



DYNAMIC[®]
LEARNING MAPS

Results from External Review during the 2014–2015 Academic Year

Technical Report #15-01

11/10/2015

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

Clark, A., Swinburne Romine, R., Bell, B., & Karvonen, M. (2015). *Results from external review during the 2014–2015 academic year* (Technical Report No. 15-01). Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

The present publication was developed under grant 84.373X100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the authors, and no official endorsement by the U.S. Department of Education should be inferred.

Table of Contents

External Reviewers	4
External Review Process.....	8
Results of Reviews during the 2014-2015 Academic Year	9
Content Team Decisions.....	10
English Language Arts Text Review.....	12
Conclusion.....	14
References	15
Appendix: Review Criteria for Panelists	16

Results from External Review during the 2014–2015 Academic Year

The purpose of this report is to provide a summary of the outcomes of the external review of tasks, testlets, and texts during the 2014–2015 academic year for the Dynamic Learning Maps™ (DLM®) alternate assessment system. The report includes a description of the reviewers who were selected to participate in the external review process, a summary of the external review recommendations for individual tasks and testlets, and a summary of the decisions made by the DLM content teams as a result of the feedback provided by external reviewers.

The external review process occurs between two distinct steps in the test development cycle: after an internal review of tasks and testlets and before the final decision is made regarding which tasks and testlets to field-test. Panels are comprised of three members, including one member responsible for the review of content, one for the review of accessibility, and one for the review of bias and sensitivity criteria. In 2014–2015, the external review of English language arts (ELA) texts was conducted separately from the external review of tasks and testlets. Results from the text review are summarized at the end of this report.

External Reviewers

Volunteer reviewers were recruited during September of 2014 from DLM partner states to review testlets (each consisting of a few tasks plus an engagement activity) intended for administration during the 2014–2015 academic year in the DLM alternate assessment system. A recruitment letter was prepared by DLM staff and disseminated to state partners for distribution. The letter included a brief overview of the external review process, an explanation of the incentive plan (described below), and a request for interested individuals to complete an online Qualtrics survey to provide their qualifications and contact information.

During the 2014–2015 year, an incentive plan was put in place to entice reviewers and to prevent steep attrition rates as witnessed the previous year.¹ The incentive plan included the following requirements for reviewers to receive a payment of \$20 per assignment:

- Reviewers were required to complete all assignments sent, with payment at every five assignments. For example, if a reviewer completed all seven assignments sent, they received \$100.
- If there were fewer than five assignments available for a panel, the panelists were required to complete all assignments sent to be paid \$20 per assignment.

¹ There was an overall attrition rate of 62% for the 2013–2014 review window.

- Reviewers who received five or fewer assignments and did not complete all of them did not receive payment.

DLM received information from a total of 475 individuals interested in serving as external reviewers during the 2014–2015 academic year. These volunteers included classroom teachers and other local educators, representatives from state departments of education, and university faculty members.

Of the pool of 475 volunteers, 118 were ineligible for serving as external reviewers. Reasons for ineligibility included not providing qualifying information or volunteering too late to be placed on a complete panel. This attrition rate of 25% brought the pool of available reviewers to 357.

Of the 357 volunteers eligible to participate in the external review process, the average number of years of experience in preschool through 12th-grade education was 11 years. A total of 35% of volunteers indicated they had previous review experience for another large-scale assessment, and 42% had served as external reviewers for DLM during the 2013–2014 academic year. Table 1 describes the number of eligible reviewers by state.

Table 1

Number of Eligible Reviewers by State

State	<i>n</i>	%
Alaska	7	2.0
Colorado	0	0.0
Illinois	17	4.8
Iowa	6	1.7
Kansas	42	11.8
Michigan	9	2.5
Mississippi	18	5.0
Missouri	121	33.9
New Hampshire	1	0.3
New Jersey	42	11.7
North Carolina	5	1.4
North Dakota	4	1.1
Oklahoma	11	3.1
Pennsylvania	0	0.0
Utah	12	3.3
Vermont	15	4.2
Virginia	4	1.1
Washington	0	0.0
West Virginia	6	1.7
Wisconsin	32	9.0
Other*	5	1.4
Total	357	

*From non-DLM states. Because external reviewers were recruited through DLM state partners, it was assumed that these reviewers were intentionally recruited.

Of the 357 volunteers eligible to serve as external reviewers, 135 did not complete training and therefore did not qualify for placement on a review panel. This resulted in a reviewer pool of 222 and an attrition rate of 47%.

The 222 reviewers were assigned to one of three types of review panels: (1) accessibility, (2) bias and sensitivity, or (3) content review. In order for reviewers to qualify for the accessibility review panel, they had to have at least one year of experience

(preschool through grade 12) working with students with significant cognitive disabilities or at least one year of experience with alternate assessments based on alternate achievement standards. Priority for the bias and sensitivity review panel was given to reviewers who selected a race other than white in the volunteer survey, but other reviewers were included in the panel as needed. To qualify for the content review panel, reviewers had to have at least one year of educational experience (preschool through grade 12) in ELA or mathematics. Regardless of panel type, all reviewers had to have access to a secure computer and agree to complete at least two rounds of reviews.

A total of 60 reviewers were assigned to the accessibility review panel, 70 were assigned to the bias and sensitivity review panel, and 92 were assigned to the content review panel. The content panel had the most review criteria to evaluate, so their review time per testlet was longer than that of the other two review panel types and more panelists were needed.

A panel from these three review groups was created for each of the two testing models (integrated and year-end). Reviewers were assigned to the panels based on the state in which they teach. Of the 222 reviewers who completed training, 148 were placed on integrated integrated-model panels, and 74 were placed on year-end-model panels. Integrated model panelists reviewed only testlets measuring a single Essential Element, which consist of three to five tasks. Year-end model panelists primarily reviewed testlets measuring multiple Essential Elements, which consist of three to eight tasks. However, when needed, year-end reviewers also reviewed testlets designed for instructionally embedded assessment, which are available in all states regardless of the assessment model.

Of the 222 reviewers assigned to panels, 136 completed two or more review assignments, reflecting an attrition rate of 71% of the original 475 volunteer reviewers and a 39% attrition rate at this stage. In all, 109 reviewers completed all assigned reviews. Because of the attrition after assignments were made and the need to complete reviews in time for field testing, a small group of reviewers, called “Power Reviewers,” were selected to review for all three panel types. From the pool of 24 Power Reviewers used in 2013–2014, six Power Reviewers were selected based on the quality of their previous work and their availability for 2014–2015 reviews. They completed training and received review bundles for all three panel types. Power Reviewers received an honorarium based on the volume of testlets they reviewed. In addition, two part-time hourly paid reviewers were selected to increase the pool of reviewers who would complete assignments on time and with a high standard of quality. These hourly reviewers met the qualifications for all three panel types and completed the training for all three panels.

Table 2 presents the number of reviewers by content area and panel type.

Table 2

Number of Reviewers by Content Area (N = 109)

Review Panel	English	
	Language Arts	Mathematics
Accessibility	18	15
Bias and Sensitivity	17	22
Content	17	20
All Three Panels*	3	5
Total	55	62

*Includes Power Reviewers and paid hourly reviewers

External Review Process

Prior to receiving their first external review assignment, volunteer reviewers completed a set of training activities to become familiar with the DLM external review process. The first training module included a general overview of the process. Following completion of the first module, reviewers submitted a test security agreement. Next, reviewers completed a module specific to the type of reviews they had been assigned (accessibility, bias and sensitivity, or content). A final module covered the procedures for completing reviews. After completion of all three modules, reviewers completed a quiz and a practice activity to familiarize themselves with the review process. Volunteers who completed the security agreement, quiz, and practice activity received testlets to review. Those who did not complete all three requirements, including a minimum score on the quiz, did not receive content to review. The quiz allowed multiple opportunities to achieve a passing score.

External reviewers remotely reviewed tasks and testlets by previewing them on their own computers and completing an online survey about each set of tasks. During the process of reviewing each task or testlet, reviewers asked themselves the following question: “Does this task or testlet meet minimal standards for acceptability based on my panel’s criteria?” “Acceptability” was defined as meeting minimum standards for field testing readiness. Based on their response to this question, reviewers made one of three recommendations:

1. Accept: The task or testlet is within acceptable limits for field testing.
2. Revise: The task or testlet violates one or more of the review criteria; however, the task or testlet has potential merits and can be acceptable for field testing after revisions to address the criteria.
3. Reject: The content of the task or testlet is fundamentally flawed; revision would not bring the task or testlet to acceptable limits.

If a recommendation for acceptance was made for the task or testlet, no further information was needed from the reviewer. If a recommendation for revision was made, the reviewer’s comment identified the problem as well as a proposed solution. In the case of a recommendation for rejection, the reviewer’s comment identified the problem.

Results of Reviews during the 2014–2015 Academic Year

The majority of the content reviewed during the 2014–2015 academic year was included in the three field-testing events. On a limited basis, content for the upcoming 2015–2016 school year was also reviewed. The reviewers’ recommendations for tasks and testlets are presented in Table 3 and Table 4. Because multiple panelists reviewed each task and testlet, these values represent the count of recommendations rather than the number of unique tasks and testlets reviewed.

For ELA, the percent of recommendations for acceptance ranged from 85% to 93%. The rate at which content was recommended for rejection ranged from < 1% to 1%.

Table 3

ELA External Review Recommendations

	Task						Testlet					
	Access		Bias		Content		Access		Bias		Content	
Accept	5,496	88%	5,421	90%	5,690	85%	1,654	91%	1,640	93%	1,716	89%
Revise	670	11%	613	10%	901	14%	151	8%	127	7%	200	10%
Reject	67	1%	15	<1%	65	1%	17	1%	3	<1%	17	1%
Total	6,233		6,049		6,656		1,822		1,770		1,933	

For mathematics, the percent of recommendations for acceptance ranged from 88% to 92%. The rate at which content was recommended for rejection ranged from <1% to 1%.

Table 4

Mathematics External Review Recommendations

	Task						Testlet					
	Access		Bias		Content		Access		Bias		Content	
Accept	9,172	90%	9,969	92%	9,101	91%	2,914	88%	2,411	91%	2,232	91%
Revise	877	9%	726	7%	885	9%	260	11%	219	8%	218	9%
Reject	125	1%	106	1%	57	<1%	27	1%	16	1%	12	<1%
Total	10,174		10,801		10,043		3,201		2,646		2,462	

Content Team Decisions

Because multiple reviewers examined each task and testlet, external review ratings were compiled across panel types. Content teams reviewed and summarized the recommendations provided by the external reviewers for each task and testlet. Decision options were broken into five categories:

1. No pattern of similar concerns; accept as-is
2. Pattern of minor concerns, will be addressed
3. Major revision needed
4. Reject
5. More information needed

Content teams documented the decision category for each task and testlet as well as the reason for a decision of 2 through 5, using the codes provided in the Appendix for each of the three panel types.

Following this process, content teams made a final decision to accept, revise, or reject each of the tasks and testlets. The number of accept, revise, and reject decisions made by each content team are included in Table 5 and Table 6.

Table 5

ELA Decisions Based on External Review

	Task		Testlet	
Accept	6,551	96%	1,709	87%
Revise	277	4%	254	13%
Reject	16	<1%	4	<1%
Total	6,844		1,967	

The ELA content team retained more than 99% of tasks and testlets sent out for external review. Of the tasks and testlets that were revised, most required only minor changes (e.g., minor rewording with the concept remaining unchanged) as opposed to major changes (e.g., stem or option replacement). Of the revisions made to ELA content, 257 (93%) task revisions and 200 (79%) testlet revisions were considered minor.

Of the 277 ELA **tasks** that were revised, 158 (57%) were flagged for a content issue. A total of 116 (42%) were flagged for an accessibility issue. No tasks were flagged solely for a bias and sensitivity issue. A total of 3 tasks (1%) were flagged for having an issue in more than one area (i.e., accessibility, bias and sensitivity, and/or content).

Of the 16 ELA **tasks** that were rejected, all 16 (100%) were flagged for a content issue; none were flagged for an accessibility issue, and none were flagged for a bias and sensitivity issue. No tasks were flagged for having an issue in more than one area.

Of the 254 ELA **testlets** that were revised, 144 (57%) were flagged for a content issue, 84 (33%) were flagged for an accessibility issue, and 6 (2%) were flagged for a bias and sensitivity issue. A total of 18 testlets (7%) were flagged for having an issue in more than one area.

Of the 4 ELA **testlets** that were rejected, all 4 (100%) were rejected for a content issue; none were rejected for an accessibility issue, and none were rejected for a bias and sensitivity issue. No testlets were flagged for having an issue in more than one area.

Table 6

Mathematics Decisions Based on External Review

	Task		Testlet	
Accept	8,824	96%	2,050	92%
Revise	344	4%	173	8%
Reject	7	<1%	2	<1%
Total	9,175		2,225	

The mathematics content team retained more than 99% of tasks and testlets sent out for external review. As with ELA, most revisions made to mathematics tasks and testlets were minor, including 284 (83%) task revisions and 155 (90%) testlet revisions.

Of the 344 mathematics **tasks** that were revised, 220 (64%) were flagged for a content issue, 112 (33%) were flagged for an accessibility issue, and 1 (<1%) was flagged for a bias and sensitivity issue. A total of 11 tasks (3%) were flagged for having an issue in more than one area.

Of the 3 mathematics **tasks** that were rejected, 3 (100%) were rejected for a content issue; none were rejected for an accessibility issue, and none were rejected for a bias and sensitivity issue. No tasks were flagged for having an issue in more than one area.

Of the 173 mathematics **testlets** that were revised, 105 (61%) were flagged for a content issue, 50 (29%) were flagged for an accessibility issue, and 1 (<1%) was flagged for an exclusive bias and sensitivity issue. A total of 17 tasks (10%) were flagged for having an issue in more than one area.

The one mathematics **testlet** that was rejected was rejected for a content issue.

English Language Arts Text Review

As a part of the external review process, reviewers also evaluated the DLM ELA texts prior to their being included in single-EE testlets.² Criteria and review processes were discussed with member states during the spring of 2014. This led to a revision of text review procedures for batches of texts reviewed between June of 2014 and November of 2014. For a summary of the previous review batches, see *Results from External Review during the 2013-2014 Academic Year*.

² Texts for multi-EE testlets are constructed at the same time as the items.

For single-EE testlets, the external review of texts was conducted separately from the external review of testlets. Each state was asked to recruit one reviewer to divide the review responsibilities so that there would be one evaluation per text, per state. DLM staff solicited reviewers who had experience reviewing ELA content or experience with students with significant cognitive disabilities.

Reviewers completed asynchronous, online training activities that included a description of the ELA text-construction process, a short video introduction to ELA texts, and a review of the criteria used in text development. Reviewers were provided with a short summary of how DLM ELA texts are created and used, a description of the internal guidelines used for text development, selection guidelines for photographic illustrations, and the *DLM Bias and Sensitivity Guidelines for ELA Texts*.

Seven batches of texts were distributed for review during the 2014–2015 cycle. All texts were provided via a secure file-sharing platform, and a spreadsheet supplied descriptions of how texts were assigned to learning map nodes for each review batch. Reviewers evaluated texts using criteria related to content, accessibility, and bias and sensitivity. Reviewers were asked to identify texts that did not meet the text development guidelines. When reviewers judged that a text did not meet the criteria, they submitted feedback through an online survey tool. Reviewers selected criteria related to their feedback, entered free-response comments about the problem, and suggested possible revisions. Reviewers also had an opportunity to recommend outright rejection of the text and to explain their reasoning.

Each batch of texts was distributed with a review deadline. After each review window closed, DLM staff evaluated the collective set of recommendations for each text. In cases where reviewers identified problems, texts were revised to enhance language clarity, to improve the logical presentation of content, and to remove bias or sensitivity issues with photos. When multiple reviewers raised the same issue regarding sensitive content in a text, the text was revised to eliminate the problem, where possible. In cases where revision would compromise the use of the text to measure its assigned node(s), the text was rejected.

A summary of reviews of ELA texts is presented in Table 7. Across all batches, 260 texts were available for review. Only 4 (2%) were rejected based on reviews.

Table 7

Summary of Reviews and Results for Each Group of Texts

	Batch*						
	D	E	F	G	H	I	J
Number of states submitting reviews	6	7	4	8	3	9	6
Total texts available for review	9	76	21	44	34	34	42
Number of texts receiving a comment	9	60	19	41	27	34	33
Total comments received	47	112	52	115	71	111	56
Number of texts recommended for rejection	7	2	1	0	2	3	3
Number of texts revised based on reviews	2	51	15	29	18	6	14
Number of texts rejected based on reviews	1	1	0	0	2	0	0

*For batches A through C, see the 2013-2014 report.

Conclusion

The vast majority of content externally reviewed during the 2014–2015 academic year was either accepted outright or was accepted with revisions. Less than 1% of English language arts and mathematics tasks and testlets were rejected. This finding supports the DLM system’s approach of evidence-centered design, whereby tasks and testlets are specifically created using Essential Element Concept Maps with accessibility, bias and sensitivity, and content considerations in mind.

Recruitment for external review for the 2015–2016 academic year was modified based on the findings from 2014–2015. In order to combat attrition, a smaller pool of reviewers was solicited to complete more reviews more frequently. In addition, reviewers were specifically sought for review of science testlets and testlets designed for students who are blind or who have visual impairments. Because of their success in 2014–2015, Power Reviewers and Hourly Reviewers will be used again in 2015–2016. Incentives will also be kept in place for reviewers participating in external review during 2015–2016.

References

Clark, A., Karvonen, M., & Swinburne Romine, R. (2014). *Results from external review during the 2013-2014 academic year* (Technical Report No. 14-02). Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

Appendix: Review Criteria for Panelists

CRITERIA FOR ACCESSIBILITY PANELS

Tasks

1. The text within the item provides an appropriate level of challenge and maintains a link to grade-level content without introducing unnecessary, confusing, or distracting verbiage. The text uses clear language and minimizes the need for inferences and prior knowledge to comprehend the content.
2. Graphics are clear and do not introduce confusion. Graphics can be presented in tactile form.

Testlets

3. The testlet is instructionally relevant to students for whom it was written and is grade-level appropriate.
4. The testlet does not introduce barriers for students with (a) limited working memory, (b) communication disorders dependent on spoken English grammatical structures, or (c) limited implicit understandings of others' emotions and intentions.

CRITERIA FOR BIAS & SENSITIVITY PANELS

Tasks (all bias criteria)

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

Testlets (sensitivity criterion)

5. Testlet is free of content that is controversial, disturbing, or emotionally charged due to issues of culture, region, gender, religion, ethnicity, socioeconomic status, occupation, or current events.

CRITERIA FOR CONTENT PANELS

Tasks

1. The item assesses the content of the targeted node.
2. The level of Depth of Knowledge required in the node matches the Depth of Knowledge identified for the item.
3. The content of the item is technically correct (wording and graphics).
4. Item answer options contain only one correct answer (the key), distractors are incorrect and not misleading, and nothing in the item cues the correct response.

5. The item type is logical and appropriate for the content being assessed, and the graphics contribute to the quality of the item.

Testlets

6. The testlet is instructionally relevant to the students for whom it was written and is grade-level appropriate.
7. Embedded items are placed within the story text at logical places, and conclusion items are placed at the end (ELA only).