



**DYNAMIC**<sup>®</sup>  
LEARNING MAPS

Results from External Review during the 2015–2016 Academic Year

---

Technical Report #16-05

12/19/2016

**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., Beitling, B., Bell, B., & Karvonen, M. (2016). *Results from external review during the 2015–2016 academic year* (Technical Report No. 16-05). Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

## Table of Contents

External Reviewers.....	4
External Review Process.....	7
Results of Reviews during the 2015–2016 Academic Year.....	8
Test Development Team Decisions .....	10
Conclusion.....	13
References.....	15
Appendix: Review Criteria for Panelists .....	16

## Results from External Review during the 2015–2016 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 2015–2016 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 test development 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 regarding which tasks and testlets to field test. Three types of panels are formed, with a member from each reviewing each testlet. The three review panels include one for the review of content, one for the review of accessibility, and one for the review of bias and sensitivity criteria. For results from the 2013–2014 academic year, see Clark, Karvonen, & Swinburne Romine (2014). For results from the 2014–2015 academic year, see Clark, Swinburne Romine, Bell, & Karvonen (2015).

### External Reviewers

Volunteer reviewers were recruited during September of 2015 from DLM partner states to review testlets (each consisting of a few tasks plus an engagement activity) intended for administration during the 2015–2016 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. Recruitment for 2015–2016 was targeted at reviewers for science assessment content and for testlets designed for students who are blind or visually impaired.

During the 2015–2016 year, an incentive plan was put in place to attract reviewers and to prevent steep attrition rates as witnessed in prior years.<sup>1</sup> The incentive plan included a payment of \$30 per assignment, which was a \$10 increase from the previous year.

---

<sup>1</sup> There was an overall attrition rate of 71% for the 2014–2015 review window and 62% for the 2013–2014 review window.

DLM received information from a total of 136 individuals interested in serving as external reviewers during the 2015–2016 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 136 volunteers, 27 were ineligible for serving as external reviewers. Reasons for ineligibility included not providing qualifying information or volunteering too late to be placed on a panel. This attrition rate of 20% brought the pool of available volunteers to 109.

Of the 109 remaining volunteers, the average number of years of experience in preschool through 12th-grade education was 14 years. A total of 40% of volunteers indicated they had previous review experience for another large-scale assessment, and 45% had served as external reviewers for DLM in previous years. Table 1 describes the number of volunteers by state.

Table 1

*Number of External Review Volunteers by State*

State	<i>n</i>	%
Alaska	2	1.8
Illinois	6	5.5
Iowa	19	17.4
Kansas	23	21.1
Michigan*	1	0.9
Mississippi	2	1.8
Missouri	12	11.0
North Carolina	2	1.8
Oklahoma	25	22.9
Pennsylvania	1	0.9
Utah	1	0.9
Vermont	6	5.5
West Virginia	2	1.8
Wisconsin	7	6.4
Total	109	

\*Non-DLM state. Because volunteers were recruited through DLM state partners, it was assumed that this volunteer was intentionally recruited.

Of the 109 people who volunteered to serve as external reviewers, 38 did not complete training and therefore did not qualify for placement on a review panel. This resulted in an eligible reviewer pool of 71 and an overall attrition rate of 48%.

Eligible 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 English language arts (ELA), mathematics, or science. Regardless of panel type, all reviewers had to have access to a secure computer and agree to complete at least two rounds of reviews.

Of the 71 eligible reviewers, 58 were placed on review panels. A total of 21 reviewers were assigned to the accessibility panel, 17 were assigned to the bias and sensitivity panel, and 20 were assigned to the content 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 panels.

Reviewers were selected from each of these panels to review content area testlets in science, ELA, and mathematics. In ELA and mathematics, panelists were assigned separately for each of the two testing models (integrated and year-end) based on the state in which they teach. Of the 58 reviewers who completed review assignments, 35 were placed on integrated-model panels (22 ELA and 13 mathematics), 14 were placed on year-end-model panels (10 ELA and 4 mathematics), and 9 were placed on science 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 58 reviewers assigned to panels, 46 completed one or more review assignments, reflecting an attrition rate of 66% of the original 136 volunteer reviewers and a 21% attrition rate at this stage. In all, 27 reviewers completed all assigned reviews.

Because of the attrition rates observed in prior years, a small group of reviewers, called “Power Reviewers,” were retained from the 2014–2015 year to review for all three panel types. The three Power Reviewers were selected based on the quality of their previous work, and their availability for 2015–2016 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 for the 46 reviewers who completed reviews, the three Power Reviewers, and the two hourly reviewers.

Table 2

*Number of Reviewers by Content Area (N = 51)*

Review Type	English Language Arts	Mathematics	Science
Accessibility	7	5	2
Bias and Sensitivity	8	5	3
Content	10	5	1
All Three Types*	3	5	5
Total	28	20	14

\*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 2015–2016 Academic Year**

Most of the content reviewed during the 2015–2016 academic year was included in the instructionally embedded and spring windows. On a limited basis, content for the upcoming 2016–2017 school year was also reviewed. The reviewers’ recommendations for tasks and testlets are presented in Table 3, Table 4, and Table 5. 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 92%. The rate at which content was recommended for rejection ranged from 1% to 3%.



Table 3

*ELA External Review Recommendations*

Type	Task						Testlet					
	Access		Bias		Content		Access		Bias		Content	
Accept	4,890	89%	5,080	92%	5,539	86%	1,556	87%	1,558	88%	1,768	85%
Revise	514	9%	380	7%	752	11%	191	11%	202	11%	271	13%
Reject	121	2%	65	1%	175	3%	31	2%	20	1%	47	2%
Total	5,525		5,525		6,466		1,778		1,780		2,086	

For mathematics, the percent of recommendations for acceptance ranged from 87% to 94%. The rate at which content was recommended for rejection ranged from <1% to 1%.

Table 4

*Mathematics External Review Recommendations*

Type	Task						Testlet					
	Access		Bias		Content		Access		Bias		Content	
Accept	2,284	92%	2,494	94%	2,486	90%	562	88%	630	93%	628	87%
Revise	196	8%	143	5%	248	9%	76	12%	49	7%	84	12%
Reject	9	<1%	6	<1%	41	1%	1	<1%	1	<1%	7	1%
Total	2,489		2,643		2,775		639		680		719	

For science, the percent of recommendations for acceptance ranged from 82% to 91%. The rate at which content was recommended for rejection ranged from 1% to 3%.

Table 5

*Science External Review Recommendations*

Type	Task						Testlet					
	Access		Bias		Content		Access		Bias		Content	
Accept	837	85%	965	88%	826	82%	269	86%	319	91%	279	88%
Revise	132	13%	113	10%	154	15%	40	13%	27	8%	29	9%
Reject	21	2%	19	2%	33	3%	4	1%	3	1%	11	3%
Total	990		1,097		1,013		313		349		319	

**Test Development Team Decisions**

Because reviewers from each panel examined each task and testlet, external review ratings were compiled across panel types. The test development team for each content area 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

The test development team for each content area 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, each test development team 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 test development team are included in Table 6, Table 7, and Table 8.

Table 6

*ELA Decisions Based on External Review*

Decision	Task		Testlet	
Accept	3,555	91%	895	71%
Revise	329	8%	231	19%
Reject	13	<1%	130	10%
Total	3,897		1,256	

The ELA test development team retained more than 97% 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 answer option replacement). Of the revisions made to ELA content, all 329 task revisions and 225 (97%) testlet revisions were considered minor.

Of the 329 ELA **tasks** that were revised, 285 (87%) were flagged for a content issue. A total of 31 (9%) were flagged for an accessibility issue. A total of 11 (3%) were flagged for a bias and sensitivity issue. A total of 2 tasks (1%) were flagged for having an issue in more than one area (i.e., accessibility, bias and sensitivity, and/or content).

Of the 13 ELA **tasks** that were rejected, 10 (77%) were flagged for a content issue, and 3 (23%) were flagged for a bias and sensitivity issue. No tasks were flagged for an accessibility issue or for having an issue in more than one area.

Of the 231 ELA **testlets** that were revised, 199 (86%) were flagged for a content issue, 19 (8%) were flagged for an accessibility issue, and 10 (4%) were flagged for a bias and sensitivity issue. A total of 3 testlets (1%) were flagged for having an issue in more than one area.

Of the 130 ELA **testlets** that were rejected, 11 (8%) were flagged for a content issue, 1 (1%) was flagged for an accessibility issue, and 118 (91%) were flagged for a bias and sensitivity issue. No testlets were flagged for having an issue in more than one area.

Table 7

*Mathematics Decisions Based on External Review*

Decision	Task		Testlet	
Accept	572	40%	90	24%
Revise	411	29%	172	45%
Reject	451	31%	119	31%
Total	1,434		381	

The mathematics test development team retained 69% of tasks and testlets sent out for external review. As with ELA, most revisions made to mathematics tasks and testlets were minor, including 410 (99%) task revisions and 171 (99%) testlet revisions.

Of the 411 mathematics **tasks** that were revised, 189 (46%) were flagged for a content issue, 132 (32%) were flagged for an accessibility issue, and 46 (11%) were flagged for a bias and sensitivity issue. A total of 44 tasks (11%) were flagged for having an issue in more than one area.

Of the 451 mathematics **tasks** that were rejected, 321 (71%) were rejected for a content issue, 58 (13%) were rejected for an accessibility issue, and none were rejected for a bias and sensitivity issue. A total of 72 tasks (16%) were flagged for having an issue in more than one area.

Of the 172 mathematics **testlets** that were revised, 80 (47%) were flagged for a content issue, 35 (20%) were flagged for an accessibility issue, and 14 (8%) were flagged for a bias and sensitivity issue. A total of 43 testlets (25%) were flagged for having an issue in more than one area.

Of the 119 mathematics **testlets** that were rejected, 87 (73%) were rejected for a content issue, 14 (12%) were rejected for an accessibility issue, and none were rejected for a bias and sensitivity issue. A total of 18 testlets (15%) were flagged for having an issue in more than one area.

Table 8

*Science Decisions Based on External Review*

	Task		Testlet	
Accept	551	86%	147	73%
Revise	87	13%	55	27%
Reject	4	1%	0	0%
<b>Total</b>	<b>642</b>		<b>202</b>	

The science test development team retained more than 99% of tasks and testlets sent out for external review. As with ELA and mathematics, most revisions made to science tasks and testlets were minor, including 85 (98%) task revisions and 52 (95%) testlet revisions.

Of the 87 science **tasks** that were revised, 34 (39%) were flagged for a content issue and 53 (61%) were flagged for an accessibility issue. No tasks were flagged for a bias and sensitivity issue or for having an issue in more than one area.

Of the 4 science **tasks** that were rejected, 2 (50%) were rejected for a content issue, 1 (25%) was rejected for an accessibility issue, and 1 (25%) was rejected for a bias and sensitivity issue. No tasks were flagged for having an issue in more than one area.

Of the 55 science **testlets** that were revised, 20 (36%) were flagged for a content issue, 32 (58%) were flagged for an accessibility issue, and 2 (4%) was flagged for a bias and sensitivity issue. Additionally, 1 testlet (2%) was flagged for having an issue in more than one area.

### Conclusion

The external review process provides a useful review of content by outside panelists in the areas of content, accessibility, and bias and sensitivity. Based on this review, most tasks and testlets reviewed during the 2015–2016 academic year were either accepted outright or accepted with revisions. In ELA and science, only around 1% of tasks and testlets were rejected, providing support for 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 2016–2017 academic year will follow the same method used in 2015–2016, including the use of a smaller pool of reviewers to complete more reviews more frequently. Because of their success in 2014–2015, Power Reviewers and Hourly Reviewers were used in 2015–2016 and will be retained for the 2016–2017 academic year as well. The increased reviewer incentive from \$20 to \$30 per assignment for 2015–2016 will also be kept in place for reviewers participating in external review during 2016–2017.

## 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.

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.

## Appendix: Review Criteria for Panelists

### CRITERIA FOR ACCESSIBILITY PANEL

#### 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 PANEL

#### 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 PANEL

#### 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).