0
Hispanic	660	60.0	270	24.5	150	13.6	20	1.8	170	15.5
English learning (EL) participation										
Not EL eligible or monitored	3,610	55.0	1,680	25.6	1,100	16.8	170	2.6	1,270	19.4
EL eligible or monitored	250	56.8	120	27.3	60	13.6	10	2.3	70	15.9

Note. Counts were randomly rounded to the nearest 10.

7.3. Mastery Results

As described above, the student performance levels are determined by applying cut points to the total number of linkage levels mastered. In this section, we summarize student mastery of assessed EEs and linkage levels.

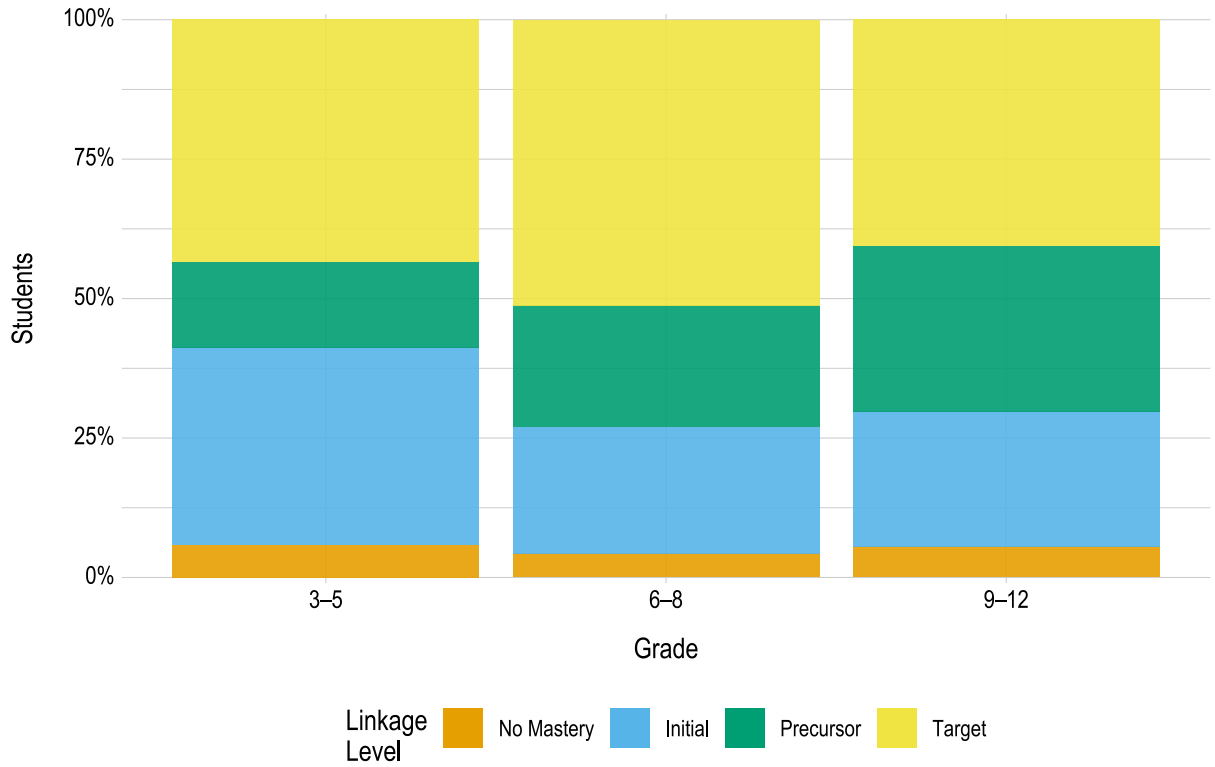
7.3.1. Linkage Level Mastery

Scoring for DLM assessments determines the highest linkage level mastered for each EE. The linkage levels are (in order): Initial, Precursor, and Target. A student can be a master of zero, one, two, or all three linkage levels, within the order constraints. For example, if a student masters the Precursor level, they also master the Initial linkage level. This section summarizes the distribution of students by highest linkage level mastered across all EEs. For each student, the highest linkage level mastered across all tested EEs was calculated. Then, for each grade, the number of students with each linkage level as their highest mastered linkage level across all EEs was summed and then divided by the total number of students who tested in the grade. This resulted in the proportion of students for whom each level was the highest linkage level mastered.

Figure 7.1 displays the percentage of Pennsylvania students who mastered each linkage level as the highest linkage level across all assessed EEs in science. For example, across all 3-5 science EEs, the Target level was the highest level that students mastered 43% of the time. The percentage of students who mastered as high as the Target linkage level ranged from approximately 41% to 51%.

Figure 7.1

Students' Highest Linkage Level Mastered Across Science Essential Elements by Grade



7.4. Data Files

DLM assessment results were made available to DLM state education agencies following the spring 2022 administration. Similar to prior years, the General Research File (GRF) contained student results, including each student's highest linkage level mastered for each EE and final performance level for science for all students who completed any testlets. In addition to the GRF, the states received several supplemental files. Consistent with prior years, the special circumstances file provided information about which students and EEs were affected by extenuating circumstances (e.g., chronic absences), as defined by each state. State education agencies also received a supplemental file to identify exited students. The exited students file included all students who exited at any point during the academic year. In the event of observed incidents during assessment delivery, state education agencies are provided with an incident file describing students impacted; however, no incidents occurred during 2021–2022.

Consistent with prior delivery cycles, state education agencies were provided with a 2-week window following data file delivery to review the files and invalidate student records in the GRF. Decisions about whether to invalidate student records are informed by individual state policy. If changes were made to the GRF, state education agencies submitted final GRFs via Educator Portal. The final GRF was used to generate score reports.

7.5. Score Reports

Assessment results were provided to state education agencies to report to parents/guardians, educators, and local education agencies. Individual Student Score Reports summarized student performance on the assessment. Several aggregated reports were provided to state and local education agencies, including reports for the classroom, school, district, and state. No changes were made to the structure of aggregated reports during spring 2022. One change to the Individual Student Score Reports is summarized below. For a complete description of score reports, including aggregated reports, see Chapter 7 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

7.5.1. Individual Student Score Reports

Because of continued impacts of the COVID-19 pandemic on instruction and assessment, during 2021–2022, state education agencies were given the option to add a cautionary statement to Individual Student Score Reports, which indicated that the results may reflect the continued effects of the COVID-19 pandemic on student performance. However, Pennsylvania did not opt to include the cautionary statement on individual student score reports.

7.6. Quality-Control Procedures for Data Files and Score Reports

Changes to the quality-control procedures were made only to the extent of accommodating the revised score reports for 2021–2022 (i.e., checking to be sure the cautionary statement was correctly applied for states who opted to include it on score reports). For a complete description of quality-control procedures, see Chapter 7 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

7.7. Conclusion

Results for DLM assessments include students' overall performance levels and linkage level mastery decisions for each assessed EE and linkage level. During spring 2022, assessments were administered to

7,010 students in Pennsylvania. Between 15% and 40% of Pennsylvania students achieved at the At Target or Advanced levels across all grades.

Following the spring 2022 administration, three data files were delivered to state education agencies: GRF, special circumstance code file, and exited students file. Lastly, state education agencies could opt to include cautionary text to score reports to aid in interpretation.

8. Reliability

The Dynamic Learning Maps® (DLM®) Alternate Assessment System uses diagnostic classification models to produce student score reports. As such, evidence for the reliability of results is based on methods that are commensurate with the models used to produce score reports. For reliability evidence from 2021–2022, including a complete description of the simulation-based methods used to calculate reliability for DLM assessments and the psychometric background for these methods, see Chapter 8 of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022).

9. Training and Professional Development

To support the instruction and the implementation of the Dynamic Learning Maps® (DLM®) Alternate Assessment System, training is offered for state and local education agency staff and test administrators. Additionally, optional professional development is provided for teachers and other staff.

For a complete description of facilitated and self-directed training for DLM assessments, including a description of training for state and local education agency staff, see Chapter 10 of the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

For a description of the optional professional development available for the Dynamic Learning Maps® (DLM®) Alternate Assessment System during 2021–2022, see Chapter 9 of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022).

10. Validity Argument

The Dynamic Learning Maps® (DLM®) Alternate Assessment System is based on the core belief that all students should have access to challenging, grade-level academic content. Therefore, the DLM assessments provide students with the most significant cognitive disabilities the opportunity to demonstrate what they know and can do. It is designed to map students' learning after a full year of instruction.

The DLM system completed its seventh operational administration year in 2021–2022. The chapters of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022) provide updated evidence from the 2021–2022 year to support the propositions and assumptions that undergird the assessment system as described at the onset of its design in the DLM theory of action. Chapter 10 of the *2021–2022 Technical Manual Update—Science* (DLM Consortium, 2022) summarizes that manual's contents and describes plans for future studies. For a complete summary of evidence collected for the DLM theory of action, also see the *2015–2016 Technical Manual—Science* (DLM Consortium, 2017).

11. References

DLM Consortium. (2021a). *Test Administration Manual 2021–2022*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2017). *2015–2016 Technical Manual—Science*. University of Kansas, Center for Educational Testing and Evaluation.

Dynamic Learning Maps Consortium. (2018a). *2016–2017 Technical Manual Update—Science*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2018b). *2017–2018 Technical Manual Update—Science*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2019). *2018–2019 Technical Manual Update—Science*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2021b). *2020–2021 Technical Manual Update—Science*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2021c). *Accessibility Manual 2021–2022*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2021d). *Educator Portal User Guide*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Dynamic Learning Maps Consortium. (2022). *2021–2022 Technical Manual Update—Science*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Matthews, G. J., & Harel, O. (2011). Data confidentiality: A review of methods for statistical disclosure limitation and methods for assessing privacy. *Statistics Surveys*, *5*, 1–29. https://doi.org/10.1214/11-SS074

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

NGSS Lead States. (2013). *Next Generation Science Standards: For States, by States*. The National Academies Press.