



DYNAMIC[®]
LEARNING MAPS

*2015 Integrated Model Standard
Setting: English Language Arts and
Mathematics*

Technical Report #15-02

12/17/2015

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

Karvonen, M., Clark, A., & Nash, B. (2015). *2015 integrated model standard setting: English language arts and mathematics* (Technical Report No. 15-02). 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.

Contents

List of Figures	6
List of Tables	6
Executive Summary	7
Chapter 1: Introduction	11
Overview of DLM Assessment Design.....	11
Assessment Content	11
Assessment Design and Delivery	16
Scoring	17
Performance Levels and Policy Performance Level Descriptors.....	17
Organization of the Report.....	18
Chapter 2: Standard Setting Methods	19
Rationale and General Approach	19
Determining Mastery and Developing Profiles	20
Define Mastery at the Node Level.....	21
Determine Linkage Level Mastery	21
Develop Profiles of EE/Linkage Level Mastery	23
Profile Selection	23
Profile Rating Procedures	24
Statistical Analysis of Panelist Ratings	25
Impact Data	26
Evaluation Procedures.....	26
Procedural Criteria	26
Internal Criteria	27
External Criteria.....	27
Chapter 3: Standard Setting Panel Meeting Preparation.....	28
Panelist Recruitment	28
Selection of Panel Participants.....	28
Forming Panels	28
Panelist Characteristics	29
Panel Facilitator Training	31
Chapter 4: Standard Setting Panel Meeting Procedures	32
Panelist Training.....	32
Advance Panelist Training	32

On-Site Panelist Training.....	33
Materials	35
Panelist Resources	35
Training Materials.....	35
Range-Finding Materials.....	35
Pinpointing Materials.....	36
Rating Forms	37
Evaluation Forms	37
Procedures	37
Range Finding	38
Pinpointing.....	39
Independent Evaluations of Panel-Recommended Cut Points.....	40
Chapter 5: Results	41
Panel-Recommended Cut Points	41
Convergence.....	42
Evaluations	42
Independent Panelist Evaluations of Panel-Recommended Cut Points.....	43
Panelist Evaluations of the Meeting	46
Technical Advisory Panel (TAC) Evaluation of Panel Process.....	49
Impact Data Based on Panel-Recommended Cut Points	49
Statistical Adjustment	50
Procedure.....	50
Adjusted Cut Points.....	51
Impact Data Based On Adjusted Cut Points.....	52
Final Results	54
Chapter 6: Future Steps.....	54
References.....	55
Appendix A: DLM 2014-2015 Scoring Model Description.....	56
Appendix B: Rationale for Determining Mastery and Developing Profiles.....	58
Step One: Mastery at the Node Level	58
Step Two: Linkage Level Mastery.....	59
Appendix C: Sample Profile Based on Judgments about Linkage Levels: Grade 7 ELA	60
Appendix D: Number of Essential Elements and Number of Linkage Levels Available per Grade and Subject, Integrated Model	65
Appendix E: Standard Setting Panelist Recruitment Letter and Survey.....	66

Appendix F: Panels and Assignments.....	71
Appendix G: Panel Training and Materials.....	72
Appendix H: Example Rating Forms for Range-Finding and Pinpointing	168
Appendix I: Independent Evaluation and Meeting Evaluation Forms.....	170
Appendix J: Original and Revised Standard Setting Procedures	173
Appendix K: Convergence Plots for Range-Finding and Pinpointing Ratings.....	174
Appendix L: TAC Resolution on DLM Standard Setting.....	191
Independent Evaluations of Panel-Recommended Cut Points.....	196
Panelist Evaluations of the Meeting	197

List of Figures

Figure 1. IM ELA Impact Data Using Adjusted Cut Points	10
Figure 2. IM Mathematics Impact Data Using Adjusted Cut Points.....	10
Figure 3. Section of the DLM ELA Learning Map for the Conceptual Area, Constructing Understandings of Text. Nodes associated with Essential Elements are circled.	13
Figure 4. Relationship between nodes at linkage levels and items in testlets.	16
Figure 7. Hypothetical student’s node mastery probabilities for a single linkage level.	22
Figure 6. IM English Language Arts Impact Data Using Adjusted Cut Points.....	54
Figure 7. IM Mathematics Impact Data Using Adjusted Cut Points.....	54

List of Tables

Table 1. IM Adjusted Cut Point Recommendations	9
Table 2. Dynamic Learning Maps Alternate Assessment Claims and Conceptual Areas	15
Table 3. Examples of Node Mastery Probabilities and Final Linkage Level Mastery Classifications	22
Table 4. Number of Panelists Selected per State.....	29
Table 5. Number of Panelists per Panel.....	29
Table 6. Panelist Demographic Characteristic	30
Table 7. Years of Experience	30
Table 8. Number of Panelists Who Taught Students in each Disability Category	30
Table 9. Panelist Self-Assessments after Completing Advance Training	33
Table 10. Number of Panelists and Order in Which Cut Points Were Set by Panel.....	38
Table 11. Final ELA and Math IM Panel Cut Point Recommendations	41
Table 12. Standard Errors for ELA Final Pinpointing Ratings	43
Table 13. Standard Errors for Math Final Pinpointing Ratings	43
Table 14. Panelist Comfort with Group Recommended Cut Points.....	43
Table 15. Panels with Median Independent Cut Point Different from Group-Recommended Cut Point ..	45
Table 16. Standard Errors for IM ELA Independent Panelist-Recommended Cut Points	46
Table 17. Standard Errors for IM Math Independent Panelist-Recommended Cut Points	46
Table 18. Percentages of IM Panelist Responses to Evaluation Items	48
Table 19. Percent of IM Students at Each Performance Level.....	49
Table 20. IM Adjusted Cut-Point Recommendations.....	52
Table 21. Percent of Students at Each Performance Level Based on IM Adjusted Cut Points	53
Table 22. Percent of Students Estimated as Node Masters Using Administered Testlet Data Based on the Probability that Indicates Mastery.....	58
Table 23. Number of Nodes per Linkage Level	59

Executive Summary

The Dynamic Learning Map™ Alternate Assessment standard setting event was conducted from June 15 – 18, 2015, in Kansas City, Missouri, following the first operational testing year in English language arts (ELA) and mathematics. The standard setting was a Dynamic Learning Maps (DLM®) consortium-wide event with the purpose of setting a set of recommended cut points for each of the consortium's two testing models: integrated and year end. Standards for the integrated model states are based on data collected during instructionally embedded and spring testing windows, whereas the standards for year-end model states are based on the spring testing window alone. The report that follows includes the findings for the integrated model. See Technical Report #15-03 for results for only the year-end model. State partners had the flexibility of either adopting the DLM-recommended cut points for their own state needs or developing their own independent set of cut points.

Panels consisting of representatives from partner states convened to recommend cut points. Separate panels were formed for each grade or course and by content area. Three cut points were determined at each grade level for each content area to differentiate between four performance levels.

A standard setting approach was implemented to classify student performance into four different levels: emerging, approaching the target, at target, and advanced. The approach was derived from existing methods, including generalized holistic and body of work, but modified to fit DLM's assessment design and scoring system. For DLM, the standard setting approach leveraged the nodes in the learning maps and the statistical modeling approach for determining student mastery classifications. DLM used a student profile approach to classify student mastery into performance levels. Profiles provided a holistic view of student performance across the Essential Elements (i.e., content standards) and linkage levels (i.e., levels of complexity for each Essential Element). Cut points were determined by evaluating the total number of linkage levels mastered, similar to assigning a cut point along a scale score continuum.

Student profiles were developed to show student mastery (mastered/not mastered) for each of the five linkage levels for each Essential Element. There were three steps to determine an overall classification of student mastery. The first two steps used criteria for determining node mastery classifications based on students' item responses and linkage level mastery classifications based on aggregations of node mastery. The third step was to calculate total numbers of linkage levels mastered in the subject. Profiles were then selected based on these values to be used as exemplars for standard setting.

Panelists were recruited to participate in the standard setting event from all DLM partner states, across all assessed grade levels, and within both content areas. The majority of

panelists were educators with experience in ELA or mathematics and/or in teaching students with significant cognitive disabilities. Six integrated model panels were created from the pool of volunteers, with representation spread across the states. Once panel selections were complete, panelists completed an online training module before the on-site standard setting event. This training provided a general overview of the DLM assessment system and was supplemented by additional on-site training on the standard setting panel procedures. Once on site, panelists were familiarized with the standard setting materials and methods, and then were given folders containing exemplars of student profiles to practice the rating process.

The standard setting process followed two basic steps: range finding and pinpointing. The purpose of range finding was for panelists to assign general divisions between performance levels after reviewing a limited set of exemplar profiles. After panelists determined the ranges of profiles where cut points were likely to be found, they moved on to the pinpointing process. During pinpointing, additional profiles were provided at levels within the range determined from the range-finding process. The purpose of pinpointing was for panelists to evaluate the additional exemplar profiles and hone in on specific cut points to distinguish the four performance levels. Within the range-finding and pinpointing phases, panelists had multiple opportunities to make independent evaluations. Further, panelists were asked to provide feedback as to their comfort level with their group's recommended cut points and independently indicate a final recommended cut point if they were dissatisfied with the group's results.

By the end of the last day of the standard setting event, all panel-recommended cut points had successfully been identified. In all but one instance, the median individual recommended cut points and the group recommended cut point were the same for the integrated model. This suggests that overall the group process was effective for using expert judgment to classify student profiles into the DLM performance levels and identify corresponding cut points. Evaluations based on panelists' experience with DLM standard setting and DLM Technical Advisory Committee (TAC) members' review of processes and outcomes provide further evidence that the methods and process used were effective for achieving the goals of the meeting.

Following the standard setting event, a statistical adjustment technique was applied to reduce the impact of panelist sampling on the cut points. Impact data was used to evaluate the distributions of students in each performance level category, with and without the DLM staff-recommended adjustments. The adjusted cut points and impact data were then presented to the DLM TAC as well as the state partners for review. After review, the TAC provided support for the statistical adjustment technique, and the state partners from both models accepted the adjusted cut points.

The final set of adjusted cut points and impact data follow by subject area for the integrated model.

Table 1. IM Adjusted Cut Point Recommendations

Content Area and Grade	Emerging/ Approaching	Approaching/ Target	Target/ Advanced	Minimum Required Linkage Levels
ELA				
3	9	16	30	40
4	11	18	37	45
5	8	16	33	40
6	9	20	36	45
7	13	27	42	55
8	17	27	41	55
9-10	18	31	41	50
11-12	15	30	40	50
Math				
3	7	17	25	30
4	9	21	31	40
5	9	20	29	35
6	8	17	24	30
7	12	21	29	35
8	11	20	30	35
9	8	15	22	30
10	5	13	22	30
11	7	15	23	30

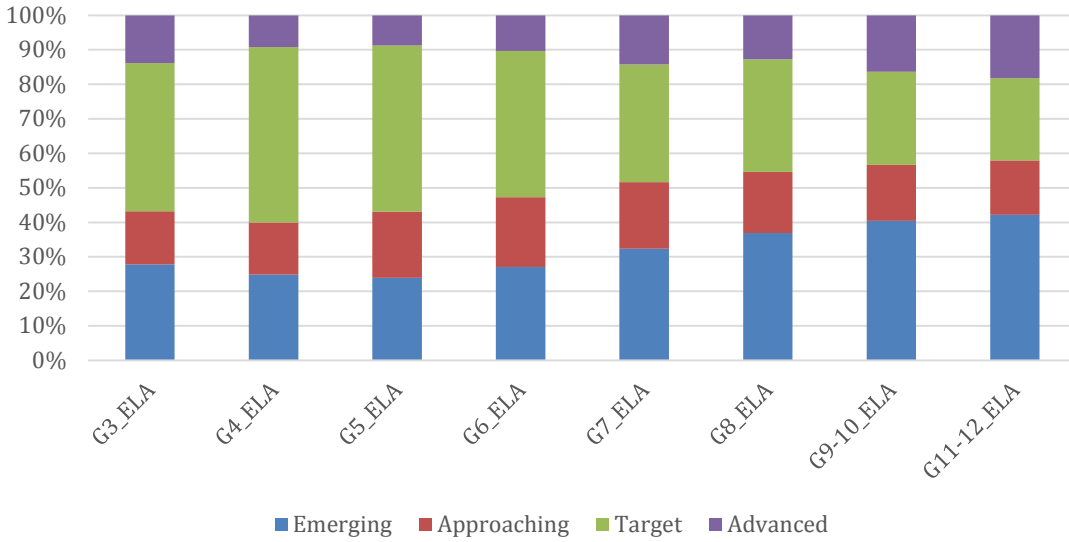


Figure 1. IM ELA Impact Data Using Adjusted Cut Points

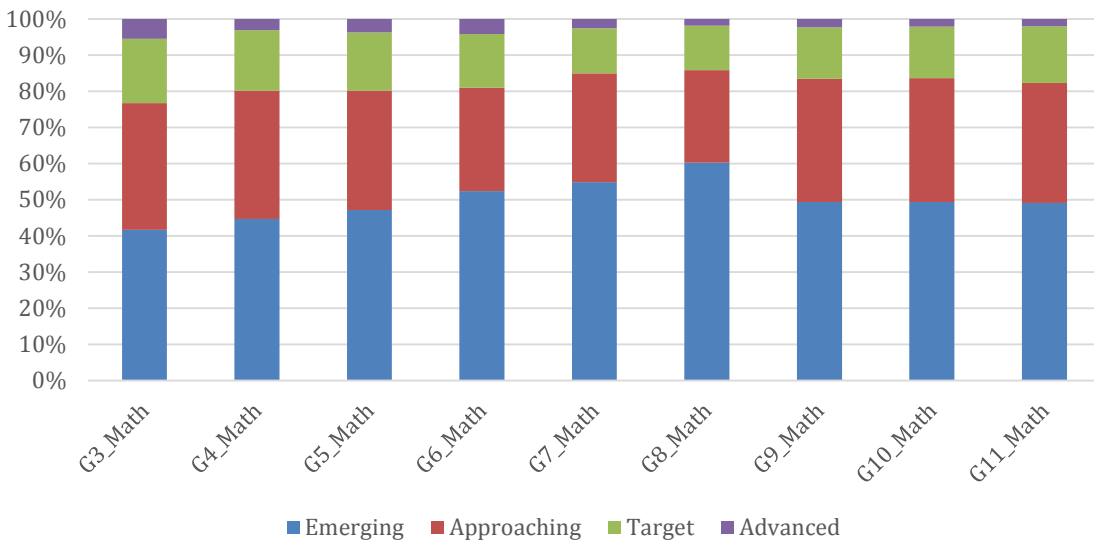


Figure 2. IM Mathematics Impact Data Using Adjusted Cut Points

Chapter 1: Introduction

The standard setting process for the DLM assessments in ELA and mathematics consisted of the development of the performance level descriptors (PLDs), the four-day standard setting meeting, and the follow-up evaluation of impact data and cut points. This report provides an overview of the DLM assessment system including the development of the PLDs and details the methods, preparation, procedures, and results of the standard setting meeting, including the follow-up evaluation of the impact data and cut points.

The purpose of the standard setting activities was to derive recommended cut points for placing students into four performance levels based on results from the 2014-2015 DLM assessments in English language arts and mathematics. The intended audiences for this standard setting technical report are DLM's technical advisory committee (TAC), DLM state partners' state boards of education, and federal peer review committee members.

The 2014-2015 school year was the first fully operational testing year for the DLM assessments in ELA and mathematics. The consortium operational testing window ended on June 12, 2015, and standard setting was conducted from June 15 – 18, 2015, in Kansas City, Missouri. The standard setting event was a DLM consortium-wide event with the purpose of establishing a set of cut points for each testing model. Although state partners voted on acceptance of final cut points for each model, individual states had the option to adopt the consortium cut points or develop their own independent cut points.

There are two groups of states within the DLM consortium who use two testing models with differences in blueprint coverage of the Essential Elements: the integrated model (IM) and the year-end model (YE). The same standard setting methods were used for both models but used separate panels consisting of representatives from either IM or YE states. All judgments about cut points were made separately for each model. Results for the integrated model are included in this report.

Overview of DLM Assessment Design

Assessment Content

The DLM assessment system is based on large, fine-grained learning maps. Learning maps not only specify assessment content, but also the relationships and learning pathways among different concepts, knowledge, and cognition. These learning maps provide a framework to support inferences about student learning needs and support multiple pathways to learning.

Nodes in the maps represent discrete knowledge, skills, and understandings in either ELA or mathematics, as well as important foundational skills. As of June 2015, there are 1,900 nodes in the ELA map, 2,399 nodes in the mathematics map, and 141 foundational nodes

that are associated with both content area maps. The maps go beyond traditional learning progressions to include multiple and alternate pathways by which students may develop content knowledge. As of June 2015, there are 5,039 connections in the ELA map and more than 5,200 connections in the mathematics map.

Within these fine-grained learning maps, there is a subset of nodes that serve as content standards for the assessment. In DLM these nodes are called the Essential Elements (EEs). The DLM EEs are specific statements of knowledge and skills linked to the grade-level expectations identified in college and career readiness standards. The purpose of the EEs is to build a bridge from those content standards to academic expectations for students with the most significant cognitive disabilities.

The EEs specify academic targets, while the learning map clarifies how students can reach those targets. Figure 3 shows a high-level overview of a section of the ELA learning map that deals with constructing understandings of text. Within this large conceptual area, there are a number of nodes associated with EEs, outlined in red. Multiple pathways are available to move from one EE to another within this single conceptual area.

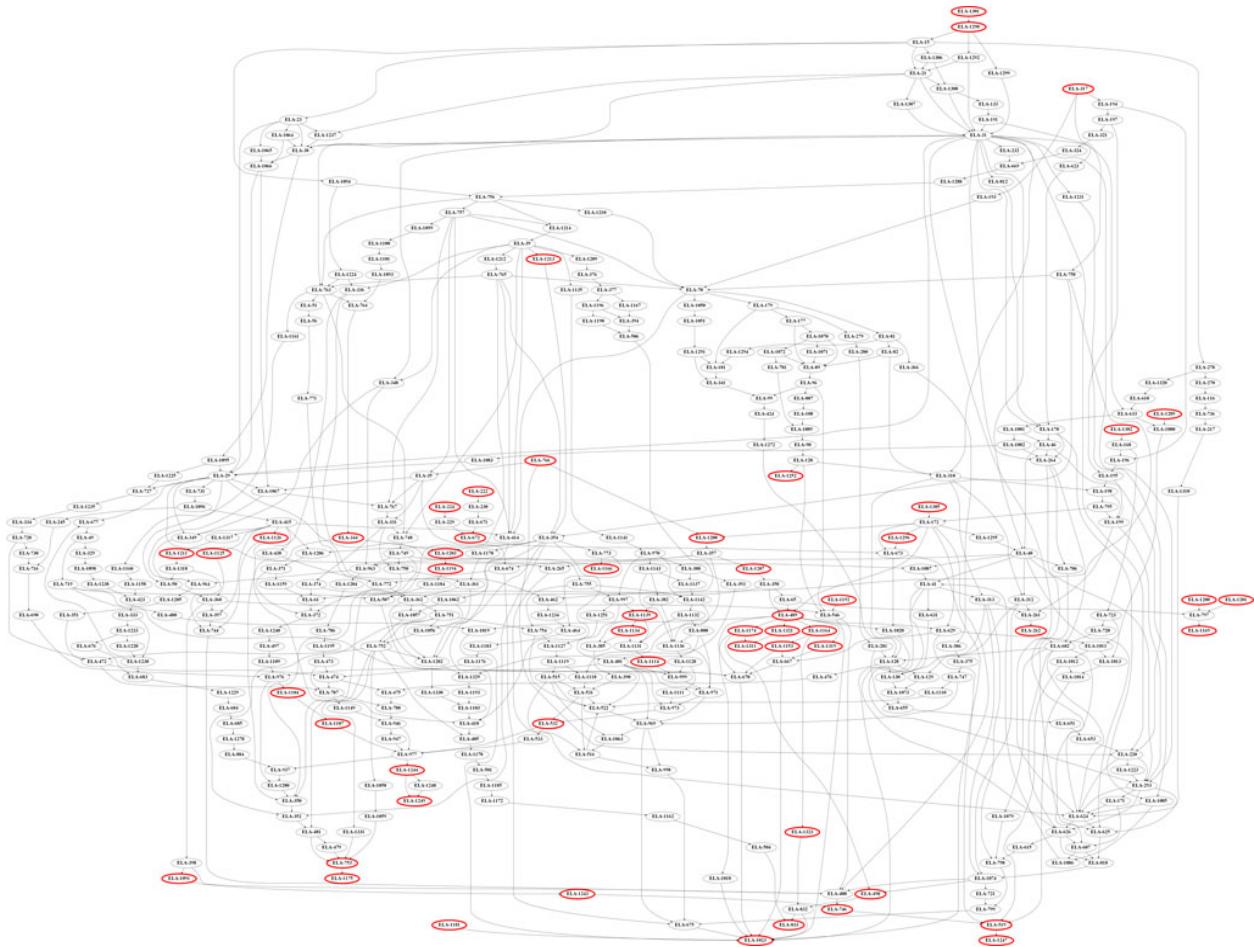


Figure 3. Section of the DLM ELA Learning Map for the Conceptual Area, Constructing Understandings of Text. Nodes associated with Essential Elements are circled.

For each EE, small collections of nodes are identified and grouped into linkage levels. A linkage level contains one or more nodes that represent critical points in the development of the knowledge and skills represented in the EE. The fourth level, the Target, aligns directly with the EE. There are three levels below the Target: Initial Precursor, Distal Precursor, and Proximal Precursor. The fifth level, the Successor, extends beyond the EE and toward the grade-level content standard for students without significant cognitive disabilities. See the following example of the relationship between an EE and nodes at different linkage levels.

DLM Essential Element	Node Linkage Progression
M.EE.7.G.5 Recognize angles that are acute, obtuse, and right	<p>Successor Node:</p> <ul style="list-style-type: none"> ● Compare angles to right angle <p>Target Nodes:</p> <ul style="list-style-type: none"> ● Recognize acute angles ● Recognize right angles ● Recognize obtuse angles <p>Proximal Precursor:</p> <ul style="list-style-type: none"> ● Recognize angle <p>Distal Precursor:</p> <ul style="list-style-type: none"> ● Recognize point ● Recognize ray <p>Initial Precursor:</p> <ul style="list-style-type: none"> ● Recognize attribute values

To organize the contents of the learning maps, there are four broad claims in each subject. Claims are statements about what students are expected to learn and be able to demonstrate. The claims organize the learning map so that related EEs are meaningfully linked together. Claims are divided into smaller groups of nodes called Conceptual Areas. The nodes and EEs in a conceptual area are more closely related than those in the larger claim. Conceptual areas are organized around common cognitive processes. There are nine conceptual areas in each subject. Table 2 presents the DLM claims and conceptual areas.

Table 2. Dynamic Learning Maps Alternate Assessment Claims and Conceptual Areas

Subject	Major Claims	Conceptual Areas
Mathematics	Students demonstrate increasingly complex understanding of number sense.	Understand number structures (counting, place value, fraction)
		Compare, compose, and decompose numbers and sets
		Calculate accurately and efficiently using simple arithmetic operations
	Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles.	Understand and use geometric properties of two- and three-dimensional shapes
		Solve problems involving area, perimeter, and volume
	Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures.	Understand and use measurement principles and units of measure
		Represent and interpret data displays
Students solve increasingly complex mathematical problems, making productive use of algebra and functions.	Use operations and models to solve problems	
	Understand patterns and functional thinking	
English Language Arts	Students can comprehend text in increasingly complex ways.	Determine critical elements of text
		Construct understandings of text
		Integrate ideas and information from text
	Students can produce writing for a range of purposes and audiences.	Use writing to communicate
		Integrate ideas and information in writing
	Students can communicate for a range of purposes and audiences.	Use language to communicate with others
		Clarify and contribute to discussion
	Students can investigate topics and present information.	Use sources and information
		Collaborate and present ideas

Assessment Design and Delivery

Each content-area/grade-level assessment is designed to assess a specific set of Essential Elements. The EEs included in each blueprint can be found at <http://dynamiclearningmaps.org/>.

DLM assessments are delivered in testlets. Each testlet is comprised of items that align with nodes at a particular linkage level, as illustrated in Figure 4.

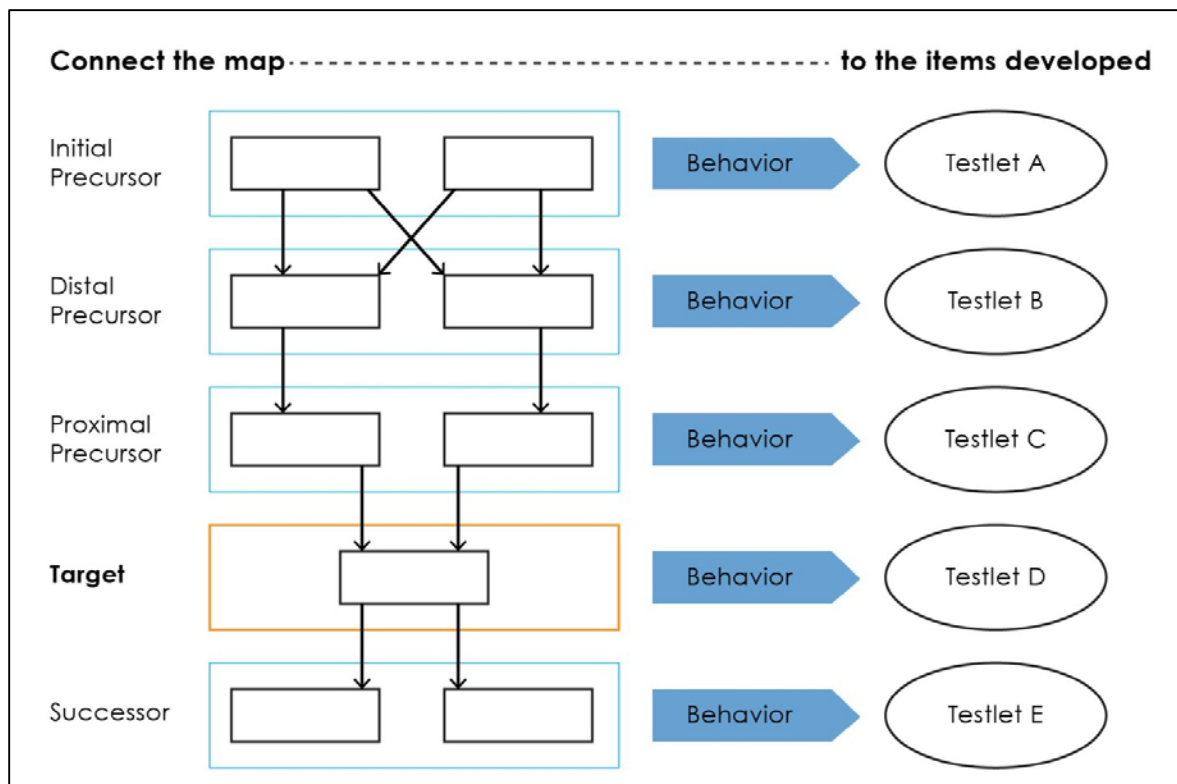


Figure 4. Relationship between nodes at linkage levels and items in testlets.

In reading and math, testlets are based on nodes for one or more EEs. Each testlet contains an engagement activity and three to eight questions. Writing testlets cover multiple EEs. Each writing testlet is a structured writing activity with several items the educator completes about the student's writing process and product.

For the integrated model, teachers select EEs and linkage levels during the instructionally embedded assessment to cover the set of requirements specified in the blueprint. During the spring window, the system selects a total of five testlets for the student. The system first ensures blueprint coverage has been met during the embedded window, then selects EEs the student previously tested on to fill the remaining testlet slots. For ELA, all students are assessed on writing during the spring window.

Scoring

Diagnostic Classification Modeling (DCM) is used to translate student responses to items into judgments about student mastery for each linkage level. For 2014-2015, students were considered masters of a linkage level if either: (1) their posterior probability from the DCM was greater than or equal to 0.8, or (2) the proportion of items that they answered correctly within the linkage level was greater than or equal to 0.8. These mastery status values were then aggregated within and across EEs to obtain the total number of linkage levels mastered within the content area.¹ Although the total number of mastered linkage levels is not a raw or scale score and should not be interpreted as an interval scale, the number of linkage levels mastered across EEs assessed was the metric translated into performance levels. Profiles used for standard setting were categorized by the number of linkage levels mastered across EEs in the subject. Further details on the development of profiles and the profile evaluation process are provided in subsequent sections.

Performance Levels and Policy Performance Level Descriptors

DLM state partners developed policy PLDs through a series of conversations and draft PLD reviews between July and December 2014. In July 2014, the state partners discussed general concepts that should be reflected in the PLDs and reviewed several examples of descriptors for three, four, and five performance levels. In fall 2014, the state partners indicated the number of levels they would require and gave feedback on additional iterations of PLDs that had been revised based on previous input. By December 2014, the PLDs were finalized. All states participating in the 2014-2015 operational assessment required four performance levels. The final version of policy PLDs are summarized below. The consortium-level definition of proficiency was at target.

¹ More information about the psychometric model used for 2014-15 operational scoring is provided in Appendix

Performance Level Descriptors
The student demonstrates <i>emerging</i> understanding of and ability to apply content knowledge and skills represented by the Essential Elements.
The student’s understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements is <i>approaching the target</i> .
The student’s understanding of and ability to apply content knowledge and skills represented by the Essential Elements is <i>at target</i> .
The student demonstrates <i>advanced</i> understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements.

Policy PLDs served as anchors for the standard setting panelists during the panel process. This procedure is described in Chapter 2. The planned process for developing grade- and content-specific PLDs is described further in Chapter 6.

Organization of the Report

The remaining chapters of this report are organized into the following categories: methods, which includes a description of the overall approach and procedures; meeting preparation steps, which include panelist recruitment and training; results, which include panel-recommended cut points; and evaluations for panel recommendations, panelist evaluations of the meeting, calculation of impact data, statistical adjustment procedures, and final recommended cut points.

Chapter 2: Standard Setting Methods

Rationale and General Approach

There is a history of selecting a standard setting method based on the type of assessment. Because DLM is a unique alternate assessment system, the approach to standard setting was developed to be consistent with DLM's design while still relying on established methods, best practices recommended in the literature, and the *Standards on Educational and Psychological Testing* (2014).

There are several assessment design features that impacted DLM's standard setting approach. A student-based standard setting approach was judged to be more appropriate than an item-based approach for the following reasons:

- Each grade/subject has hundreds of items and there are multiple testlet forms that are designed to be fungible. Using an item-based approach with such a large set of items is time and cost prohibitive.
- Modeling is used to confirm the structure of the learning map to support the order of linkage levels. Item difficulty statistics are not used to ensure correct ordering of content.
- DLM assessments are adaptive across testlets. Considering adaptive delivery and multiple forms for each EE/linkage level, it would be rare for students to receive completely identical testing experiences.
- A student-based approach supports the panelists' ability to make judgments about the student's mastery of the full range of skills rather than performance on a limited subset of items.

For DLM, the standard setting approach leverages the learning map and mastery classifications. The panel process draws from several established methods, including generalized holistic (Cizek & Bunch, 2006) and body of work (Kingston & Tiemann, 2012) but is unique to DLM. Other holistic approaches, such as the performance profile method (Perie & Thurlow, 2011), which takes into account the specific content mastered, would have been difficult to apply due to the large number of Essential Elements being evaluated and DLM's goal of reporting an overall performance level for each subject rather than subscores.

DLM's standard setting approach relied on aggregation of dichotomous classifications of mastery of the knowledge and skills across EEs in the blueprint. This is different from assessments that use score scales, where standard setting involves identifying cut scores that are imposed on a theoretical, unidimensional continuum of knowledge in a subject.

Drawing from the generalized holistic and body of work methods, DLM used a profile approach to classify student mastery into performance levels. Profiles provided a holistic view of student performance by summarizing across the Essential Elements and linkage

levels. Cut points were determined by evaluating the total number of linkage levels mastered. Although the number of linkage levels mastered is not an interval scale, the process for identifying DLM cut points is roughly analogous to assigning a cut point along a scale score continuum.

Before making a final decision whether to use the profile approach, the DLM TAC reviewed a preliminary description of the proposed methods. At the TAC's suggestion, DLM staff conducted a mock panel process using this profile-based approach to evaluate the feasibility of the rating task and the likelihood of obtaining sound judgments using this method.

Although DLM's standard setting approach is a unique hybrid of existing methods, the guidance in the *Standards for Educational and Psychological Testing* and recommended practices for developing, implementing, evaluating, and documenting the standard setting was followed (Cizek, 1996; Hambleton, Pitoniak, & Copella, 2012). For example, this report summarizes the rationale and procedures used to establish cut points (Standard 5.21), including evidence that the judgmental process should be designed so that the participants providing the judgments can bring their knowledge and experience to bear in a reasonable way (Standard 5.23).

The following steps were used in the process and are described more fully in subsequent sections of this report.

1. Determining mastery and developing profiles
 - a. Define mastery at the node level
 - b. Determine linkage level mastery
 - c. Develop Profiles of EE/Linkage Level Mastery
2. Profile selection
3. Panelist profile ratings
4. Statistical analysis of panelist ratings
5. Review of impact data
6. Adjusting cut points for cross-grade consistency

Determining Mastery and Developing Profiles

Because of the unique nature of the DLM assessment, student performance on nodes in the learning map must be aggregated to create profiles of student learning. There were three steps in the standard setting process to go from node posterior probabilities derived from the DCM approach to an overall performance classification in the subject. The first two steps addressed mastery thresholds that can be applied wholesale, and the third step set performance level cut points using the panel process. Descriptions of the criteria used in each step are provided in the sections that follow. Further detail on the rationale for the criteria in steps one and two may be found in Appendix B.

Define Mastery at the Node Level

Mastery classifications were derived from applying an agreed-upon criterion to students' posterior probabilities from the DCM calibration. These posterior probabilities are referred to as node mastery probabilities.

A student's classification as master or non-master is made with a level of certainty that is reflected in the posterior probability. The acceptable level of certainty (i.e., the posterior probability) must be identified before additional judgments can be made in the standard setting process. Maximum uncertainty occurs when the probability is 0.5, and maximum certainty when the probability approaches 0 or 1. Considering the risk of false positives and negatives, the threshold used to determine mastery classification was 0.8. That is, students with node mastery probabilities ≥ 0.8 were considered masters of the node while students with probabilities < 0.8 were considered non-masters of the node. This threshold was determined based on conversations with the TAC and informed by examination of distributions of probabilities. See Appendix B for rationale behind a 0.8 mastery threshold.

For each node, a mastery status of 0 or 1 is obtained based on the student's probability of node mastery. Using 0.8 as the cutoff for **node** mastery, all students with a probability greater than or equal to 0.8 would receive a **node** mastery status of 1, or mastered. All students with a probability lower than 0.8 would receive a **node** mastery status of 0, or not mastered.

Determine Linkage Level Mastery

The dichotomous node mastery statuses described in the previous section were then summed for every node the student was assessed on at the linkage level and divided by the total number of nodes the student was assessed on at the linkage level to obtain the proportion of nodes mastered at the linkage level. This proportion represents **linkage level mastery**. The threshold used to determine **linkage level mastery** was set at 0.75.

Similar to node mastery, a mastery status of 0 or 1 was obtained for each linkage level. Using 0.75 as the cutoff for **linkage level mastery**, all students with a proportion of nodes mastered [at the linkage level] greater than or equal to 0.75 would receive a **linkage level mastery** status of 1, or mastered. All students with a proportion of nodes mastered lower than 0.75 would receive a **linkage level mastery** status of 0, or not mastered.

As an example, a hypothetical student was assessed on four nodes for one linkage level. The student's hypothetical probabilities of node mastery are included in each example node in the figure that follows.



Figure 5. Hypothetical student's node mastery probabilities for a single linkage level.

Based on these values, using 0.8 as the **node mastery** threshold would mean the hypothetical student is considered a master of three out of the four nodes assessed at this linkage level. The student's linkage level mastery would be calculated as 0.75. Using a threshold of 0.75 as the **linkage level mastery** threshold, the student would be considered a master of the linkage level.

Table 3 displays examples of how node mastery probabilities and proportions of nodes mastered are used to determine final linkage level mastery classifications.

Table 3. Examples of Node Mastery Probabilities and Final Linkage Level Mastery Classifications

Student	Node	Node Mastery Probability	Node Mastery Classification	LL Mastery	LL Mastery Classification
IIIIII	M-803	0.78	No (o)	3 of 4 nodes = 0.75	Yes (1)
	M-1015	0.89	Yes (1)		
	M-551	0.84	Yes (1)		
	M-742	0.82	Yes (1)		
222222	M-803	0.81	Yes (1)	1 of 2 nodes = 0.50	No (o)
	M-1015	0.59	No (o)		

Develop Profiles of EE/Linkage Level Mastery

The threshold values from step one and step two were applied to create **profiles of student mastery**, which summarize linkage level mastery by EE. Profiles were created using data for each content area, grade, and testing model (IM or YE). Each profile listed all the EEs from the blueprint, along with the conceptual area for each. Profiles contained between eight and twenty EEs, depending on the blueprint. The five levels of mastery were included as columns on the profile, ranging from the Initial Precursor linkage level up to the Successor level. Green shading indicated that a linkage level was mastered (the threshold was met) for students matching that profile.

Appendix C provides an example of one profile for a seventh grade student in English language arts for the integrated model. In this example, the profile was created to demonstrate the possible skills mastered for a student who has demonstrated mastery of 34 linkage levels, as evidenced by the green shading in 34 boxes.

The maximum linkage level mastery value was determined by the model's blueprint. For IM, the number of minimum required EEs in the blueprint times five linkage levels was considered the highest linkage level value. Although it is possible that IM students were assessed on and potentially demonstrated mastery of more levels than this, the use of the most common profiles during standard setting did not warrant their inclusion. Appendix D lists the number of EEs and minimum linkage levels expected for each grade and subject.

Profile Selection

In order to select exemplar profiles for inclusion in standard setting, a program was written in *R* to first calculate node mastery based on a threshold of 0.8, and then to calculate linkage level mastery based on a threshold of 0.75. Finally, the program determined the highest linkage level the student mastered for each EE and summed them to get the total linkage level mastery value. As a quality control effort, psychometric staff members ensured that the results of the program were expected based on the input data. Specifically, checks were conducted to determine that the program accurately applied mastery thresholds and correctly determined the highest linkage level mastered by the student.

Profiles were available for all students who participated in the spring window by May 15, 2015 ($n_{IM} = 14,278$). A program was written in C++ to identify the frequency with which each precise profile (i.e., pattern of linkage level mastery) occurred in this population. Based on these results, the three most common profiles were selected for each possible total linkage level mastery value (i.e., total number of linkage levels mastered) for each grade, content area, and model. For example, the program identified the three most common ways to have mastered 34 linkage levels for IM grade three mathematics. To

ensure that the exemplar profiles were not overly similar, the program identified profiles where different linkage levels were mastered for at least three EEs.

In instances where data was not available at a specific linkage level value, (e.g. no students mastered exactly 47 linkage levels for a grade and content area), profiles were based on simulated data. Simulated profiles were not distinguishable from those based on real student data. In March 2015, the TAC suggested that content experts be involved in the process of simulating data for any levels with no real data available. This recommendation was then taken to the DLM states for consortium-wide approval. All 12 states responding to the poll indicated that the recommendation was satisfactory for simulated profiles to be included in standard setting. As a result, DLM content teams used adjacent profiles for reference and created simulated profiles that represented likely patterns of mastery. Fewer than 10% of all the profiles developed were simulated.

Profile Rating Procedures

Exemplar profiles of student mastery were compiled in folders for panelist ratings. Two types of folders were prepared for standard setting: range-finding folders and pinpointing folders. After panelists familiarized themselves with performance levels during training, the range-finding process followed. The purpose of range finding was for panelists to assign general divisions between performance levels after reviewing a limited set of profiles from points along the distribution of total linkage levels mastered. These samples were comprised of profiles at intervals of five linkage levels mastered (e.g., a total of 5, 10, 15, and 20 linkage levels mastered). After panelists determined the ranges of linkage levels mastered where cut points were likely to be found, panelists completed the pinpointing process. The purpose of pinpointing was for panelists to evaluate the additional exemplar profiles with the goal of identifying specific cut points. Profiles for seven adjacent levels within the range determined from the range-finding process were distributed to the panelists for pinpointing. For both the range-finding and pinpointing phases, panelists completed an independent round of ratings, reviewed their results and discussed them, then completed a second round of independent ratings. The results at the end of the second pinpointing round served as the group-recommended cut points. Finally, panelists independently evaluated the group-recommended cut points and indicated their level of comfort with each cut point. Additional detail about these procedures is provided in the Procedures section of Chapter 4.

Statistical Analysis of Panelist Ratings

Both the range-finding and pinpointing exercises utilized logistic regression analyses to identify appropriate ranges and calculate cut points, respectively. Logistic regression models the relationship between an independent variable, number of linkage levels mastered in this case, and the probability of being classified into a category, such as performance level **approaching** or above.

The primary goal of using logistic regression as the analytical procedure is to identify the number of linkage levels mastered where the likelihood of being assigned to a given performance level equals or exceeds that of being assigned to the next performance level or higher (where $p = .50$). In other words, rather than determining the likelihood of resulting in a specific level, given a number of linkage levels mastered, the goal was to find the likelihood of being assigned to a level **or higher**, given a number of linkage levels mastered. For example, if logistic regression indicated that the likelihood of panelists assigning a profile with 20 linkage levels mastered to performance level **approaching** or higher is 7 out of 12 (about 58%), it could be concluded that 20 linkage levels mastered would be a good cut point to differentiate **emerging** from **approaching**.

For pinpointing, the range of profiles was calculated by taking the value determined during range-finding plus and minus three for a total of seven different profiles each representing a different number of linkage levels mastered. Using this narrowed but more informative range of exemplar profiles, logistic regression was again used during the pinpointing process to determine the point at which the probability of being assigned to each performance category or higher was 0.5. The predicted values from this process were used as the recommended cut points for each level.

In some cases, the logistic regression analysis did not yield a useful result. Because this analysis largely depends on identifying areas of maximum disagreement between panelists across two performance categories to identify the point at which the probability is 0.5, logistic regression failed for any case where all of the panelists within a group had unanimous agreement on profile ratings. In these cases, on-site psychometricians reviewed the panelist group ratings and visually identified where the obvious inflexion point occurred. The value where the shift in ratings moved from one category to the adjacent category was used as the recommended group cut point.

The regression analyses to obtain the cut points were carried out in Excel using the same facilitator workbooks in which the original data were tallied and transformed to logistic functions. The facilitator workbooks are discussed in more detail in the Procedures section of Chapter 4.

The panelists' independent evaluations of the group-recommended cut points were summarized and evaluated using descriptive statistics. The purpose of evaluating the

independent ratings was to identify any places where the median independent recommended cut points differed from the group recommended cut points.

Impact Data

While impact data was not available during the standard setting meeting, it was included after the meeting as part of the evaluation process with the DLM TAC and state partners. Impact data was calculated by grade, subject, and model based on total number of linkage levels mastered. The percent of students who would be classified at each performance level based on the panelists' recommended cut points was calculated and presented first to the TAC and then to the state partners.

State partners for each model served as the policy group reviewing impact data. The state partners, who are members of the DLM consortium governing board, have varying roles within the special education and assessment departments in the state education agencies. These partners were not only knowledgeable of the DLM assessment system, but also of their own states' educational policies and student populations. State partners discussed recommended cut points and impact data with their internal stakeholders and reviewed input from the DLM TAC before participating in consortium-level discussions. Additional details regarding recommended cut points, impact data, and cut point adjustments are provided in Chapter 5.

Evaluation Procedures

The standard setting procedures were evaluated using procedural, internal, and external criteria as described by Hambleton & Pitoniak (2006). Each category contains several sub-categories. Relevant sub-categories are addressed individually.

Procedural Criteria

Explicitness. The standard setting process was explicitly defined prior to the standard setting event. Facilitators used a guide with detailed instructions for each step in the process. As part of the training for the event, all facilitators went through a mock standard setting using the intended process to ensure that there was an understanding of how the process should occur.

Practicability. To evaluate the use of the intended standard setting approach, a mock panel convened to test the process and evaluate its ease of use and likelihood of generating the intended results. In instances where the outlined procedures were inadequate (e.g., the logistic regression failed due to unanimous panelists recommendations), solutions were quickly implemented without creating confusion for the facilitators or panelists.

Implementation of Procedures. The selection of panelists was completed in the most objective way possible while also ensuring adequate coverage of content areas and grade levels. During the panel meeting, staff used a step-by-step guide to ensure fidelity of implementation. Where procedures were revised, staff and panelists were trained on the revisions. Additionally, DLM staff members who were not facilitating specific panels observed the standard setting event to verify that the specified procedures were being implemented correctly. Panelist selection and assignment is described in Chapter 3. The training of the panelists is detailed in Chapter 4.

Panelist Feedback. After receiving training for the standard setting event, nearly all panelists reported “Good” or “Excellent” understanding of important and relevant ideas. This included the purpose of standard setting, how DLM assesses content knowledge, and how scores are calculated and reported. Notably, no panelists reported “Poor” understanding for any of the key ideas assessed. Further details are presented in Chapter 4.

Documentation. When developing this standard setting method, documentation was kept on the proposed techniques, associated rationales, and TAC and state feedback. Documentation has also been kept on all stages of the process, including panelist recruitment and selection, training, and implementation. This technical report is largely based on source documentation.

Internal Criteria

Consistency Within Method. This standard setting event covered 40 sets of standards. It was not possible to replicate panels due to resource limitations. Instead, the variability of panelists’ final pinpointing ratings and their final independent ratings were reported. Standard errors are presented in Chapter 5.

Interpanelist Consistency. Due to the nature of the standard setting method used (i.e., logistic regression to identify areas of maximum disagreement as potential cut points), interpanelist consistency was not the desired outcome. However, there was an expectation that panelists would converge towards an increasingly narrow range of profiles to identify the cut point. Evidence of convergence is described in Chapter 5.

External Criteria

Reasonableness of Performance Levels. The panel-recommended and adjusted cut points, with the corresponding impact data, were presented to the TAC and state partners to ensure their reasonableness. Further details of this process may be found in Chapter 5.

Chapter 3: Standard Setting Panel Meeting Preparation

Panelist Recruitment

DLM drafted and distributed a recruitment letter to participating DLM states in March 2015. The recruitment letter is included in Appendix E. Participating states for standard setting included those that were operational in 2014-2015, or intended to be operational in 2015-2016. States were responsible for distributing the letter within their state to recruit potential panelists. Some states elected to distribute the list narrowly to constrain the number of potential panelists to only those they recommended. Others distributed the call more broadly within the states.

Panelists sought were those with both content knowledge and expertise in the education and outcomes of students with significant cognitive disabilities, including teachers as well as school and district administrators. Other subject matter experts, such as higher education institution faculty or state/regional educational staff, were also suggested for consideration. Employers were considered at the high school level only, specifically targeting companies that employ individuals with disabilities.

All potential panelists were asked to complete a Qualtrics survey. Survey items included basic demographic information as well as areas of expertise and years of experience. In addition, volunteer panelists were asked to indicate whether they were willing to commit to advance training (up to four hours during the first two weeks in June) and whether they would be available to attend the on-site meeting during the week of June 15, 2015. See the survey in Appendix E.

DLM received 429 total responses to the Qualtrics survey. This included volunteers from 5 integrated states, including one state transitioning from year-end to integrated, whose volunteers were eligible to serve on panels for either model.

Selection of Panel Participants

All Qualtrics responses were evaluated in April 2015 to assign volunteers to panels. Some states provided pre-approved panelist lists to guide panelist selection for their state. Diversity of experience and level of expertise were given priority in the selection of panelists. General priorities were weighed when selecting from the available panelists on the state-provided lists.

Forming Panels

Fourteen panels were created from the pool of volunteers, with representation as spread across the states as possible. Note that some states were limited in the total number of

available panelists due to fewer survey responses, incomplete survey data, declines to participate once notified, or the targeted recruitment (i.e., state pre-approval) of a limited number of qualified individuals. Similarly, because only two states participate in high school end-of-instruction (EOI) assessments (Mississippi and Oklahoma), larger numbers were necessary to fill the complete panels from these two states.

The specific standard setting assignments for each panel are listed in Appendix F.

When making panel assignments, the goal was for each panel to include at least six individuals. To allow for attrition, 8-9 panelists were assigned to each panel, balancing across states where possible. The number of panelists per panel is listed in Table 4 and Table 5.

Table 4. Number of Panelists Selected per State

State	Count
Iowa	14
Kansas	7
Missouri	12
North Dakota	8
Utah	4
TOTAL	45

Table 5. Number of Panelists per Panel

Grade Band Panel	Count
ELA	
3-5	8
6-8	7
HS	7
Math	
3-5	7
6-8	8
HS	8
Total	45

Panelist Characteristics

The 99 panelists who participated in standard setting represented varying backgrounds, as summarized in Table 6. Most of the selected panelists were classroom teachers. Panelists had a range of years of experience with mathematics, English language arts, and working with students with significant cognitive disabilities. The maximum, minimum,

and average years of experience are presented in Table 7. The number of panelists who taught or worked with students in each disability category are displayed in Table 8 (note that more than one disability category could be selected).

Table 6. Panelist Demographic Characteristic

	Count
Gender	
Female	41
Male	4
Race	
African American	1
American Indian/Alaska Native	1
Asian	0
Hispanic/Latino	0
Native Hawaiian/Pacific Islander	1
White	42
Not Disclosed	1
Professional Role	
Classroom Teacher	30
Building Administrator	7
District Staff	1
University Faculty/Staff	2
Other	5
Total	45

Table 7. Years of Experience

	<i>M</i>	Min	Max
Students with Significant Cognitive Disabilities	16.3	0.0	41.0
Mathematics	16.0	0.0	47.0
English Language Arts	16.0	0.0	40.0

Table 8. Number of Panelists Who Taught Students in each Disability Category

Disability	Integrated
Blind/Visually Impaired	23
Deaf/Hard of Hearing	23
Emotional Disability	44

Mild Cognitive Disability	43
Multiple Disabilities	42
Orthopedic Impairment	21
Other Health Impairment	41
Severe Cognitive Disability	42
Specific Learning Disability	40
Speech Impairment	37
Traumatic Brain Injury	28

Nearly half of the participants had experience with setting standards for other assessments (16). Some panelists already had experience with DLM, either from writing items (6) or externally reviewing items and testlets (8). Only three total panelists reported having less than one year or no experience with alternate assessments: one was a classroom teacher with 15 years of experience working with students with significant cognitive disabilities, one was a building administrator, and one was district staff.

Panel Facilitator Training

All staff, including facilitators, room leads, and supporting staff, participated in a one-hour orientation meeting regarding the purposes and outcomes of standard setting. Staff reviewed a high-level overview of the procedure. Following orientation, facilitators read a description of the training range-finding and pinpointing procedures. During the next training session, panel facilitators received a detailed agenda and scripts to be used for the standard setting process. A mock run-through of the standard setting process was held with all facilitators the week prior to the actual standard setting meeting. Facilitators practiced leading a group using the agenda and scripts and learned how to enter panelist ratings in the facilitator workbook. The agenda and scripts were adjusted prior to the standard setting panel meeting based on this run-through. A final training was held the Friday before the standard setting meeting to cover any remaining questions and review updated materials. Debriefs were also held each day of the panel meeting.

Chapter 4: Standard Setting Panel Meeting Procedures

Panelist Training

Advance Panelist Training

All panelists participated in a training module in advance of the standard setting meeting. The purpose of this training was to give panelists a general overview of the DLM assessment system ahead of time so that on-site training could focus on the panelists' specific grade/content area assignment and panel procedures. After introducing the purpose of standard setting and expectations for confidentiality, the advance training addressed the following topics:

1. Students who take DLM assessments
2. Content of the assessment system, including learning maps, Essential Elements, claims and conceptual areas, linkage levels, and alignment
3. Accessibility by design, including the framework for DLM's cognitive taxonomy and strategies for maximizing accessibility of the content; the use of the Personal Needs and Preferences (PNP) profile to provide accessibility supports during the assessment; and the use of First Contact Survey to determine linkage level assignment
4. Assessment design, including item types, testlet design, and sample items from various linkage levels in both subjects
5. An overview of the assessment model, including test blueprints and the timing and selection of testlets administered.
6. A high-level introduction to two topics that would be covered in more detail during on-site training: the DLM approach to scoring and reporting and the steps in the standard setting process.

The advance training was available online, on demand during the ten days prior to the standard setting meeting. All panelists completed the required training before arriving for the on-site panel meeting.

After viewing the training videos, panelists completed a survey where they rated their understanding of key topics. The results are summarized in Table 9.

Panelists reported feeling most comfortable with areas referencing the characteristics of students taking DLM assessments and what a student is expected to do when taking a testlet. Since most panelists were also teachers who administered DLM assessments, these were likely areas where they had direct experience. Panelists reported being less comfortable with the more technical aspects of the structure of the learning map, how testlets measure content, and calculation and reporting of results.

Table 9. Panelist Self-Assessments after Completing Advance Training

Understanding of:	Fair	Good	Excellent
Characteristics of students who take DLM assessments	0	17	39
Purpose of standard setting	0	24	32
Structure of the learning map	2	35	19
Expectations for maintaining security of information	0	7	49
Essential Elements and linkage levels	2	26	28
How testlets measure the intended content	0	30	26
How testlets are made accessible to students across DLM population	0	18	38
What a student is expected to do during a DLM testlet	0	13	43
How results are calculated and reported	2	41	13

Note: There were four options (i.e. Poor, Fair, Good, and Excellent) in the survey. Since no panelists chose “Poor” for any item in the survey, this column was omitted from this table in order to improve readability.

Panelists also rated their overall preparation for the next phase of training and whether their understanding was sufficient to make judgments about student results. All panelists ranked themselves as either **very prepared** (50) or **somewhat prepared** (6) for the next training at standard setting, and 100% of panelists believed their knowledge to be sufficient to make judgments about student performance and assessment results.

On-Site Panelist Training

Additional panelist training was conducted on site. The purposes of on-site training were twofold: (1) to review advance training concepts that panelists had indicated less comfort with, and (2) to prepare panelists for their responsibilities during the panel meeting. Since the majority of panelists indicated a high degree of comfort with advance training concepts, the first part of on-site training was a high-level review of expectations for confidentiality and test security, the organization of academic content, and testlet design.

Prior to training on the standard setting procedures, panelists were prompted to ask questions about any of the topics from the advance training.

Training on the standard setting panel procedures included the following topics:

1. How results are calculated and displayed in mastery profiles for standard setting, including guidance about appropriate interpretations of the contents of mastery profiles
2. An overview of the standard setting process including the policy PLDs, terms used during the standard setting process, the key question panelists would ask themselves when completing ratings, and the range-finding and pinpointing procedures
3. An overview of the week's activities, from training to final evaluation
4. Roles and responsibilities of everyone present for the panel meeting
5. Discussion of the contents and use of the policy PLDs
6. Presentation of the resource materials panelists should refer to when familiarizing themselves with mastery profiles

After the large group presentation on these topics, the trainer introduced the practice activity to be completed at each panel table. The training activity consisted of range finding using training profiles for just a few total linkage levels mastered (e.g., 5, 10, 15, 20). Each table trained using sample profiles for the first grade/course for which the panelists would be setting standards. Table facilitators walked panelists through the process of using their resource materials to familiarize themselves with the Essential Elements and linkage levels for that grade/course. Once panelists were ready, the facilitator then introduced the contents of the training folder (i.e., the training profiles and rating forms) and reminded panelists how to complete the rating form. Once all panelists completed the practice activity, there were opportunities to debrief at the table and through group discussion for all tables in the room.

Since all panels were expected to work on range finding during the first day, more in-depth training on the pinpointing procedure was reserved for the second day. Training on the second day also covered procedures for capturing information to be used for grade/content-specific PLDs.

After the second day facilitator and staff debrief, additional training was conducted on the third day to provide guidance on expected consistency of ratings for profiles with the same total linkage levels mastered and reminders about profiles as samples of performance. As procedures evolved during the week (see the Procedures section), third day training also included a short explanation about changes in procedures and set expectations for the timeline for panels to complete their remaining tasks.

Additional detail about on-site training is provided in the agenda and training slides in Appendix G.

Materials

Panelist Resources

Each panelist received a resource notebook with materials to use in training and during the rating process. The resource notebook contained:

- a standard setting flowchart,
- an annotated sample mastery profile,
- a performance level descriptor handout,
- hints for making ratings,
- instructions for completing rating forms,
- diagrams of the elements of the DLM system, and
- a glossary of DLM and standard setting terms.

When familiarizing themselves with each grade/subject's EEs and linkage levels, panelists also used the following resources:

- Node reference booklets that listed the nodes assessed at each linkage level for every Essential Element for the grade and content area. The booklets included node observation statements in addition to node names.
- A blank mastery profile for that grade/subject (i.e., one that contained Essential Elements and linkage level descriptions but no mastery shading).
- The blueprint for that grade/subject.

Panelists also had access to sample testlets for any Essential Element/linkage level assessed in a grade/subject. Upon request, facilitators displayed sample testlets in the online content management system. Testlets displayed as a preview of the way a student would see the testlet.

Training Materials

Training folders were prepared with exemplar profiles of student mastery for grade-, content-, and model-specific panels. The training folders included six exemplar profiles: two profiles with 10 levels mastered, two profiles with 20 levels mastered, and two profiles with 30 levels mastered. Two examples were included at each linkage level mastery amount to show how students with the same number of linkage levels might achieve that number by mastering different EEs or linkage levels. The training folders also contained sample rating sheets.

Range-Finding Materials

Range-finding folders were prepared with exemplar mastery profiles from across the range of student performances for the specific grade, content area, and model being reviewed. The number of profiles varied depending on the number of linkage levels on the blueprint. Exemplar profiles were provided in five-number increments. For example,

in a grade/subject with 10 Essential Elements and therefore 50 linkage levels, the range-finding folder included profiles for students who mastered 5, 10, 15, 20, 25, 30, 35, 40, and 45 linkage levels.

Profiles were ordered in the folder according to the total number of linkage levels the student mastered. Appendix D details the number of linkage levels per grade and content area. There were two exemplar profiles for each available level of mastery. In the previous example for a grade/content area with 50 possible linkage levels, a total of 18 profiles would be included in the folder spanning the 9 possible linkage level values included.

All exemplar profiles were numbered prior to being included in the folders to ease discussion.

Pinpointing Materials

The pinpointing folders contained profile exemplars for a reduced set of levels. For each cut point, exemplar profiles were included at seven levels, including the number closest to the suggested cut point determined in range finding and three above and below that number. For example, if range finding identified that the cut point should be somewhere around 30 linkage levels mastered, the folder would contain profiles with 27, 28, 29, 30, 31, 32, and 33 linkage levels mastered. A folder contained three profiles for each number of linkage levels mastered (i.e., multiple ways students have actually demonstrated the same number of linkage levels mastered), for a total of 21 profiles at the seven levels.

Any profiles that were used in range finding were re-used in pinpointing (e.g. the two profiles reviewed for 30 linkage levels mastered during range finding were also included in the pinpointing folder, with one additional profile).

Rating Forms

Rating forms for each of the range-finding and pinpointing processes were provided in the panelists' folders. One range-finding rating form and one pinpointing rating form were provided for each model, subject, and grade level set of cut points. Each form contained columns for round one (first) and round two (final) ratings. Example range-finding and pinpointing rating forms are provided in Appendix H.

Evaluation Forms

Two types of evaluation forms were provided to panelists: panelists' independent evaluations of group recommended cut points and panelists' evaluations of the overall standard setting training and meeting. The independent cut point evaluation forms were provided after groups completed each grade level set of cut points. The meeting evaluations were provided to panelists on the closing day of the standard setting meeting. Example independent and meeting evaluation forms are provided in Appendix I.

Procedures

Both the range-finding and pinpointing procedures consisted of two rounds of ratings. Panelists reviewed the exemplar profiles, independently rated each profile for round one ratings, discussed ratings as a group and then independently rated each profile again for round two ratings. Throughout both range finding and pinpointing, panelists were instructed to use their best professional judgment and consider all students with significant cognitive disabilities to determine which performance level best described each profile.

Each panel had at least two, and up to three, groups of grade-level cut points to set. The order in which they set the cuts for each grade level was based on the grades that they were assigned to. Panels in either the third through fifth or sixth through eighth grade levels began with the grade level that was adjacent to the next panel. That is, the panels assigned to grades 3-5 began with fifth grade while the panels assigned to grades 6-8 began with sixth grade. High school panels began with the lowest grade level. Table 10 displays the order in which each panel set their grade-level cut points.

Details of the final procedures used for determining cut points is provided in the subsequent sections. Further information regarding fidelity of the final procedures to the planned procedures can be found in Appendix J.

Table 10. Number of Panelists and Order in Which Cut Points Were Set by Panel

Panel	# of panelists	Order		
		First	Second	Third
ELA IM 3-5	8	ELA IM 5	ELA IM 4	ELA IM 3
ELA IM 6-8	7	ELA IM 6	ELA IM 7	ELA IM 8
ELA IM 9-12	8	ELA IM 9-10	ELA IM 11-12	--
Math IM 3-5	7	Math IM 5	Math IM 4	Math IM 3
Math IM 6-8	8	Math IM 6	Math IM 7	Math IM 8
Math IM 9-12	8	Math IM 10	Math IM 9	Math IM 11

Range Finding

During the range-finding process, panelists reviewed a limited set of profiles to assign general divisions between the performance levels. The goal of range finding was to locate ranges (in terms of number of linkage levels mastered) where panelists agreed that approximate cut points should exist.

These are the procedures the panelists followed for range-finding.

1. Panelists independently evaluated the profiles in the range-finding folder and identified the performance level that best described each profile. They recorded their decision for each exemplar profile on their rating sheet.
2. Once all panelists completed their ratings, the facilitator obtained the performance level recommendations for each profile by a raise of hands. The facilitator recorded the counts in the facilitator workbook, which was projected for the group to view. One panelist at each table was assigned to check that the values were entered correctly to ensure accurate data entry.
3. After table discussion of how panelists chose their ratings, the panelists were given the opportunity to adjust their independent ratings if they chose. A second round of ratings were recorded and shared with the group. Again, the facilitator entered values in the Excel sheet, and the designated panelist confirmed their accuracy.
4. Using the round two ratings, built-in logistic regression functions calculated the probability of a profile being categorized in each performance level conditioned on number of linkage levels mastered, and the most likely cut points for each performance level were identified. In instances where the logistic regression function could not identify a value (e.g. the group unanimously agreed on the categorization of profiles to performance levels), psychometricians evaluated the results to determine the approximate cut point based on the panelist recommendations.
5. Psychometricians reviewed every workbook prior to the group beginning the pinpointing process to ensure no errors were present and to check that the logistic regression had successfully determined a reasonably appropriate approximate cut point.

Pinpointing

During pinpointing, panelists reviewed additional profiles to refine the cut points. The goal of pinpointing was to pare down to specific cut points in terms of number of linkage levels mastered within the general ranges determined in range finding, not relying on conjunctive or compensatory judgments.

These are the procedures the panelists followed for pinpointing.

1. Folders were distributed to the panel containing the profiles for the seven levels, including and around the cut point value identified during range finding.
2. Panelists independently evaluated the profiles in each folder and assigned each a performance level – those in the higher level and those in the lower level. Panelists entered their recommendations on their pinpointing rating sheet.
3. Once all panelists completed their ratings, the facilitator obtained the recommendations for each profile by a raise of hands. These counts were entered into the projected facilitator Excel sheet. The identified panelist checker confirmed all values were entered correctly.
4. After discussion of the ratings, a second round of rating commenced. Panelists were given the opportunity to adjust their independent ratings if they chose.
5. The facilitator collected final ratings by show of hands. The checker confirmed values were entered correctly.
6. Using the second round's ratings, built-in logistic regression functions calculated the probability of a profile being categorized in each performance level conditioned on number of linkage levels mastered, and the most likely cut points for each performance level were identified. In instances where the logistic regression function could not identify a value (e.g. the group unanimously agreed on the categorization of profiles to performance levels), psychometricians evaluated the results to determine the final recommended cut point based on the panelist recommendations.
7. Psychometricians reviewed every workbook at the close of the pinpointing process to ensure values were obtained accurately.

Independent Evaluations of Panel-Recommended Cut Points

Once pinpointing was complete and the group had a final set of recommended cut points, panelists used cut point evaluation forms to indicate the degree to which they were comfortable with the cut points the group recommended. If panelists did not feel comfortable with the group cut point, they were also asked to indicate what cut point they would be comfortable with. Panelists completed their cut point evaluation forms independently and turned them in to the table facilitators without discussing as a group.

DLM staff entered the data from the evaluation forms into a spreadsheet after the standard setting meeting adjourned.

Chapter 5: Results

This chapter summarizes the final panel-recommended cut points, evaluation evidence regarding the panel process, impact data, and the final results.

Panel-Recommended Cut Points

Table 11 includes a summary of the cut point recommendations reached by the panelists following the range-finding and pinpointing process. Note that the linkage levels in the last column are not to be interpreted as a percent correct or as an interval level scale.

Table 11. Final ELA and Math IM Panel Cut Point Recommendations

Grade	Emerging/ Approaching	Approaching/ Target	Target/ Advanced	Minimum Required Linkage Levels
ELA				
3	12	18	29	40
4	8	16	38	45
5	9	16	35	40
6	9	21	36	45
7	14	27	41	55
8	18	29	43	55
9-10	18	32	42	50
11-12	21	30	40	50
Math				
3	7	17	24	30
4	9	23	33	40
5	8	19	29	35
6	10	18	25	30
7	12	20	28	35
8	14	24	30	35
9	7	13	23	30
10	6	15	24	30
11	8	14	22	30

Convergence

The purpose of range-finding and pinpointing was to identify the specific number of linkage levels mastered that would differentiate placement of students between each of the four performance levels. Through each round of discussion and ratings, panelists narrowed in on the range in which the cut point could be identified. Due to the nature of the statistical analysis method used, inter-panelist consistency was not the desired outcome for a single round; however, there was an expectation that panelists would converge toward an increasingly narrow range of profiles to identify the cut point. To illustrate the degree to which panelists converged upon an agreed upon cut point, box and whisker plots are displayed in Appendix K. These plots convey the median, first and third quartiles, and range of the frequencies with which each number of linkage levels mastered was classified into each of the four performance levels.

Overall, the plots support the claim that the panel process worked as intended. In general, the ranges of profiles categorized into each performance level narrowed from round one to round two during both range finding and pinpointing.

Evaluations

Standard Errors of Pinpointing Ratings

Following the standard setting event, standard errors were computed to evaluate the results. This method was based on the frequency distributions of panelists' final pinpointing ratings and was accomplished by dividing the standard deviation of the frequencies of panelists' final pinpointing ratings by the square root of the number of total ratings. Table 12 and Table 13 display the standard errors for the distribution of final pinpointing ratings.

Table 12. Standard Errors for ELA Final Pinpointing Ratings

	G3	G4	G5	G6	G7	G8	G9-10	G11-12
Emerging	0.1105	0.1288	0.1685	0.1737	0.1387	0.1514	0.1414	0.0847
Approaching	0.1700	0.3020	0.4771	1.2480	0.3459	0.3927	0.3393	0.2023
Target	0.2958	0.5842	0.3743	0.3027	0.9788	0.5319	0.4497	0.3205
Advanced	0.1074	0.1585	0.2483	0.2065	0.1542	0.1588	0.1406	0.1619

Table 13. Standard Errors for Math Final Pinpointing Ratings

	G3	G4	G5	G6	G7	G8	G9	G10	G11
Emerging	0.1049	0.1266	0.1227	0.1455	0.1022	0.1292	0.1429	0.1532	0.1260
Approaching	0.2437	0.4419	0.4148	0.1983	0.1811	0.2399	0.2107	0.2240	0.1713
Target	0.1655	0.2827	0.2188	0.2020	0.3022	0.1668	0.2332	0.3115	0.1884
Advanced	0.0781	0.1053	0.4122	0.1594	0.1369	0.1393	0.1456	0.2329	0.1243

Independent Panelist Evaluations of Panel-Recommended Cut Points

A short questionnaire was created to evaluate panelists' comfort with the final panel-recommended cut points. Once pinpointing was complete for the grade level, questionnaires were distributed to each panelist. The panelists each rated their comfort level with the three cut points their group set for each grade or course. Panelists responded with whether or not they agreed with the group-determined cut point and indicated what they would independently suggest be applied for each cut point. Table 14 summarizes panelist responses to this questionnaire. Note that the percent included in the table is based on all three cut points. Panelists rated their comfort three times: once for the Emerging/Approaching cut, once for the Approaching/Target cut, and once for the Target/Advanced cut.

Table 14. Panelist Comfort with Group Recommended Cut Points

Content Area and Grade	N Panelists	N Ratings (n Panelists x n Cut Points Evaluated)		% Agreement
		n "Yes" Ratings		
ELA				
3	8	24	24	100.0
4	8	24	24	100.0
5	8	24	19	79.2
6	7	21	20	95.2
7	7	21	21	100.0
8	7	21	21	100.0
9-10	7	21	20	95.2
11-12	7	21	21	100.0

Math

3	7	21	21	100.0
4	7	21	21	100.0
5	7	21	21	100.0
6	8	24	24	100.0
7	8	24	24	100.0
8	8	24	24	100.0
9	8	24	24	100.0
10	8	24	24	100.0
11	8	24	22	91.7

Across all panelists, panels, grades/courses, and cut points ($N=861$), 95.9% of panelists ($n = 826$) indicated that they were comfortable with the group-recommended cut point. Only 4.1% of responses ($n = 35$) indicated a discomfort with a group-recommended cut. Complete panelist agreement with the recommended cut point was found in 99 out of 120 cuts (82.5%) across all models and subjects, grades, and courses. Independent panelist comfort with all three recommended cut points was found for 25 out of 40 cut point panels (62.5%). Most recommendations for a change to the cut point were for only one of the three cut points for a given panel, and most often, the recommended changes differed from the initial recommendation by only a single point. Unfortunately, data was not collected on the reasons for disagreement; however, future standard setting methods will incorporate this valuable information.

To further evaluate panelist comfort with the recommended cut points, the median panelist-recommended cut point was compared to the group-recommended cut point. In all instances but two, the median- and group-recommended cut points were the same. Table 15 summarizes the two instances where the median differed from the group-recommended value. In both instances, the median panelist recommendation was lower than the group-recommended cut point for the cut between the **approaching** and **target** performance levels.

Table 15. Panels with Median Independent Cut Point Different from Group-Recommended Cut Point

Content Area/Model	Grade	Performance Cut Point	Group Recommended Cut Point	Median Panelist Independent Cut Point
ELA IM	5	AP/T	26	23

Standard Errors of Independent Recommended Cut Points

Kingston and Tiemann (2012) outline a method for computing standard errors when using Body of Work methods (or close variants), which is based on variability in the individual panelist recommended cut points. The standard deviation of independent cut points were divided by the square root of the number of panelists. The standard errors of the individual panelist cut points and the panels' final pinpointing ratings by grade level are shown in Table 16 and Table 17.

Please note, there are a couple of instances where a panelist agreed with the panel-recommended cut point as noted in the tables above but suggested a different cut point than what the panel recommended as noted in the tables that follow. For example, there was one panelist on the IM ELA Grade 7 panel who agreed with all three of the panel-

recommended cut points but also recommended a cut point of that was one point higher or lower for each of the three levels.

Table 16. Standard Errors for IM ELA Independent Panelist-Recommended Cut Points

Grade	App+	Target+	Adv+
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.00	1.04	0.00
6	0.29	0.71	0.00
7	0.14	0.14	0.14
8	0.00	0.00	0.00
9-10	0.14	0.00	0.00
11-12	0.00	0.00	0.00

Table 17. Standard Errors for IM Math Independent Panelist-Recommended Cut Points

Grade	App+	Target+	Adv+
3	0.00	0.00	0.00
4	0.00	0.00	0.00
5	0.00	0.00	0.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	0.00	0.00	0.00
10	0.00	0.00	0.00
11	0.42	0.00	0.00

Panelist Evaluations of the Meeting

Panelists completed evaluations at the conclusion of the standard setting meeting. The evaluation included self-evaluation of readiness to rate and understanding of the tasks plus evaluation of outcomes. Panelists rated their responses to the 27 questions on a Likert scale, choosing either “Strongly Disagree” (SD), “Disagree” (D), “Agree” (A), or “Strongly Agree” (SA). For the last three questions, “Not applicable” was an additional option.

Table 18 shows that, overall, the majority of panelists for both models agreed or strongly agreed that the meeting was well organized; they understood their tasks and felt

confident to complete their tasks, and they thought the cut points they made were defensible and valid. Furthermore, panelists believed that the meeting was a good experience in terms of professional development and for planning instruction with students with significant cognitive disabilities. The questions that evaluate panelists' perceptions of the results of the standard setting process are highlighted in the Table 5.6-IM.

Table 18. Percentages of IM Panelist Responses to Evaluation Items

Question	SD	D	A	SA
1. The overall goals of the standard setting panel meeting were clear.			27	73
2. The panel meeting was well organized.			24	76
3. The training and practice exercises provided the information I needed to complete my tasks.		7	40	53
4. It was clear what knowledge, skill, or ability a student would need to demonstrate to achieve a certain profile.		2	57	41
5. I considered the performance level descriptors when I rated each profile.			32	68
6. I considered the assessment items when I rated each profile.			44	56
7. I considered the other panelists' opinions when I rated each profile.		2	36	61
8. I considered my experience in the field when I rated each profile.			27	73
9. I understood how to rate each profile.			47	53
10. I had enough time to complete the tasks.			24	76
11. I felt confident when rating the profiles.		2	47	51
12. The procedure for recommending cut points was free from bias.			40	60
13. Overall, I was satisfied with the ratings made by panelists in my group.			20	80
14. I would defend the group's At Target decisions against criticism that they are too high.			49	51
15. I would defend the group's At Target decisions against criticism that they are too low.			42	58
16. I would defend the group's Advanced decisions against criticism that they are too high.			47	53
17. I would defend the group's Advanced decisions against criticism that they are too low.			42	58
18. I would defend the group's Approaching Target decisions against criticism that they are too high.			50	50
19. I would defend the group's Approaching Target decisions against criticism that they are too low.			42	58
20. I am confident that the meeting produced valid cut-point recommendations.			36	64
21. Overall, I believe my opinions were considered and valued by the group.			20	80
22. Overall, my group's discussions were open and honest.			13	87
23. Participating in the process increased my understanding of DLM.			4	96
24. Overall, I valued the panel meeting as a professional development experience.			2	98
25. This experience will help me plan and provide instruction for my students with significant cognitive disabilities.			17	83
26. This experience will help me use DLM more effectively.			16	84

Technical Advisory Panel (TAC) Evaluation of Panel Process

The DLM TAC chair was on-site for the duration of the standard setting event. The goal was to observe the process and provide feedback to the TAC and consortium states regarding any relevant observations of the event. Overall, the DLM TAC chair believed that the standard setting meeting was well planned and implemented, the staff were helpful to the panelists, and the panelists worked hard to set standards. The full TAC evaluated the evidence about the standard setting process, including the TAC chair's observations, panelist evaluations, and the relationship between panel and independent cut points. The TAC accepted the resolution about the adequacy, quality of judgments, and extent to which the process met professional standards. A copy of the memorandum and resolution is provided in Appendix L.

Impact Data Based on Panel-Recommended Cut Points

Using the panel-recommended cut points, impact data was calculated to present to the TAC as well as to the states to use in their decision-making process for adoption of the cut points. The impact data was calculated using the linkage level mastery status and total number of linkage levels mastered on each tested EE for all students. Duplicate student records were then removed using the following rule:

Remove duplicates when the following fields were all identical across rows: student ID, state, grade level, and number of linkage levels mastered by subject area.

This step prevented the same student's linkage level mastery status from being used multiple times in the calculation of the impact data. This means that if a student was rostered to multiple teachers, the data were only included once. Students who were rostered in the system but did not test on any EEs were not excluded from the data file. However, because these students had no scores, their inclusion did not influence the frequency distributions of the impact data. Once duplicate records were removed, the frequency distributions of students at each performance level were calculated by grade level, content area, and model.

Table 19 shows the frequency distributions for the panel-recommended cut points.

Table 19. Percent of IM Students at Each Performance Level

Content Area and Grade	Emerging (%)	Approaching (%)	Target (%)	Advanced (%)	Target/Adv (%)
-------------------------------	---------------------	------------------------	-------------------	---------------------	-----------------------

ELA						
3	32.5	17.3	33.6	16.6	50.2	
4	21.3	13.7	58.0	7.0	65.0	
5	27.0	16.2	52.6	4.3	56.9	
6	27.1	22.3	40.2	10.3	50.6	
7	32.4	17.6	34.1	15.9	50.0	
8	39.2	20.9	30.9	9.0	39.9	
9-10	41.2	17.7	27.4	13.7	41.1	
11-12	45.2	10.4	26.2	18.2	44.4	
Math						
3	41.7	35.1	17.8	5.5	23.3	
4	44.7	40.2	13.0	2.1	15.2	
5	42.8	35.5	18.0	3.7	21.8	
6	60.7	23.0	12.9	3.4	16.3	
7	54.9	28.4	12.9	3.8	16.7	
8	69.9	23.8	4.6	1.8	6.4	
9	49.4	29.2	19.8	1.6	21.4	
10	49.4	39.5	10.2	0.8	11.1	
11	55.2	23.9	18.2	2.7	20.8	

Statistical Adjustment

Procedure

Despite evaluative evidence that was generally supportive of the panel-recommended cut points, these recommendations are based on the work of single panels. Each panel is a sample of possible experts. In theory, some variability in recommended cut points would be expected with a different sample, and each sample's recommendation would be an estimate of the true cut point.

To mitigate the effect of sampling error and issues related to a system of cut points across a series of grade levels, many testing programs have borrowed strength by considering impact data in the grade at question **and** contiguous grades. The logic is that under most circumstances (especially when there is no significant shift in demographics), students in bordering grades should have similar distributions within performance levels. Dramatically different distributions are likely due to sampling error and not differences in true cut points.

Statistical adjustments were made to the panel-recommended cut points in an effort to systematically smooth distributions within the system of cut points being considered. Adjustments were applied to the panel-recommended cut points, separately for the year-end and integrated models. No adjustments were made for EOI because the standards

assessed and students taking these assessments were assumed to be very different from one course to another.

The following steps were applied to each subject within each grade level.

1. Create a frequency distribution of the number of linkage levels mastered (from low to high). This step was done separately for each grade/subject. The number of possible linkage levels varies considerable from one grade to another.
2. Calculate cumulative proportions from low to high.
3. Perform a probit transformation (z-score associated with the cumulative proportion of students) for each number of linkage levels mastered. Because at the top of the distribution (proportion equal to 1) a finite z-score cannot be calculated. To perform subsequent calculations, z-scores were defaulted to 3.5.
4. Find the z-score associated with the raw cut point of interest (for example, Approaching/Target) for each grade level.
5. Create a weighted rolling average of z-scores for the cut-point of interest, using a weight of 0.4 for the grade of interest, 0.2 for contiguous grades, and 0.1 for the next grades.

$$\frac{\sum w_i Z_i}{\sum w_i}$$

At the ends (grades 3, 4, 11, 12) there cannot be a symmetric set of five grade levels involved in the rolling average.

- a. For ELA in the integrated model, the end of the grade continuum was 9-10 and 11-12.
6. Using the table of probit-transformed cumulative proportions, look up the raw number of linkage levels mastered for which the z-score is closest to the weighted rolling average of z-scores. The closest z-score was selected instead of the lowest z-score to prevent systematically decreasing the proportion of students in the higher category over the system of cut-points.

Adjusted Cut Points

Table 30 summarizes the adjusted cut points that used the methods described above and the impact data for those adjusted cut points. Frequency distributions for the impact data of the adjusted cut points were calculated using the same process as described for the panel-recommended cut points.

The approach used did decrease the between-grade variability without systematically raising or lowering the cut points over the entire system of grade-level tests. While the EOI assessments may have benefitted from smoothing, no reasonable approach for doing so was evident, since there was no reason to expect that the students in one course were in any way representative of the students in the other courses. Also, sample sizes for the EOI courses were relatively small—as low as 97 for Algebra 2.

Table 20. IM Adjusted Cut-Point Recommendations

Content Area and Grade	Emerging/ Approaching	Approaching/ Target	Target/ Advanced	Minimum Required Linkage Levels
ELA				
3	9	16	30	40
4	11	18	37	45
5	8	16	33	40
6	9	20	36	45
7	13	27	42	55
8	17	27	41	55
9-10	18	31	41	50
11-12	15	30	40	50
Math				
3	7	17	25	30
4	9	21	31	40
5	9	20	29	35
6	8	17	24	30
7	12	21	29	35
8	11	20	30	35
9	8	15	22	30
10	5	13	22	30
11	7	15	23	30

Impact Data Based On Adjusted Cut Points

Using the adjusted cut points, impact data was calculated to present to the TAC as well as to the consortium states to use in their decision making process for adoption of the cut points. The impact data was calculated using the same process used with the panel-recommended cut points.

Table 21 shows the frequency distributions for the panel-recommended cut points.

Table 21. Percent of Students at Each Performance Level Based on IM Adjusted Cut Points

Content Area and Grade	Emerging (%)	Approaching (%)	Target (%)	Advanced (%)	Target/Adv (%)
ELA					
3	28.6	15.6	42.6	13.3	55.9
4	26.5	13.2	51.1	9.2	60.3
5	24.8	18.5	48.2	8.5	56.7
6	28.5	20.2	41.9	9.4	51.3
7	32.3	20.2	35.1	12.4	47.5
8	38.6	16.8	32.2	12.4	44.6
9-10	43.0	15.3	26.4	15.4	41.7
11-12	46.7	14.3	22.3	16.6	38.9
Math					
3	42.6	35.0	18.9	3.6	22.5
4	45.2	35.1	16.8	3.0	19.8
5	47.7	33.0	15.8	3.5	19.3
6	53.5	28.7	14.1	3.7	17.8
7	57.4	28.1	12.0	2.5	14.5
8	60.5	25.8	12.2	1.6	13.8
9	54.8	28.9	14.0	2.3	16.3
10	44.5	41.1	12.4	2.0	14.4
11	49.4	32.9	15.7	2.0	17.7

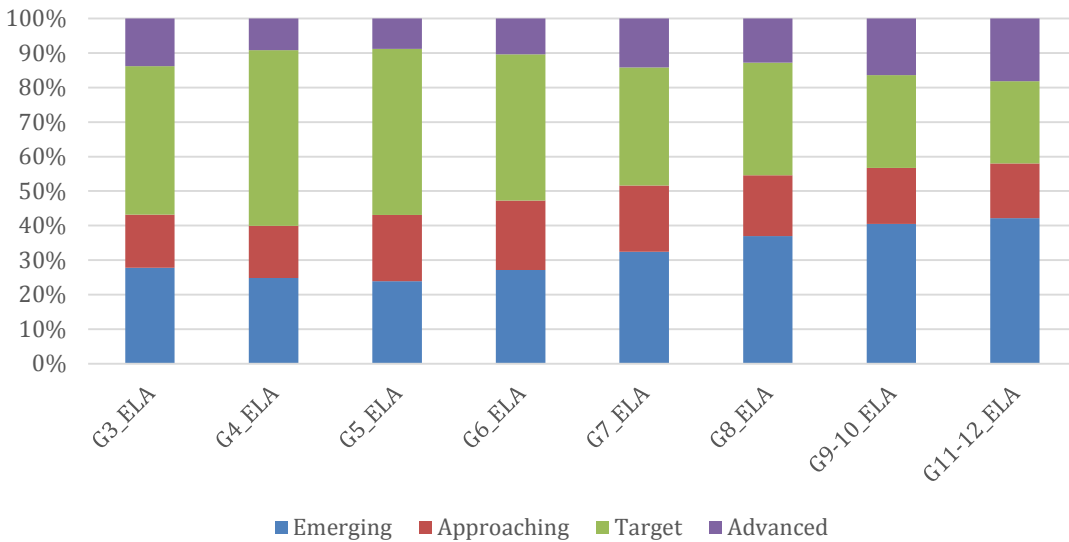


Figure 6. IM English Language Arts Impact Data Using Adjusted Cut Points

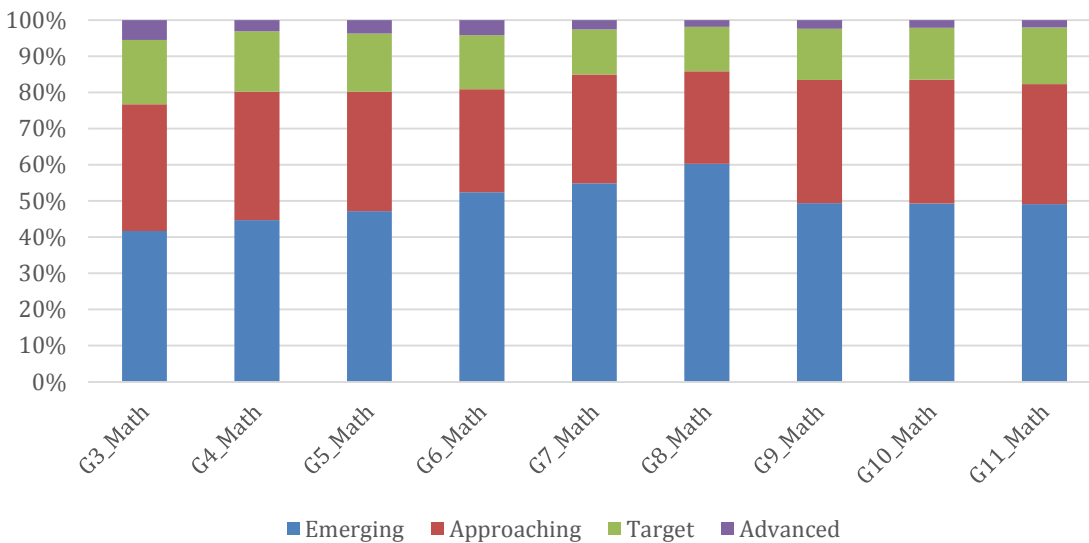


Figure 7. IM Mathematics Impact Data Using Adjusted Cut Points

Final Results

The panel-recommended cut points, adjusted cut points, and associated impact data for both sets of cut points were presented to the TAC and partner states for review. The TAC accepted the DLM adjustment method and resulting adjusted cut points. Following the states' review process and discussion with the DLM team, the states using each model voted to accept the DLM-recommended adjusted cut points as the final consortium cut points with no further adjustment.

Chapter 6: Future Steps

This technical report describes the steps in standard setting from developing policy-level PLDs through consortium adoption of cut points. Since the chosen standard setting approach was student-based rather than item-based, grade and content-specific PLDs were not developed for use during the panel process. Instead, grade and content-specific PLDs will be developed from the work done by panelists as they evaluated profiles. Starting with raw notes about critical skills and understandings for each performance level and the associated rationales, DLM test development content teams will draft PLDs for each grade and content area. These drafts will go through rounds of review and input from the partner states before they are finalized.

References

Bridgeman, B. (2013). *Human Ratings and Automated Essay Evaluation*. In M. D. Shermis & J. Burstein (Eds.), *Handbook of Automated Essay Evaluation* (221-232). New York, NY: Routledge.

Cizek, G. J. (1996). Standard-setting guidelines. *Educational Measurement*, 15, 13-21.

Cizek, G. J., & Bunch, M. B. (2006). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 433-470). New York, NY: American Council on Education/Praeger.

Hambleton, R. K., Pitoniak, M. J., & Copella, J. M. (2012). *Essential steps in setting performance standards on educational tests and strategies for assessing the reliability of results*. *Setting performance standards: Foundations, methods, and innovations*, 47-76.

Karvonen, M., Clark, A., & Nash, B. (2015). *2015 year-end model standard setting: English language arts and mathematics* (Technical Report No. 15-03). Lawrence, KS: University of Kansas, Center for Educational Testing and Evaluation.

Kingston, N. M. & Tiemann, G. C. (2012). *Setting performance standards on complex assessments: The body of work method*. In G. J. Cizek (Ed.). *Setting performance standards: Concepts, methods, and perspectives* (2nd ed.). New York, NY: Routledge.

Appendix A: DLM 2014-2015 Scoring Model Description

Essential Elements (EEs) are academic content standards for students with significant cognitive disabilities. For each tested EE, assessments are available at one of five linkage levels representing the relative progression toward the academic standard. For each part of the assessment, the student receives a testlet at a linkage level. Each level contains items aligned to one or more nodes in the learning map.

The original intent was for the student's responses to allow for modeling of mastery probabilities of all nodes, including those at the non-tested linkage levels. However, due to testlet design and delivery, many EEs had test items that were administered to non-overlapping sets of students, creating statistical issues with concurrent calibration of all items across all nodes for an EE. Therefore, the DLM scoring model used for operational purposes in 2014-15 was constructed based on information obtained from students at each linkage level separately and then aggregated to produce student linkage level mastery estimates.

Students taking testlets at a linkage level within an EE were considered masters of that linkage level if one of two conditions were met:

1. The posterior probability of mastery determined from the diagnostic classification model estimated for the linkage level was greater than or equal to 0.8.
2. The proportion of items answered correctly within the linkage level was greater than or equal to 0.8.

Students were considered masters by meeting either condition in order to prevent consequences associated with false negatives. Linkage levels were treated hierarchically in that masters of higher linkage levels (based on the two criteria above) were automatically assumed to be masters of lower linkage levels. Students who did not demonstrate mastery at any linkage level were assumed to be masters of linkage levels at least two categories below the highest linkage level where they tested. Students who did not meet mastery criteria and whose highest level tested was either the Initial Precursor or Distal Precursor levels were considered non-masters of all linkage levels.

The diagnostic classification model used to classify students within each linkage level was the "Noisy Inputs, Deterministic Or gate" (NIDO) model (e.g., Rupp, Templin, & Henson, 2010; Templin, 2006). In this model, all items from each linkage level within each EE are treated as measuring one binary latent variable that represents mastery status for a student. All items within a linkage level are treated as exchangeable or fungible, a condition made necessary due to many items not being administered to large numbers of examinees. Fungibility (from the NIDO model) means that within a linkage level, all item parameters are constrained to be equal, providing the same item intercept and main effect parameters.

References:

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic Measurement: Theory Methods and Applications*. New York: Guilford (Chapter 6, p. 135).

Templin, (2006). [Computer Program Manual]. CDM User's Guide. Lawrence, Kansas.

Appendix B: Rationale for Determining Mastery and Developing Profiles

Step One: Mastery at the Node Level

During the March 2015 TAC meeting, DLM staff proposed a node mastery threshold of 0.8 be applied to all nodes assessed in the learning map. This value was proposed because a probability further from 0.5 and closer to 1.0 allows for greater certainty in making claims about node mastery status for the student, and the negative educational consequences of ending instruction on a topic before mastery are significant. Furthermore, it is often used by teachers in writing IEP goals (e.g. 4/5 trials or 80%).

To investigate the impact of using a threshold of 0.8, DLM field test data from January 2014 through February 2015 was evaluated using items with at least 200 student responses to ensure stability. Table 22 shows the percent of students who would be classified as node masters and non-masters based on varied probabilities of node mastery. Using a value of 0.8 as the threshold indicates roughly 38% of students would be considered masters of the nodes they were assessed on.

Table 22. Percent of Students Estimated as Node Masters Using Administered Testlet Data Based on the Probability that Indicates Mastery

	Probability of Node Mastery				
	0.5	0.6	0.7	0.8	0.9
Masters	0.55	0.51	0.45	0.38	0.30
Non-masters	0.45	0.49	0.55	0.62	0.70

TAC members discussed the optimal threshold for node mastery as part of the March 2015 TAC meeting. Through discussion, the TAC members reached consensus that a probability of 0.8 was the preferred threshold for node mastery. This value was preferred due to being further from 0.5 (the point of maximum uncertainty) and therefore allows greater certainty in the mastery determination. The TAC members noted that 0.8 is a little over one standard deviation above 0.5, which added additional support for their recommendation.

This recommendation was then taken to the DLM states for approval. All 12 states who responded to the poll indicated that a threshold of 0.8 was acceptable for determining node mastery.

Based on feedback from states and the TAC, a threshold value of 0.8 was applied to the non-bifactor, node-based model for standard setting.

Step Two: Linkage Level Mastery

To further evaluate the impact of selecting 0.75 or 0.8 as the threshold for linkage level mastery, the number of nodes assessed per linkage level was considered. Table 23 lists the number of nodes per linkage level for mathematics and ELA, respectively. Most ELA linkage levels (94%) assess a single node in the learning map. As such, students would be required to have mastered the one assessed node to be considered a linkage level master, regardless of whether the linkage level threshold was set at 0.75 or 0.8.

However, for mathematics, the spread of nodes assessed per linkage level has greater variation. Most linkage levels assess between one and three nodes (89%). For those linkage levels assessing more than three nodes, the difference between a threshold of 0.75 and 0.8 has more impact. If 0.75 is selected as the threshold, students would need to have mastered only three out of four nodes at a linkage level, including four nodes to be considered a master of the linkage level. By contrast, a threshold of 0.8 would require students to have mastered all four nodes. By selecting 0.75 as the threshold, more students have potential to be considered masters of the linkage level. By selecting 0.8 as the threshold, greater content knowledge (as evidenced by node mastery) would be required. Since this was the first operational year of the DLM system, it was decided that 0.75 would be used as the criterion in an effort to consider more students as masters of the linkage level.

Table 23. Number of Nodes per Linkage Level

Number of Nodes	Number of Linkage Levels	
	Math	ELA
1	188	697
2	207	47
3	81	1
4	30	N/A
5	14	N/A
6	5	N/A
7	1	N/A
8	5	N/A
9	1	N/A
10	2	N/A
11	1	N/A

Appendix C: Sample Profile Based on Judgments about Linkage Levels: Grade 7 ELA

Shaded cells indicate linkage levels where a student's node mastery translates into a judgment of mastery. This example student mastered 34 of 90 linkage levels.

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C1.1	RL.7.5	Understands that objects have a function	Identifies a concrete detail in early informational texts	Determines how the title fits the structure of the text	Determines how a fact, step, or event fits into the overall structure of the text	Determines how a word, phrase, sentence, or paragraph fits into the overall structure of an informational text
C1.2	RL.7.1	Differentiates between text and pictures	Identifies the characters, setting, and major events of a story	Identifies words in a narrative to answer a question about explicit information	Analyzes a narrative to identify where information is explicitly stated and where inferences should be drawn	Determines what a narrative states explicitly and implicitly
C1.2	RL.7.2	Matches a picture with the real object	Identifies concrete details in a familiar story (characters, objects)	Identifies the overall goal or main idea of a single episode	Identifies events that are related to the theme of a narrative	Identifies the relevant events contributing to the theme or central idea of a narrative

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C1.2	RL.7.4	Demonstrates understanding of words for absent objects and persons	Relies on syntactic clues to determine the meaning of an unknown word when the definition is directly stated in the sentence	Determines the meaning of multiple meaning words in text (support of context)	Determines the meaning of simple idioms and figures of speech	Determines the connotative meaning of words and phrases in a text
C1.2	RI.7.1	Differentiates between text and pictures	Identifies illustrations that go with a familiar text	Identifies words in the text to answer a question about explicit information	Analyzes text to identify where information is explicitly stated and where inferences must be drawn	Determines the difference between explicit and implicit information in an informational text
C1.2	RI.7.2	Matches a picture with the real object	Identifies a concrete detail in early informational texts	Identifies the implicit main idea of a paragraph in an informational text	Identifies multiple main ideas in an information text	Creates a summary for a familiar informative text
C1.2	RI.7.4	Demonstrates understanding of property words	Recognizes the literal meaning of a word or phrase in a sentence	Determines the meaning of phrases in a text using the context	Determines how word choice in an informational text is used to persuade or inform	Determines the connotative meaning of words and phrases in a text

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C1.2	RI.7.6	Identifies people associated with familiar routines	Identifies a concrete detail in early informational texts	Identifies words or phrases in an informational text that describes or demonstrates the author's point of view	Determines the author's point of view or purpose for writing in an informational text	Identifies examples demonstrating an author's or presenter's point of view in an informational text or presentation on a topic
C1.2	RI.7.8	Anticipates consequences of actions	Identifies the main idea in a single paragraph when it appears explicitly in the first sentence	Determines how a word, phrase, sentence, or paragraph fits into the overall structure of an informational text	Determines the place each claim and reason takes in the overall structure of an informational text	Describes the overall structure of informational texts
C1.3	RI.7.3	Demonstrates understanding of property words	Relates character's feelings or desires to action in a story	Identifies how a character responds to a challenge in story	Determines how two story elements (characters, settings, or major events) relate to each other	Identifies the causes of character's actions in a story
C1.3	RI.7.5	Differentiates between text and pictures	Identifies the beginning and end of a familiar story	Determines the structure of a text (e.g., story, poem, or drama)	Compares the structure of two or more texts (e.g., stories, poems, or dramas)	Compares and contrasts the structure of two or more texts (e.g., stories, poems, or dramas)

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C1.3	RI.7.3	Draws conclusions based on category knowledge (early inductive reasoning)	Identifies the key points the author makes in an informational text	Identifies two related points made in an informational text.	Identifies how two or more individuals, events, or ideas in an informational text are related	Describes how two or more details in an informational text interact or relate
C1.3	RI.7.9	Draws conclusions based on category knowledge (early inductive reasoning)	Identifies a common element in two different informational texts on the same topic	Compares and contrasts informational texts on the same topic	Compares and contrasts how two different informational texts on the same topic present details	Compares the arguments made by two different informational texts on the same topic
C2.1	L.7.2.a	Understands that objects have a function	Recognizes the first word to read on a page	Recognizes end punctuation	Uses end punctuation in written communication	Uses commas in written communication
C2.1	L.7.2.b	Identifies the first letter in their own name	Uses letters to create string (words)	Uses spelling patterns in familiar words to spell new words	Spells words phonetically using letter-sound knowledge and common spelling patterns	Spells words with inflectional endings

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C2.1	W.7.2.a	Makes a choice between two objects	Demonstrates understanding of wh- questions	Writes about a topic using pertinent tactile, visual, or multimedia information	Introduces an informational topic and write to convey both ideas and information	Introduces a topic clearly and use a clear organization to write about it including visual, factual, or multimedia information as appropriate
C2.1	W.7.2.b	Identifies words that describe familiar persons, places, things, or events	Identifies functional words to describe common persons, places, objects, or events	Strengthens the message of written work (drawing, dictation, writing) by adding more information	Provides facts, details, or other information related to the topic	Writes one or more facts or details related to the topic
C2.1	W.7.2.d	Understands that specific members comprise categories	Draws conclusions based on category knowledge	Recognizes domain-specific words in speech or text	Selects domain-specific vocabulary to use in writing about a topic	Uses domain-specific vocabulary in informative writing

Appendix D: Number of Essential Elements and Number of Linkage Levels Available per Grade and Subject, Integrated Model

Integrated Model ²				
	EEs Available	Linkage Levels Available	Minimum Required Blueprint EEs	Minimum Required Blueprint Linkage Levels
ELA				
Grade 3	17	85	8	40
Grade 4	17	85	9	45
Grade 5	19	95	8	40
Grade 6	19	95	9	45
Grade 7	18	90	11	55
Grade 8	20	100	11	55
Grade 9	n/a	n/a	n/a	n/a
Grade 10	n/a	n/a	n/a	n/a
Grade 11	n/a	n/a	n/a	n/a
Grades 9-10	19	95	10	50
Grades 11-12	19	95	10	50
Math				
Grade 3	11	55	6	30
Grade 4	16	80	8	40
Grade 5	15	75	7	35
Grade 6	11	55	6	30
Grade 7	14	70	7	35
Grade 8	14	70	7	35
Grade 9	8	40	6	30
Grade 10	9	45	6	30
Grade 11	9	45	6	30

² Note for integrated model states, not all EEs on the blueprint are required. The first two columns of this table list the maximum number of possible EEs on the integrated model blueprint. The last two columns of this table list the required number of EEs based on the integrated model blueprint.

Appendix E: Standard Setting Panelist Recruitment Letter and Survey

Dear Colleagues,

[state] is a state partner in the Dynamic Learning Maps (DLM) Alternate Assessment Consortium. DLM assessments are designed for students with significant cognitive disabilities and measure student mastery of content in English language arts and mathematics based on progressions in a learning map. The 2014-2015 academic year is the first year the DLM assessment is operational. Student responses obtained during this first operational testing window will be used to determine what level of mastery is associated with certain performance levels. This process is referred to as standard setting.

As a partner state, we have the opportunity to recruit educators to serve on one of three panels that will help set standards:

- Elementary (grades 3-5)
- Middle (grades 6-8)
- High school (grades 9-12)

We are writing to invite volunteers from [state or district] to serve on these three DLM standard setting panels. We seek educators with a broad array of perspectives and backgrounds, although we especially seek individuals with content expertise in English language arts or mathematics and in education and assessment for students with significant cognitive disabilities. Other subject matter experts and individuals who work at establishments that employ individuals with significant cognitive disabilities are also encouraged to volunteer to serve on high school panels.

We ask that volunteers commit to up to four hours of training in advance of the meeting and to attend an on-site standard setting meeting in Kansas City, MO June 15-18, 2015. Panelists must be present for the entire on-site meeting. Panelists who participate outside the scope of their usual job requirements will be paid a stipend of \$___ to complete advance training and participate in the entire on-site meeting.

Volunteers are invited to complete a background survey online following the link provided (). The deadline to volunteer to participate in a standard setting panel is Friday April 10, 2015. DLM will notify volunteers who are selected to serve on panels.

We would appreciate your assistance with recruiting volunteers to serve as standard setting panelists.

Questions about the external review process should be directed to d1m@ku.edu.

Thank you for your assistance with the recruitment process!

Sincerely,

Intro DLM Standard Setting Panel Survey

Please tell us about yourself and your interest in participating as a standard setting panel member. Thank you!

Q1 First name

Q2 Last Name

Q3 E-mail Address

Q4 Preferred Phone Number

Q5 Full Mailing Address

Street Address 1

Street Address 2

City

State

Zip

Q6 What is your current role?

- Classroom Teacher
- Building Administrator
- District Staff
- State Education Agency Staff
- University Faculty/Staff
- Community Member
- Other _____

Q7 Please adjust the bars to indicate your years of p-12 educational experience in each of the following areas.

_____ ELA

_____ Math

_____ Students with Significant Cognitive Disabilities

_____ p-12 Education Overall

Q8 Which of the following types of students with disabilities have you taught/worked with in the past ten years? (Mark all that apply)

- Blind/Low Vision
- Deaf/Hard of Hearing
- Emotional Disability
- Mild Cognitive Disability
- Multiple Disabilities

- Orthopedic Impairment
- Other Health Impairment
- Severe Cognitive Disability
- Specific Learning Disability
- Speech Impairment
- Traumatic Brain Injury
- None of the Above

Q9 Which grade(s) did you teach in 2014-15?

- Grade 3
- Grade 4
- Grade 5
- Grade 6
- Grade 7
- Grade 8
- Grade 9
- Grade 10
- Grade 11
- Grade 12
- I did not teach in 2014-15

Answer If *Which grade(s) did you teach in 2014-15?* None Is Selected

Q9b Please indicate the grade band(s) at which you believe you have expertise to participate in standard setting.

- Grades 3-5
- Grades 6-8
- Grades 9-12

Q11 How many years of experience do you have teaching at these grade levels?

_____ Years of Experience

Q12 Do you have previous experience with a standard setting process for another large-scale assessment besides DLM?

- Yes
- No

Q13 How many years of experience do you have with Alternate Assessments based on Alternate Achievement Standards (AA-AAS)?

- None
- less than 1 year
- 1-5 years
- 6-10 years
- 11+ years

Q14 Have you written items for DLM?

- Yes
- No

Q15 Have you previously served as an external reviewer for DLM?

- Yes
- No

Q16 Please list all licensures/certifications you hold.

Q17 Please check all of the following statements that apply to you.

- I have/had a leadership role in curriculum planning in my school or district.
- I have/had a leadership role in special education in my school or district.
- I have worked on my state's alternate assessment (e.g., scoring, range finding).
- I have written items for a statewide assessment.

Q18 What is your gender?

- Male
- Female

Q19 What is your ethnicity?

- Hispanic/Latino
- Non-Hispanic/Latino

Q20 What is your race? (Choose one or more)

- White
- Black/African-American
- Asian
- American Indian/Alaska Native
- Native Hawaiian/Other Pacific Islander

Q21 What state do you work in?

- AK
- CO
- IL
- IA
- KS
- MI
- MS
- MO
- NH
- NJ
- NC
- ND

- OK
- PA
- UT
- VT
- VA
- WI
- WV
- Other

Answer If *In which state do you work?* Other Is Selected

Q21b If “Other” was selected, please list the state in which you work.

Q22 Which best describes the population density in your school/workplace?

- Rural (population living outside settlements of 1,000 or less inhabitants)
- Suburban (an outlying residential area of a city of 2,000-49,000 or more inhabitants)
- Urban (city of 50,000 inhabitants or more)

Q23 Will you be able to commit to completing up to four hours of advance training prior to the on-site standard setting meeting?

- Yes
- No

Q24 Will you be able to attend the entire on-site standard setting meeting on June 15-18, 2015?

- Yes
- No

Thank you for completing the survey. DLM plans to notify volunteers who have been selected to serve on panels within 14 days after a recruitment phase ends.

Appendix F: Panels and Assignments

ELA IM 3-5 Panel

1. ELA 3 IM
2. ELA 4 IM
3. ELA 5 IM

ELA IM 6-8 Panel

4. ELA 6 IM
5. ELA 7 IM
6. ELA 8 IM

ELA IM 9-12 Panel

7. ELA 9-10 IM
8. ELA 11-12 IM

Math IM 3-5 Panel

20. Math 3 IM
21. Math 4 IM
22. Math 5 IM

Math IM 6-8 Panel

23. Math 6 IM
24. Math 7 IM
25. Math 8 IM

Math IM 9-11 Panel

26. Math 9 IM
27. Math 10 IM
28. Math 11 IM

Appendix G: Panel Training and Materials



Standard Setting Advance Training

Part 1: Introduction

June 2015

The present publication was developed under grant #4-375100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



WELCOME STANDARD SETTING PANELISTS



DLM® Consortium



ROLES AND EXPECTATIONS OF STANDARD SETTING PANELISTS



Being a Panelist

- Thoughtfully review materials using established standards and procedures
- Use professional judgment in making recommendations
- Participate as a team member
- Maintain security of all training materials
 - Do not make paper or electronic copies
 - Secure electronic devices
 - Do not talk about specifics to others
- Direct questions or concerns to DLM (d1m@ku.edu)



Talking to Others

Please DO

- Talk about what you've learned about
 - the DLM system
 - Assessment of students with significant cognitive disabilities
 - Academic content for students with significant cognitive disabilities

Please DON'T

- Talk about specific texts, items, testlets, or test results



Objectives of Advance Training

- Prepare for onsite standard setting training in June
- Understand:
 - your role as a panelist
 - the population of students
 - the DLM assessment system design
 - how content standards are addressed
 - student results and scoring



On-Site Standard Setting Training

- DLM's approach to standard setting
- Review Performance Level Descriptors
- Review examples of items and testlets
- Review student profiles
- Use materials to help determine cut points



Completing Advance Training

- One module with several activities
- Self-paced
- Must complete before arriving for panel meeting
- Self-assessment



<http://dynamiclearningmaps.org>



Standard Setting Advance Training

Part 2: Overview of DLM System Design

June 2015



The present publication was developed under grant #H323110001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



Part 2: Topics

- Students taking DLM alternate assessment
- System design
 - Learning maps, Essential Elements, claims & conceptual areas/ linkage levels, alignment
- Accessibility by design
 - Accessible content, assessment delivery, dynamic routing



STUDENTS WITH SIGNIFICANT COGNITIVE DISABILITIES TAKING DLM ALTERNATE ASSESSMENTS



DLM General Eligibility Criteria

1. The student has a significant cognitive disability
2. The student is primarily being instructed using the Essential Elements
3. The student requires extensive direct individualized instruction and substantial supports to achieve measurable gains in the grade-and age-appropriate curriculum



1. Significant Cognitive Disability

A disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior.

- *Adaptive behavior is defined as essential for someone to live independently and to function safely in daily life*



2. Content for instruction

The student is learning content based on grade level alternate achievement standards that are linked to the general education standards.

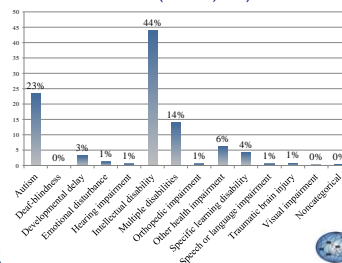


3. Individualized instruction & substantial supports

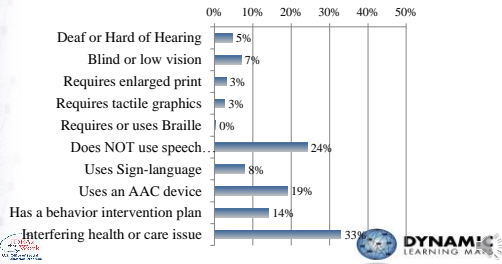
The student requires extensive direct individualized instruction and substantial supports to achieve measurable gains in the grade- and age-appropriate curriculum.



Who are the DLM students? (N = 44,782)



Other Characteristics

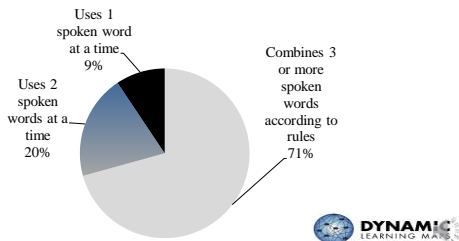


How Students Communicate

Communication Modalities	n	%
Student uses expressive speech for communication	33811	75.9
Student does not use expressive speech for communication	10759	24.1
Total	44,570	100%
For Students who do not use expressive speech for communication:		
Student uses AAC to augment or replace speech	8438	18.9
Student uses Sign Language to augment or replace speech	3435	7.9
Total	11,872	26.8%

Expressive Communication

(w/speech - 75.9% of sample)



Student Use of Computers

(N = 44,439, 89% of sample)

Computer Use	n	%
Accesses a computer independently	25012	56.3
Uses a computer with support (human or assistive technology)	17889	40.3
No opportunity to access a computer	1538	3.5

AAC system used 9 or fewer symbols

AAC system	n
Symbols offered in groups of 1 or 2	4114
Low tech communication board with 8 or fewer symbols	2406
Eye-gaze board w/ 5 or fewer symbols.	83
Voice output communication aide (VOCA) with 9 or fewer symbols	2692
Total	9,295

AAC system used more than 9 symbols

AAC system	Number
VOCA 10-40 Symbols	464
Low tech communication book with 8 or fewer symbols per page	732
Low tech communication board with 9 or more symbols	782
Dynamic Display including tablets	1925
Dynamic Display icon sequencing	350
Total	4,253

Academics

Reading

- 20% do not read words (print or Braille)
- 39% read at first grade level or below

Math

- >50% sort objects by properties
- <20% multiply or divide using numbers or objects



SYSTEM DESIGN: LEARNING MAPS, ESSENTIAL ELEMENTS, AND MORE



Some DLM Features That Impact Standard Setting Procedures

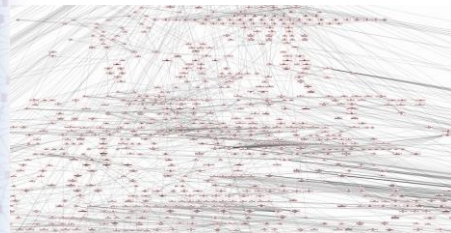
- DLM assessments are based on a fine-grained learning map
- Content complexity is organized by linkage levels
- Results are based on mastery classifications



LEARNING MAPS & ORGANIZATION OF ACADEMIC CONTENT



Learning Maps: A Unique Feature of DLM



A Portion of the Mathematics Learning Map

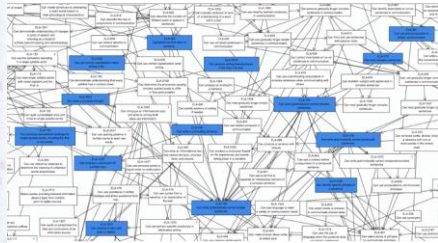
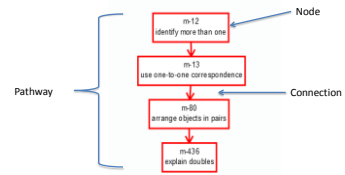


Learning Map

- A graphical representation of a learning theory to depict how English language arts and mathematical ideas develop and are connected to prior understandings
- A network of connected learning targets



Map Terminology



Quick Facts about the DLM Maps

English Language Arts

- 161 foundational nodes
- 1,645 ELA nodes
- 3,982 connections

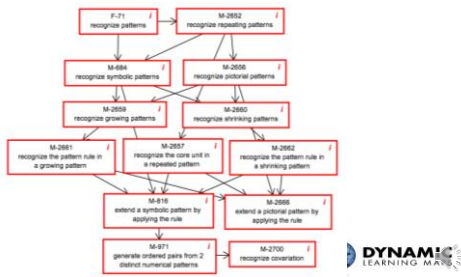
Mathematics

- 161 foundational nodes
- 2,579 mathematics nodes
- 5,360 connections

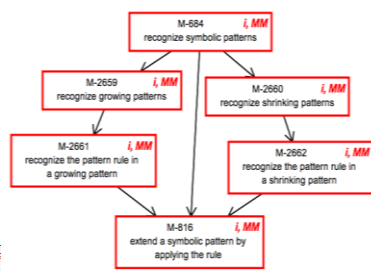
As of January 2015



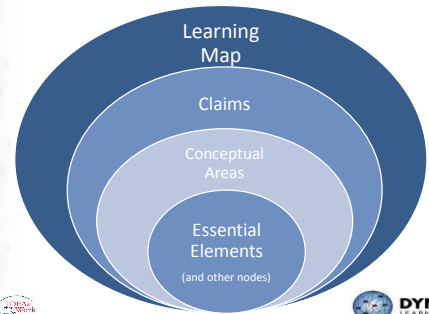
Nodes/Connections: Mathematics Example Patterns



A Closer Look at Some Pattern Nodes



ORGANIZATION OF MAP NODES FOR ASSESSMENT DEVELOPMENT AND REPORTING



ESSENTIAL ELEMENTS



What are Essential Elements (EEs)?

- Links to general education standards at reduced depth, breadth, & complexity
- Provide appropriate level of rigor and challenge
- Focus on the skills (with multiple means of demonstration)
- A starting point for defining achievement standards



Purpose of DLM Essential Elements

- To align to grade level standards without compromising learning and development
- To reflect alignment with grade level standards (horizontal) and across the grades (vertical)
- To specify learning targets for SWSCD, not describe how students engage in instruction or demonstrate understanding



Characteristics of an EE

- Aligns to the grade-level content standard
- Reflects high but reasonable expectations for SWSCD
- Reflects student needs for post-secondary life
- Scope is appropriate and manageable
 - Reflects less complexity, breadth, and depth of the same essential concepts as the grade level standard
- Written in universal terms to allow for demonstration of knowledge in a variety of ways
- Terms are consistent across EEs
- Similar complexity of EEs for same grade span



Example EE for English Language Arts

Grade Level Standard

RL.6.2 Determine a theme or central idea of a text and how it is conveyed through particular details; provide a summary of the text distinct from personal opinions or judgments.

Essential Element

EE.RL.6.2 Determine the theme or central idea of a familiar story and identify details that relate to it.



Example EE for Mathematics

Grade Level Standard

4.G.2. Classify two-dimensional figures based on the presence or absence of parallel or perpendicular lines, or the presence or absence of angles of a specified size. Recognize right triangles as a category, and identify right triangles.

Essential Element

EE.4.G.2. Distinguish between different attributes of shapes (lines, curves, angles).

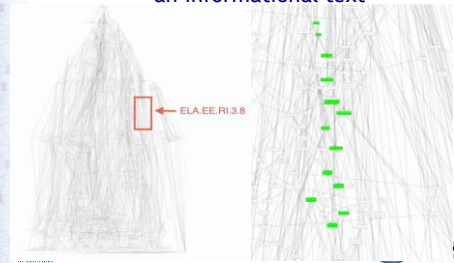


DLM Essential Elements are NOT:

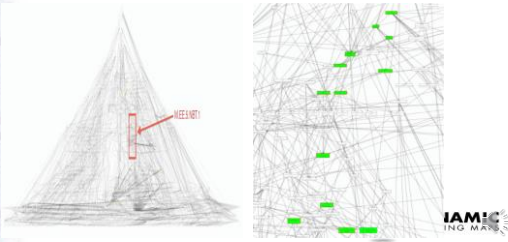
- Replacements for the general education grade level standards
- Downward extensions to pre-K
- Statements of functional skills
- Curriculum or learning progressions
- IEP goals or benchmarks



ELA: Identify two related points the author makes in an informational text

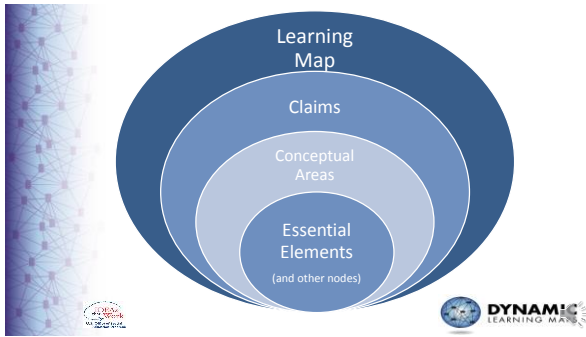


Mathematics: Compare numbers up to 99 using base ten models



CLAIMS AND CONCEPTUAL AREAS





DLM Claims are Tailored for SWSCD

- English Language Arts - Claim 1
 - Students will comprehend text in increasingly complex ways
- Mathematics - Claim 1
 - Students will demonstrate increasingly complex understandings of number sense

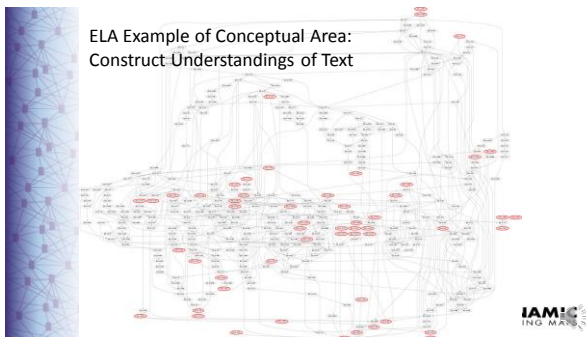
DYNAMIC LEARNING MAPS

English Language Arts	
Major Claims	Conceptual Areas
Students can <i>comprehend text</i> in increasingly complex ways.	Determine Critical Elements of Text
	Construct Understandings of Text
	Integrate Ideas and Information from Text
Students can <i>produce writing</i> for a range of purposes and audiences.	Use Writing to Communicate
	Integrate Ideas and Information in Writing
Students can <i>communicate</i> for a range of purposes and audiences.	Use Language to Communicate with Others
	Clarify and Contribute to Discussion
Students can <i>engage in research/inquiry</i> to investigate topics and present information.	Use Sources and Information
	Collaborate and Present Ideas

DYNAMIC LEARNING MAPS

Mathematics	
Major Claims	Conceptual Areas
Students demonstrate increasingly complex understanding of <i>number sense</i> .	Understand number structures (counting, place value, fraction)
	Compare, compose, and decompose numbers and sets
	Calculate accurately and efficiently using simple arithmetic operations
Students solve increasingly complex mathematical problems, making productive use of <i>algebra and functions</i> .	Use operations and models to solve problems
	Understand patterns and functional thinking
Students demonstrate increasingly complex spatial reasoning and understanding of <i>geometric principles</i> .	Understand and use geometric properties of two- and three-dimensional shapes
	Solve problems involving area, perimeter, and volume
Students demonstrate increasingly complex understanding of <i>measurement, data, and analytic procedures</i> .	Understand and use measurement principles and units of measure
	Represent and interpret data displays

DYNAMIC LEARNING MAPS



LINKAGE LEVELS

DYNAMIC LEARNING MAPS

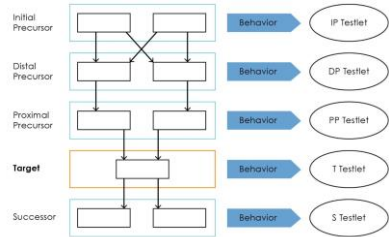
Linkage Levels - A Definition

- Linkage levels (LLs) contain one or more nodes that connect to an identified EE.
- LLs both identify important milestones en route to an EE and specify where a student is in relationship to the grade-span target.

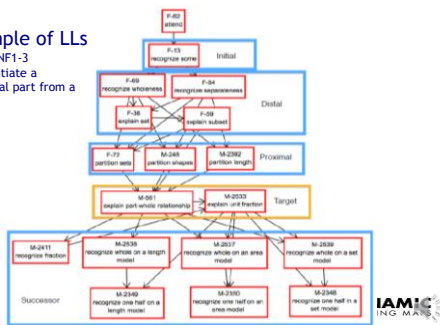


Testlets in Linkage Levels

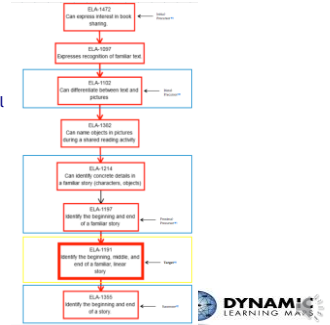
Connect the map to the items developed



Example of LLs M.EE.3.NF1-3 Differentiate a fractional part from a whole



Example of LLs ELA.EE.RL.3.5 Determine the beginning, middle, and end of a familiar story with a logical order



Important Information on LLs

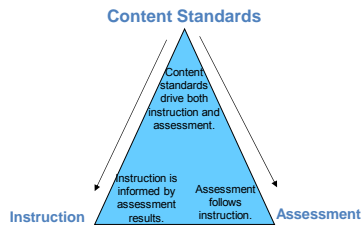
- All testlets written to the initial precursor level are administered offline to the student.
- The distance in the learning map between LLs gets larger in later grades.
- Students are assigned to a linkage level based on First Contact survey data when starting assessment.



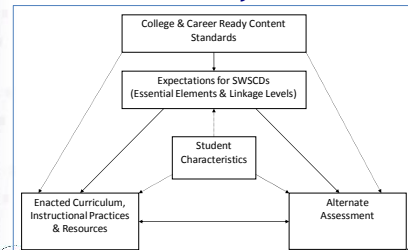
A NOTE ABOUT ALIGNMENT



Alignment in General



The Educational System in DLM



ACCESSIBILITY BY DESIGN



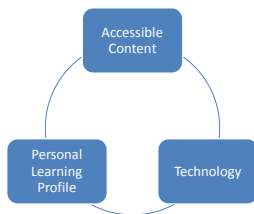
Universal Design

DLM assessments are based on principles of Universal Design (UD)

- UD considers all students from the outset and provides flexibility in the ways students respond or demonstrate knowledge and skills
- UD reduces barriers by incorporating appropriate accommodations and supports
- UD promotes high achievement expectations by engaging students in rigorous content



DLM Accessibility



ACCESSIBLE CONTENT

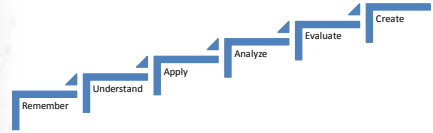


Accessible Content: Strategies

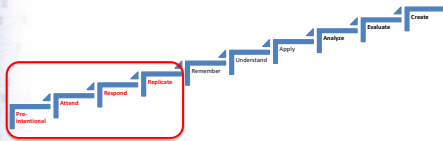
- Multiple and alternate pathways
- Multiple levels of complexity for each EE
 - linkage levels



Revised Bloom's Taxonomy (RBT)



DLM's Taxonomy



DLM Expansion of RBT

Process	Definitions
Pre-intentional	Behavior reflects a general state, but does not reflect intentional behavior. Intent is inferred by others (e.g., teachers, parents) through facial expressions, movements, or sounds.
Attend	Orients to objects, people, or activity. Indicates selective attention to stimuli in the academic learning environment.
Respond	Intentional response using any mode of expression. Indicates joint attention to materials and activities in the academic learning environment.
Replicate	Perform rote task in familiar or practiced context.



Accessible Content: Strategies

- Careful attention to unnecessary text complexity
- Vocabulary
- Items that can differentiate between a non-mastery performance and a mastery performance



Accessible Content: Strategies

- Item writing guidelines based on principles of Universal Design
 - Alternate forms where necessary
- Items tagged to accessibility features



Accessible Content: Strategies

- Accessibility review
 - Multiple times during the test development process
- Criteria



PERSONAL NEEDS AND PREFERENCES (PNP)



Personal Needs and Preferences



Three Categories of Accessibility Features

1. Supports provided within the system via PNP
2. Supports requiring additional tools/materials
3. Supports provided outside the system



Category 1: Supports Provided Within the System via PNP

- Magnification
- Invert Color Choice
- Color Contrast
- Overlay Color
- Read Aloud with Highlighting-Text to Speech (TTS)
 - Text Only
 - Text and Graphics
 - Non-Visual



Category 2: Supports Requiring Additional Tools/Materials

- Uncontracted Braille
- Single-switch system/PNP enabled
- Two-switch system
- Administration via iPad
- Adaptive equipment used by student
- Individualized manipulatives



Category 3: Supports Provided Outside the System

- Human Read Aloud
- Sign interpretation of text
- Language translation of text
- Test administrator enter responses for students
- Partner-Assisted Scanning (PAS)



Supports not available in DLM

Because students participating in DLM also have significant cognitive disabilities, these supports are not appropriate for delivery in DLM assessments.

- Sign language videos
- Standardized tactile graphics



Supports not allowed in DLM

- Repeating the question again, even after the student has selected a response, in order to prompt the student to choose a different answer
- Using physical prompts or hand-over-hand guidance
- Removing answer options or giving content hints
- Modifying the content of a performance task in a computer-administered testlet in an attempt to help the student arrive at the correct response



DYNAMIC ROUTING



First Testlet

- First Contact Survey
 - Prior academic performance
 - Expressive communication



Levels & Bands

Testlet Linkage Level	Student First Contact Complexity Band
Initial Precursor	0
Distal Precursor	1
Proximal Precursor	2
Target	3
Successor	4



How First Contact May Impact Initial Testlet Level

Student Characteristics

- Uses only 1 signed word at a time
- Does not read any words when presented in print
- Is able to sort objects by common properties (e.g. color) up to 20% of the time

Assignment of a Level

- Initial Precursor level testlet



How First Contact May Impact Initial Testlet Level

Student Characteristics

- Regularly combines 3 or more spoken words for a variety of purposes
- Is able to read print above the 3rd grade level
- Is able to add or subtract by joining or separating groups of objects 90% of the time

Assignment of a Level

- Target level testlet



Dynamic Routing

After first testlet:

- Changes in linkage levels
- Routing within the structure of the learning map



Testlet Delivery

Part 1	Part 2	Part 3	Part 4
IP	IP	IP	IP
DP	DP	DP	DP
PP	PP	PP	PP
T	T	T	T
S	S	S	S



Testlet Delivery

Part 1	Part 2	Part 3	Part 4
IP	IP	IP	IP
DP	DP	DP	DP
PP	PP	PP	PP
T	T	T	T
S	S	S	S



The present publication was developed under grant #4-3731-10301 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



Standard Setting Advance Training

Part 3: Assessment Design

June 2015

The present publication was developed under grant #4-375100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



ITEM TYPES



Computer-Administered Items

- Single selection multiple choice
- Multi-selection multiple choice
- Matching
- Sorting: drag & drop, click to place
- Select text



Single-select multiple choice

What does this text tell about?

- how to pet a nice dog
- how to feed a hungry dog
- how to find a lost dog



Single-select multiple choice

Which is a circle?



Multi-select multiple choice

Select all of the coins.



Matching items



Match the signs with their names.

-	→	Addition
=	→	Equal sign
+	→	Minus sign



Sorting

Put one shape into each box.



Select Text

Choose the word that is a number.

Sam likes . Sam has dogs. Sam with his dogs.



Teacher-Administered Items


Educator Directions:

SAY: "There are clothes in the story."
SHOW: one familiar object that belongs to the student and two unfamiliar objects.
ASK: "Where is your (familiar object name)?"

Record student response:

Indicates familiar object
Indicates unfamiliar object

Attends to other stimuli
No response

BACK  NEXT



TESTLETS



Computer-Delivered Testlets

- Tests delivered directly to students via computer (KITE system)
- Designed to allow students to interact independently with the computer, using assistive technologies or human support as necessary
- More common at upper linkage levels



Teacher-Administered Testlets

- Designed for administration to the student outside the KITE system
 - Educator directions
 - Scripted statements/interactions
- Test administrator records responses in KITE system
- Most common at lower linkage levels (with some exceptions)



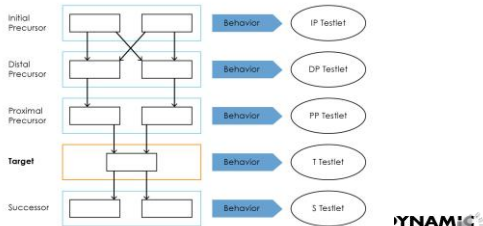
Engagement Activities

- Reading: Read entire text once before reading it again with embedded and concluding items. Shared reading is an option on the first reading.
- Writing: Activities can include choosing a topic or reviewing potential topics of interest before making a choice
- Math: Context at beginning, carried throughout



Testlets at Linkage Levels

Connect the map to the items developed



M.EE.7.G.4 Determine the perimeter of a rectangle by adding the measures of the sides

- **Initial Precursor**
 - Node: F-65 recognize attribute values
- **Distal Precursor**
 - Nodes: M-289 describe measurable attributes, M-194 recognize measurable attributes
- **Proximal Precursor**
 - Nodes: M-397 explain length, M-587 explain perimeter
- **Target**
 - Nodes: M-2413 calculate the perimeter of a rectangle by counting unit lengths on a grid, M-592 calculate perimeter by adding all the side lengths
- **Successor**
 - Node: M-1908 use coordinates to calculate perimeters of polygons



ELA: Reading Literature, Grade 3

ELA.EE.RL.3.5 Determine the beginning, middle, and end of a familiar story with a logical sequence



ELA Example

- **Initial Precursor**
 - Node: ELA-1472 Can express interest in book sharing
- **Distal Precursor**
 - Node: ELA-1102 Can differentiate between text and pictures
- **Proximal Precursor**
 - Nodes: ELA-1214 Can identify concrete details in a familiar story; ELA-1197 Identify the beginning and end of a familiar story
- **Target**
 - Node: ELA-1191 Identify the beginning, middle, and end of a familiar, linear story
- **Successor**
 - Node: ELA-1355 Identify the beginning and end of a story



Testlet Navigation

Read this sentence.

If this plan fails, we will have to go back to the drawing board.

In this sentence, what is the meaning of the underlined phrase?

start all over support the decision work faster avoid mistakes

BACK



NEXT



Responding to Items

What does this text tell about?

how to pet a nice dog
how to feed a hungry dog
how to find a lost dog

BACK



NEXT



Responding to Items

What does this text tell about?

how to pet a nice dog

how to feed a hungry dog
how to find a lost dog

BACK



NEXT



Responding to Items

What does this text tell about?

how to pet a nice dog
how to feed a hungry dog

how to find a lost dog

BACK



NEXT



Finishing the Testlet

Review: The Outstanding Ostrich

Are you done?

Red boxes mean you are not done. Ask your teacher for help.

Green checks mean you are done. You can choose end.

1	2	3	4	5	6	7	8	9
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
10	11	12						
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>						

Answered Not answered

GO BACK

END



SAMPLE ITEMS



MATHEMATICS ITEMS



6th Grade Initial Precursor Level Node F-84 Recognize separateness



Educator Directions:

SHOW: the bundled pencils.
SAY: "Here are some pencils."

SHOW: the separate pencils.
SAY: "Here are some more pencils."

SHOW: the bundled pencils and the separate pencils.
SAY: "Show me the separate pencils."



Record student response:

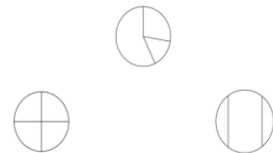
- Indicates the separate pencils
- Indicates the group of bundled pencils
- Indicates one pencil or all of the pencils
- Attends to other stimuli
- No response



6th Grade Distal Precursor Level Node M-373 Model Equal Parts



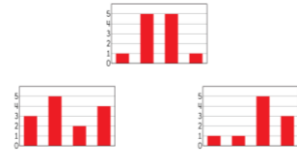
Which shape is divided into equal parts?



6th Grade Proximal Precursor Level
Node M-2679
Recognize Symmetric Distribution



Which graph has a symmetric data distribution?



6th Grade Target Level
Node 1044
Recognize Equivalent Algebraic Expressions



Ruth writes an expression.

$$9 + 0$$

Which expression is equivalent to Ruth's expression?

$$9 + 1$$

$$9 \times 1$$

$$9 \times 0$$



6th Grade Successor Level
Node 1162
Solve real world problems by representing them in algebraic equations



Dave has \$100. Dave spends some money on toys. Dave has \$40 left. Dave writes an equation to see how much money he spent. Which equation does

Dave write?

$$100 + 40 = y$$

$$y + 100 = 40$$

$$100 - y = 40$$



ELA ITEMS



F-146 Can match a real object with a picture or other symbolic representation of the object*

Initial Precursor



There is a pencil in a desk.

SHOW: the pencil, paper, and ruler.

SAY: "Show me one that is the same as the picture."



Record student response:

- Indicates the matching object
- Indicates one of the other objects
- Indicates multiple objects
- Attends to other stimuli
- No response



F-121 Can identify familiar people, objects, places, and events

Distal Precursor



Educator Directions:

SAI: "There is a hat in the story."

SHOW: one familiar object that belongs to the student that was not used in the previous item and two unfamiliar objects.

SAY: "Where is your (familiar object name)?"

Record student response:

- Indicates familiar object
- Indicates unfamiliar object
- Indicates multiple objects
- Attends to other stimuli
- No response



ELA-1204 Can answer who and what questions about details in a familiar narrative

Proximal Precursor



What did Ana find?

a fuzzy hat

a dirty shoe

a shiny rock



ELA-1678 Can answer who and what questions about details in a narrative

Target



Who fussed and whined?

Mom and Dad

Dad and Ramona

Ramona and Beezus



ELA-786 Can answer where questions about details in a narrative

Successor



Where does Mudge like to run?

in the house

in the school

in the woods



Writing Assessments

Emergent -

- combines initial and distal precursor levels
- for students who do not yet have or are working on early symbolic understanding

Conventional -

- combines proximal precursor, target and successor levels
- for students who have symbolic understanding and can use writing tools to communicate



Educator Directions:

SHOW: the object.

SAY: "What is a word that tells about the (object)?"

WAIT AND OBSERVE: Give the student time to communicate a word. If the student is able to write the word, encourage him or her to do so. After you have given the student an opportunity to write or communicate a word, select "NEXT."



Choose the highest level that describes your observation.

- Writes the word that describes the object
- Communicates a word that describes the object
- Writes or communicates a word that is unassociated with the object
- Attends to other stimuli
- No response



<http://dynamiclearningmaps.org>



The present publication was developed under grant #4-3751-10001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



THE INTEGRATED ASSESSMENT MODEL (IM)



Assessment Models

- Blueprint (content coverage)
- Testlet construction
- Frequency, timing, number of testlets



Content of the IM Assessments

- Grades 3-8 and high school
 - English Language Arts (ELA)
 - Mathematics
- Blueprints:
 - Consortium approved a subset of Essential Elements
 - Consortium set minimum requirements for breadth of coverage
 - 6-8 in mathematics
 - 8-11 in ELA



IM ELA Content Coverage Example

3rd grade:

- At least **three** EEs in C1.1 including RL and RI (eight available)
- At least **two** EEs in C1.2 from different strands (five available)
- At least **one** EE in C1.3 (two available)
- All students take **one** writing assessment (one available)



IM Math Content Coverage Example

3rd grade:

- At least **two** EEs from two conceptual areas in claim 1 (four available)
- One** EE in claim 2 (one available)
- At least **two** EEs in claim 3 (three available)
- At least **one** EE in claim 4 (three available)



Complete IM Test Blueprints

Grade 4: Available Essential Elements and minimum expectation for each student's assessment

Conceptual Area	EE	Description
SIA.1.1	EE.1.1.1	Identify the main idea or central theme of a text.
	EE.1.1.2	Identify the main idea or central theme of a text.
	EE.1.1.3	Identify the main idea or central theme of a text.
	EE.1.1.4	Identify the main idea or central theme of a text.
	EE.1.1.5	Identify the main idea or central theme of a text.
SIA.1.2	EE.1.2.1	Identify the main idea or central theme of a text.
	EE.1.2.2	Identify the main idea or central theme of a text.
	EE.1.2.3	Identify the main idea or central theme of a text.
	EE.1.2.4	Identify the main idea or central theme of a text.
	EE.1.2.5	Identify the main idea or central theme of a text.
SIA.1.3	EE.1.3.1	Identify the main idea or central theme of a text.
	EE.1.3.2	Identify the main idea or central theme of a text.
	EE.1.3.3	Identify the main idea or central theme of a text.
	EE.1.3.4	Identify the main idea or central theme of a text.
	EE.1.3.5	Identify the main idea or central theme of a text.
SIA.1.4	EE.1.4.1	Identify the main idea or central theme of a text.
	EE.1.4.2	Identify the main idea or central theme of a text.
	EE.1.4.3	Identify the main idea or central theme of a text.
	EE.1.4.4	Identify the main idea or central theme of a text.
	EE.1.4.5	Identify the main idea or central theme of a text.



How Testlets Are Constructed

All except writing

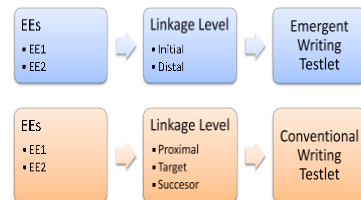
- Each testlet assesses one linkage level for one Essential Element
- Engagement activity and 3-5 items

Writing

- Each testlet assesses emerging or conventional writing
- Testlet includes nodes from multiple Essential Elements



Writing Testlets



HOW AND WHEN TESTLETS ARE DELIVERED



Instructionally Embedded

Testlets can be delivered for any available Essential Element (Minimum blueprint EEs are required)

System recommends linkage level; teacher can accept or override

Student performance contributes information to scores used for summative purposes

Student performance informs starting place for spring testing

Spring Testing

Five testlets are selected to cover a subset of the blueprint (retesting previously assessed content)

Dynamic routing supports adaptivity between testlets

Scores are used for summative purposes



Instructionally Embedded Assessment

- One testlet per EE chosen in ITI
 - Teacher chooses linkage level
- Each testlet assigned separately
- Test administrator schedules sessions within testing windows



Spring Assessment

- Purpose: to update and validate information about student performance
- Five testlets per subject
- Each testlet assigned separately
- Test administrator schedules sessions within window



Dynamic Routing

Instructionally Embedded:

- System recommends, teacher can override

Spring Window:

- System selects Essential Element and linkage level



<http://dynamiclearningmaps.org>

The present publication was developed under grant #4-3731-10301 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



Standard Setting Advance Training

Part 4: Introduction to Standard Setting

June 2015

The present publication was developed under grant #4-375100001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



ASSESSMENT RESULTS AND SCORE REPORTS



Unlike Other Assessments

- No raw scores, percentages, or scale score
- Starts from small grain size (node) and builds up to information about the whole subject
- Will be based on mastery of EEs and linkage levels assessed



Starting with the Node

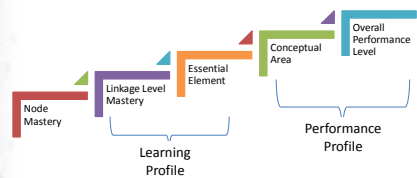
Using all student responses to items for that node:



Working Up



Reports



Learning Profile

- Shows mastery status for every linkage level for each Essential Element
 - EEs grouped by conceptual area
 - Each EE in one row
 - Linkage levels in columns
 - Shading to show linkage level mastery



Learning Profile

Area	Essential Element	Level				Success
		Initial Proficiency	Developing Proficiency	Proficient Proficiency	Target	
ELA	ENEL2	Measurable expectations in standards that are not specified together. Students will use a group of elements to write an expository text.	Exceeds the skill. Proficient on the quantity based by one or more items in performed into separate, clear with two or more repeat words.	Exceeds the skill. Proficient on the quantity based by one or more items in performed into separate, clear with two or more repeat words.	Exceeds the skill. Proficient on the quantity based by one or more items in performed into separate, clear with two or more repeat words.	Exceeds the skill. Proficient on the quantity based by one or more items in performed into separate, clear with two or more repeat words.

Linkage Level Descriptors

Essential Element Code



Performance Profile

- Number of linkage levels mastered within each conceptual area
- Overall performance in the subject
- Narrative statements about skills the student mastered



Performance Profile

Individual Student Year-End Report Performance Profile



NAME: Susie Smith SCHOOL: DIM School YEAR: 2014-15
 SUBJECT: English Language Arts DISTRICT: DIM District GRADE: 3
 REPORT DATE: 06-10-2015 STATE: DIM State STATE ID: 08691

Overall Results

Grade 3 English language arts allows students to show their achievement in 85 skills related to 17 Essential Elements. Susie has mastered 32 of those 85 skills during the 2014-15 school year. Overall, Susie's mastery of English language arts fall into the second of four performance categories: **approaching the target**. The specific skills Susie has and has not mastered can be found in her Learning Profile.



INTRODUCTION TO STANDARD SETTING



Performance Profile

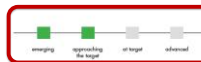
Individual Student Year-End Report Performance Profile



NAME: Susie Smith SCHOOL: DIM School YEAR: 2014-15
 SUBJECT: English Language Arts DISTRICT: DIM District GRADE: 3
 REPORT DATE: 06-10-2015 STATE: DIM State STATE ID: 08691

Overall Results

Grade 3 English language arts allows students to show their achievement in 85 skills related to 17 Essential Elements. Susie has mastered 32 of those 85 skills during the 2014-15 school year. Overall, Susie's mastery of English language arts fall into the second of four performance categories: **approaching the target**. The specific skills Susie has and has not mastered can be found in her Learning Profile.



4. Analysis of Impact Data

- Data collected by May 13, 2015
- Impact data is shared to show the number and percent of students who would be classified to each level
- Panelists review and discuss the results within and across grades
- Final ratings



5. Review of Results and Final Recommendations

- Staff compile the final judgments of panelists and make final recommendation for each level
- DLM Technical Advisory Committee evaluates the recommendations and provides feedback
- State partners review the results of the panel process, including recommended cut points within and across grades



<http://dynamiclearningmaps.org>



The present publication was developed under grant #4-373410001 from the U.S. Department of Education, Office of Special Education Programs. The views expressed herein are solely those of the author(s), and no official endorsement by the U.S. Department should be inferred.



Dynamic Learning Maps Standard Setting Panel Meeting

June 15-18, 2015



WELCOME AND INTRODUCTIONS



2



Opening Remarks

Neal Kingston
DLM Project Director



3



Introductions

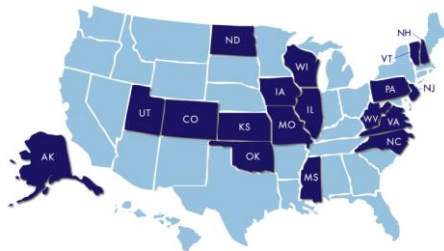
Meagan Karvonen
DLM Associate Director



4



Consortium



5



Two Testing Models

- Integrated
- Year-End
 - Includes two states that use End of Instruction (EOI) courses at the HS level



6



Welcome Panelists!

- Recruited by DLM partner states
- Over 100 panelists across both testing models
- Average of 15+ years of experience in academics and/or students with significant cognitive disabilities (range from <1 to 40+ years)



7



Welcome Panelists!

- More than half are classroom teachers
 - Others include building / district staff, university faculty, others
- Majority have some previous assessment leadership experience
 - Other standard settings
 - DLM item writing, external review



8



Brief Introductions

Groups represented:

- Panelists
- DLM staff
 - Facilitators, room staff, psychometricians, leadership
- Observers
 - State partners, Technical Advisory Committee



9



Procedures

- Time slots in agenda & breaks
- Pacing
- Index cards
 - And questions from the self-assessments



10



Housekeeping



11



TRAINING: PART 1



12



SECURITY AND CONFIDENTIALITY



13



Reminders from Advance Training

- All materials are secure - kept in the meeting room
- No electronic devices out during panel work
 - Emergencies
- No discussion of the detailed processes, deliberations, or results outside of this meeting
 - Wrap-up on last day



14



Confidentiality Agreement

- What questions do you have?
- Give signed copy to your facilitator this morning



15



DLM SYSTEM DESIGN: KEY POINTS



16



Purpose

- Refresher on most essential concepts
- Address questions/concerns from advance training
 - Areas with less confidence
 - Questions posed



17



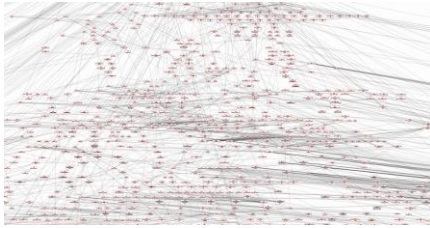
LEARNING MAPS & ORGANIZATION OF ACADEMIC CONTENT



18

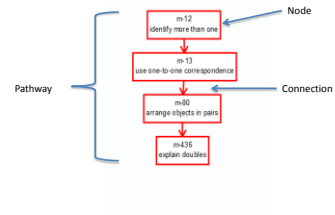


Learning Maps: A Unique Feature of DLM



A Portion of the Mathematics Learning Map

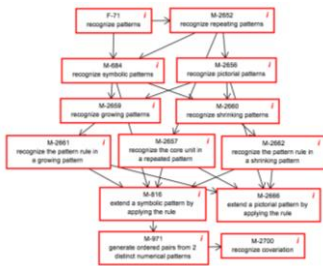
Map Terminology



20



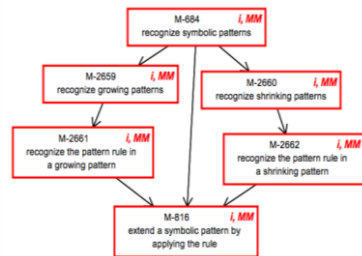
Nodes/Connections: Mathematics Example Patterns



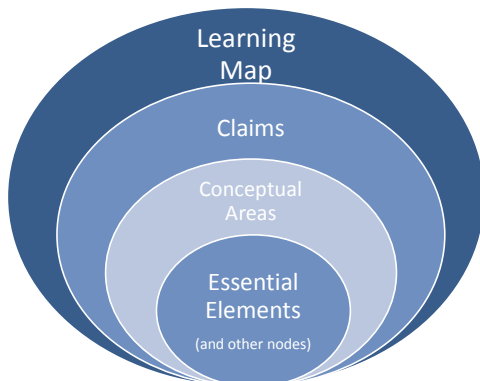
21



A Closer Look at Some Pattern Nodes



22



DLM Claims are Tailored for SWSCD

- English Language Arts - Claim 1
 - Students will comprehend text in increasingly complex ways
- Mathematics - Claim 1
 - Students will demonstrate increasingly complex understandings of number sense

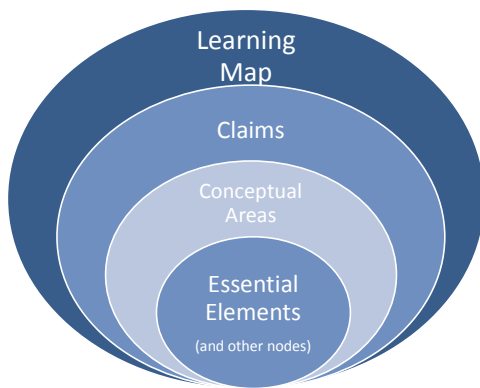


24



English Language Arts	
Major Claims	Conceptual Areas
Students can <i>comprehend text</i> in increasingly complex ways.	Determine Critical Elements of Text
	Construct Understandings of Text
	Integrate Ideas and Information from Text
Students can <i>produce writing</i> for a range of purposes and audiences.	Use Writing to Communicate
	Integrate Ideas and Information in Writing
Students can <i>communicate</i> for a range of purposes and audiences.	Use Language to Communicate with Others
	Clarify and Contribute to Discussion
Students can <i>engage in research/inquiry</i> to investigate topics and present information.	Use Sources and Information
	Collaborate and Present Ideas

Mathematics	
Major Claims	Conceptual Areas
Students demonstrate increasingly complex understanding of <i>number sense</i> .	Understand number structures (counting, place value, fraction)
	Compare, compose, and decompose numbers and sets
Students solve increasingly complex mathematical problems, making productive use of <i>algebra and functions</i> .	Calculate accurately and efficiently using simple arithmetic operations
	Use operations and models to solve problems
Students demonstrate increasingly complex spatial reasoning and understanding of <i>geometric principles</i> .	Understand patterns and functional thinking
	Understand and use geometric properties of two- and three-dimensional shapes
Students demonstrate increasingly complex understanding of <i>measurement, data,</i> and analytic procedures.	Solve problems involving area, perimeter, and volume
	Understand and use measurement principles and units of measure
	Represent and interpret data displays



What are Essential Elements (EEs)?

- Links to general education standards at reduced depth, breadth, & complexity
- Provide appropriate level of rigor and challenge
- Focus on the skills (with multiple means of demonstration)
- A starting point for defining achievement standards



28



Example for English Language Arts

- | | |
|--|---|
| <p>Common Core State Standard</p> <ul style="list-style-type: none"> • RL.6.2 Determine a theme or central idea of a text and how it is conveyed through particular details; provide a summary of the text distinct from personal opinions or judgments. | <p>Essential Element</p> <ul style="list-style-type: none"> • EE.RL.6.2 Determine the theme or central idea of a familiar story and identify details that relate to it. |
|--|---|



29



Example for Mathematics

- | | |
|--|--|
| <p>Common Core State Standard</p> <ul style="list-style-type: none"> • 4.MD.5. Recognize angles as geometric shapes that are formed wherever two rays share a common endpoint, and understand concepts of angle measurement: • An angle is measured with reference to a circle with its center at the common endpoint of the rays, by considering the fraction of the circular arc between the points where the two rays intersect the circle. An angle that turns through $1/360$ of a circle is called a "one-degree angle," and can be used to measure angles. • An angle that turns through n one-degree angles is said to have an angle measure of n degrees. | <p>Essential Element</p> <ul style="list-style-type: none"> • EE.4.MD.5. Recognize angles in geometric shapes |
|--|--|



30



DLM Essential Elements are NOT:

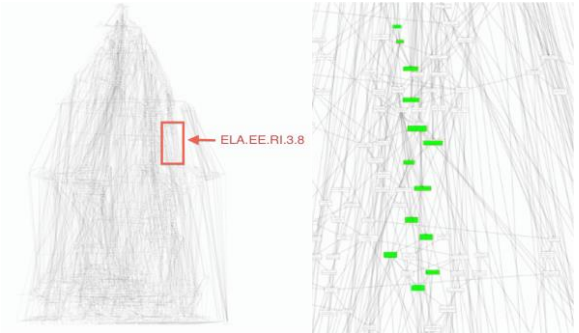
- Replacements for the general education grade level standards
- Downward extensions to pre-K
- Statements of functional skills
- Curriculum or learning progressions
- IEP goals or benchmarks



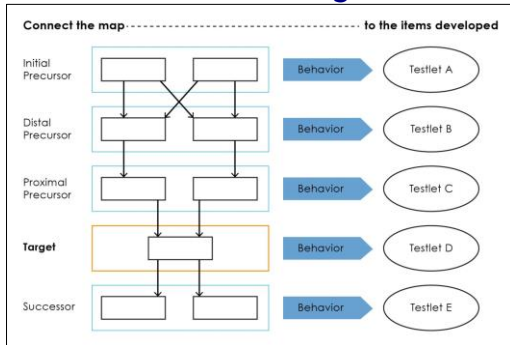
31



Identify two related points the author makes in an informational text



Testlets in Linkage Levels



Linkage Levels - A Definition

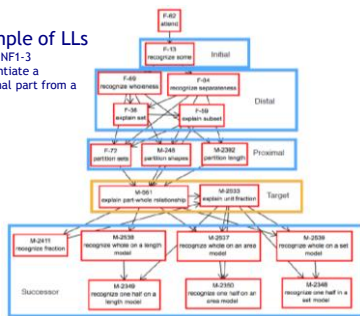
- Linkage levels (LLs) contain one or more nodes that connect to an identified EE.
- LLs both identify important milestones en route to an EE and specify where a student is in relationship to the grade-span target.



34



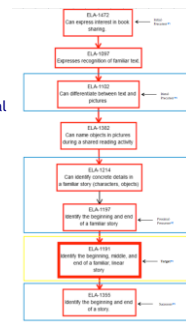
Example of LLs
M.EE.3.NF1-3
Differentiate a fractional part from a whole



35



Example of LLs
ELA.EE.RL.3.5
Determine the beginning, middle, and end of a familiar story with a logical order



36



Important Information on LLs

- All testlets written to the initial precursor level are administered offline to the student.
- The distance in the learning map between LLs gets larger in later grades.
- Students are assigned to a linkage level based on First Contact survey data when starting assessment.



37



TESTLETS



38



Computer-Delivered Testlets

- Tests delivered directly to students via computer (KITE system)
- Designed to allow students to interact independently with the computer, using assistive technologies or human support as necessary
- More common at upper linkage levels



39



Teacher-Administered Testlets

- Designed for administration to the student outside the KITE system
 - Educator directions
 - Scripted statements/interactions
- Test administrator records responses in KITE system
- Most common at lower linkage levels (with some exceptions)



40



Adaptive Delivery

System has testlets available at all linkage levels for every part of the test

Students take testlet from one level for each EE or group

EE 1	EE 2	EE 3	EE 4
IP	IP	IP	IP
<u>DP</u>	DP	DP	<u>DP</u>
PP	<u>PP</u>	<u>PP</u>	PP
T	T	T	T
S	S	S	S

Questions?

- Structure of content
- What testlets look like
- Students who take DLM assessments
- Accessibility



42



TRAINING PART 2: STUDENT RESULTS & MASTERY PROFILES



43



Purpose

- Whole standard setting process is based on information about student performance
- Important to understand **conceptually** (not statistically) so you can complete your ratings



44



What Counts?

Integrated Model

- All instructionally embedded testlets and all spring testlets

Year End and EOY

- All spring testlets



45



How DLM Assessments are Scored

- There is no raw score (total correct)
- There is no scale score
- Student has mastered or not mastered each node
 - Multiple ways to be a master of a node



46



How DLM Assessments are Scored

- Every assessment taken updates what is known about the student's mastery probability for all nodes
- At end of year, take a snapshot of the student's mastery probabilities



47



Starting with the Node: Mastery Probability

Using all student responses to items for that node:



48



How DLM Assessments are Scored

- Combine information about nodes to get to **linkage level mastery**



49



Example: 5th Grade Math (PP)

Nodes

- Recognize the hour hand
- Know hours on a clock
- Recognize the hour on a digital clock
- Recognize the minute hand
- Recognize the minute on a digital clock

Linkage Level Descriptor

- Recognize measurable attributes



50



Example: 3rd Grade ELA (DP)

Node

- Can demonstrate understanding of the names of objects or people who are not immediately present

Linkage Level Descriptor

- Understand words for absent objects and people



51



Working Up From the Node



Unit of mastery considered in DLM standard setting procedure

Mastery Profile

- Shows mastery status for every linkage level for each Essential Element
 - EEs grouped by conceptual area
 - Each EE in one row
 - Linkage levels in columns
 - Shading to show linkage level mastery



53



Mastery Profile

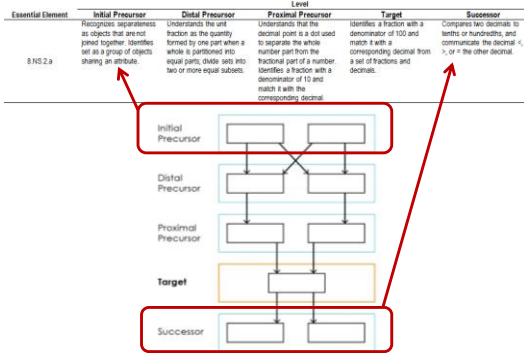
Conceptual Area

Essential Element	Initial Proficiency	Developing Proficiency	Proficient Proficiency	Target	Successor	
0211	0302a	Recognizes representations on maps for the use of a group of objects sharing an attribute.	Understands the use of maps on the spatial coordinate, one part when a map is used to show spatial relationships.	Understands that the development of a map uses a particular set of symbols and that a map is a two-dimensional representation of a three-dimensional object.	Identifies a fraction with a denominator of 10 and compares the fraction to another fraction from the set of fractions with a denominator of 10 and identifies a fraction with a denominator of 10 and compares it with the corresponding decimal.	Compares two fractions with a denominator of 10 and compares the decimal to another decimal.

Linkage Level Descriptors

Essential Element Code

Example: Linkage Level Mastery



Mastery Profiles

End of Year Learning Profile

SUBJECT: Mathematics
MODEL: Integrated

GRADE: 8
PROFILE ID: 0003

YEAR: 2014-15
TOTAL LL: 20

Area	Essential Element	Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
C1.1	8.NS.2.a	Recognize representations as objects that are not joined together. Identify sets as a group of objects sharing an attribute.	Understands the unit fraction as the quantity formed by one part when a whole is partitioned into equal parts; divide sets into two or more equal subsets.	Understands that the decimal point is a dot used to separate the whole number part from the fractional part of a number; identifies a fraction with a denominator of 10 and match it with the corresponding decimal.	Identifies a fraction with a denominator of 100 and communicate the decimal $\frac{1}{100}$, $\frac{1}{10}$, or $\frac{1}{1000}$ as the other decimal.	Compares two decimals to tenths or hundredths, and communicate the decimal $\frac{1}{10}$, $\frac{1}{100}$, or $\frac{1}{1000}$ as the other decimal.
C1.2	8.NS.2.b	Recognize representations as objects that are not joined together.	Identifies a unit model that represents one tenth, and recognize multiple tenths, such as two tenths, five tenths, eight tenths, etc.	Chooses a fraction from a set of three fractions that accurately represent a given decimal (tenths or hundredths).	Compares two decimals to thousandths, and communicate the decimal $\frac{1}{1000}$, $\frac{1}{100}$, or $\frac{1}{10}$ as the other decimal.	Compares two decimals to thousandths, and communicate the decimal $\frac{1}{1000}$, $\frac{1}{100}$, or $\frac{1}{10}$ as the other decimal.

- Green = level was mastered this year
- Total LL = number of linkage levels mastered across all tested EEs

Evidence of Mastery

- Statistical modeling of the map (relationships between nodes) tells us which other nodes the student is likely to have mastered
 - If mastered nodes at target level, then highly likely to master nodes at previous linkage levels
 - Constraint: cannot demonstrate mastery of nodes at level beyond the one that was tested

Area	Grade Level Expectations	Level				
		1	2	3	4	5
Understanding Central Elements of Text	RL.3.1 Answer who and what questions to demonstrate understanding of details in a text.	Attend to object characteristics	Identify familiar people, objects, places, or events	Answer who and what questions to identify details in a familiar story	Answer who and what questions about story details	Answer who, what, when, and where questions about story details
	RL.3.2 Associate details with events to answer from diverse cultures.	Seek about objects	Identify familiar people, objects, places, or events	Associate details with events in a familiar story	Associate details with events to diverse stories	Relevant diverse stories with key details
	RL.3.3 Identify the feelings of characters in a story.	Identify feeling states in self	Identify feeling words	Identify the feelings of characters in familiar stories	Identify the feelings of characters in a story	Identify character feelings and relate to writers
	RL.3.5 Determine the beginning, middle, and end of a familiar story with a logical order.	Express interest in book sharing	Differentiate between text and pictures	Identify details and beginning and end of a story	Determine the beginning, middle and end of a familiar story with a logical order	Identify beginning and end of a story with relevant information
	RL.3.1 Answer who and what questions to demonstrate understanding of details in a text.	Attend to object characteristics	Identify familiar people, objects, places, or events	Identify concrete details in an informational text	Identify concrete detail in an informational text	Identify explicit details in informational texts
Understanding Text	RI.3.2 Identify details in a text.	Seek about objects	Attend to object characteristics	Identify illustrations that are familiar text	Identify concrete detail in an informational text	Identify explicit details in informational texts
	RI.3.5 With guidance and support, use text features including headings and key words to locate information in a text.	Seek about objects	Identify familiar people, objects, places, or events	Identify illustrations that are familiar text	Use basic text features to find information	The specific text features to locate information
	RI.3.4 Determine words and phrases that convey their meanings in a text.	Attend to object characteristics	Understand words for absent objects and people	Identify similar or opposite word meanings	Use words to complete a sentence from a story	Identify the meaning of an unknown word using basic context
Understanding Text	RI.3.5 Determine the meaning of words and phrases in context.	Attend to object characteristics	Identify familiar people, objects, places, or events	Identify similar or opposite meaning	Determine literal meaning of words and phrases	Use words to complete meaningful sentences
	RI.3.5.a Identify words that describe the personal emotional state.	Identify feeling states in self	Understand common feeling words	Identify feeling words	Identify feeling words in personal state	Describe the internal and external character traits
Using Writing to Communicate	W.3.2a Write a topic and write about it including one fact or detail.	Identify feeling states in self	Understand common feeling words	Identify feeling words	Write about a topic by producing facts and details	Independently select a topic and produce relevant facts and details
	W.3.2 With guidance and support produce writing that expresses more than one idea.	Seek about objects	Express interest in making marks on paper	Express interest in making marks on paper	Produce writing that expresses more than one idea	Independently produce writing with multiple ideas

How to Interpret the Mastery Profile

- Profile reflects a student's mastery of the academic content for that grade level, based on evidence available through the DLM assessment
- Each block (linkage level for one EE) is either mastered or not mastered based on statistics and thresholds

How NOT to Interpret the Mastery Profile

- Doesn't tell # of items answered correctly
- The reasons for not having evidence of mastery are not part of the judgment
 - Blank does not always mean student was tested but answered incorrectly

Recap

- Student responses to items determine if node was mastered
- Node information combined into linkage level mastery
- Linkage levels are the units used in standard setting
- Panelists review profiles that show student's entire LL mastery for the grade/course



61



Questions?



62



TRAINING: PART 3



63



Why Set Standards?

- Assessment is used for accountability purposes
- States need to be able to say whether each student met minimum threshold to be considered proficient
- In DLM states, intended uses may include program and school evaluation, educator effectiveness, etc.



64



How do we know if a student was “proficient”?



65



DLM Performance Levels

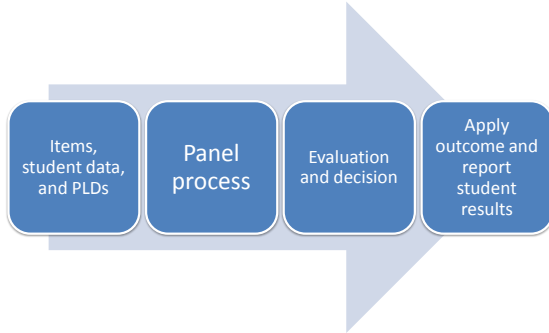
1. The student demonstrates *emerging* understanding of and ability to apply content knowledge and skills represented by the Essential Elements. (EM)
2. The student's understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements is *approaching the target*. (AT)
3. The student's understanding of and ability to apply content knowledge and skills represented by the Essential Elements is *at target*. (T)
4. The student demonstrates *advanced* understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements. (ADV)



66



High Level Overview



DLM Standard Setting

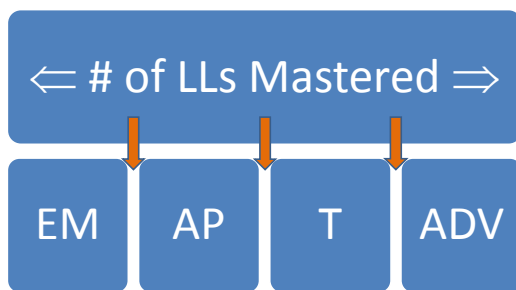
- Panels will set standards by determining where the **cut points** are between the number of linkage levels mastered.
- Anywhere from 40 to 100 linkage levels potentially mastered in one grade/subject



68



Identify Cut Points



Student Profile Method

Draws from other standard setting methods but tailored to DLM

- Iterative process
- Judgments based on patterns of student performance, not item difficulty
 - Panelists do see items though!
- Use mastery profiles
- One set of cut points per grade/course



70



The Essential Question

Which performance level best describes this profile?

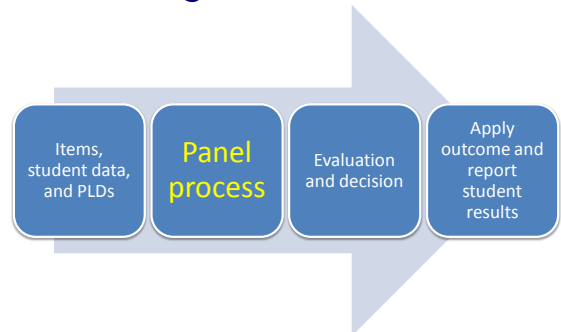
- Use your best professional judgment
- Consider all students with significant cognitive disabilities



71



High Level Overview



Process

1. Training
2. Range finding
3. Pinpointing
4. Draft grade level PLDs
5. Analysis of Impact Data
6. Review of Results and Final Decisions

During the panel process



73



Terms to Know

Cut point: Threshold between two performance levels. LL# that is the **minimum number needed to reach the higher level.**

Range finding: Identifying the **approximate LL#** for each of the cut points between adjacent performance levels

Pinpointing: Honing in on the **specific LL#** that represents the best recommendation of the threshold between two adjacent levels

1. Training

- Advance training: background knowledge
 - Overview of key concepts in DLM
- On-site training: skills needed for the task - **see agenda**
 - Reinforce advance training topics
 - Learn standard setting procedures
 - Specific to grade/content area of panel
 - Includes discussion of PLDs and practice round with materials



75



2. Range Finding

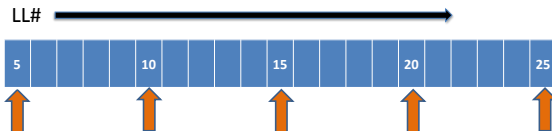
- What is the approximate LL# for each of the cut points between adjacent performance levels?
 - EM/AP
 - AP/T
 - T/ADV



76



Range Finding



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



78



3. Pinpointing

- Honing in on the specific LL# that represents the best recommendation of the threshold between two adjacent levels
 - One cut point at a time



79



Pinpointing



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



81



Pinpointing Procedure

Step 1: Approaching the Target / At Target (AP/T)

Step 2: At Target / Advanced (T/ADV)

Step 3: Emerging / Approaching the Target (EM / AP)



82



Identify Cut Points

← # of LLs Mastered →

EM

AP

T

ADV



83



4. Grade Level PLDs

- Draft descriptions of the skills that are very important for students to have
- Group discussion at the table after pinpointing is done



84



After the Panel Finishes

- State partners and DLM Technical Advisory Committee review impact data
 - If the panel recommendations were adopted, what percent of students would achieve at each level in 2014-15?
 - Review for the whole consortium and by state



85



After the Panel Finishes

- States have processes for internal review and decisions on adoption of consortium-wide outcome
- Outcomes are applied to student performance for 2014-15 for score reporting



86



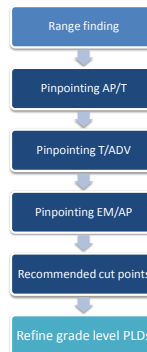
A LITTLE MORE SPECIFIC...



87



Panel Process



- Entire process completed for one grade (or course) before starting the next
 - Most tables will do this three times
- Several steps within each phase

Range Finding

- Orientation to profiles
- Panelists assign performance levels to each profile (round 1 rating)
- Summary information is shared and group discusses
- Panelists make final ratings of each profile (round 2)
- Statistics are used to identify the approximate cut point
 - Where is the greatest disagreement?



89



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



90



Materials

- Profiles across spectrum for number of linkage levels mastered (LL#)
 - Multiple profiles provided at each number mastered
- EE lists and mini-maps
- Node booklets
- Example testlets
- PLDs and other resources



Area	Grade Level Expectation	Level					
		1	2	3	4	5	
Determining Critical Elements of Text	RI.3.1	Answer who and what questions to demonstrate understanding of details in a text	Attend to object characteristics	Identify familiar people, objects, places, and events	Answer who and what questions and identify details in a familiar story	Answer who and what questions about story details	Answer who, what, when, and where questions about story details
	RI.3.2	Analyze details with events in stories from diverse cultures	Seek about objects	Identify familiar people, objects, places, or events	Answer details with events in a familiar story	Associate details with events in diverse stories	Recall diverse stories with key details
	RI.3.3	Identify the feelings of characters in a story	Identify feeling states in self	Identify feeling words	Identify the feelings of characters in familiar stories	Identify the feelings of characters in a story	Identify character feelings and relate to actions
	RI.3.5	Determine the beginning, middle, and end of a familiar story with a logical order	Express interest in book sharing	Differentiate between text and pictures	Identify details and end of a story	Determine the beginning, middle, and end of a familiar story with a logical order	Identify beginning and end of a story
	RI.3.1	Answer who and what questions to demonstrate understanding of details in a text	Attend to object characteristics	Identify familiar people, objects, places, or events	Identify concrete details in an informational text	Answer who and what questions to demonstrate understanding of details in a text	Identify words related to explicit information
Using Writing to Communicate	RI.3.2	Identify details in a text	Seek about objects	Attend to object characteristics	Identify illustrations for familiar text	Identify concrete details in informational text	Identify explicit details in informational texts
	RI.3.5	With guidance and support, use text features including headings and key words to locate information in a text.	Seek about objects	Identify familiar people, objects, places, or events	Identify illustrations that with a familiar text	Use basic text features to find information	Use specific text features to locate information
	RI.3.4	Determine words and phrases that complete literal sentences in a text.	Attend to object characteristics	Understand words for object characteristics	Identify similar or opposite word meanings	Use words to complete a sentence from a story	Identify the meaning of an unknown word using basic context
	EE.3.5.a	Determine the literal meaning of words and phrases in context	Attend to object characteristics	Identify familiar people, objects, places, or events	Identify similar or opposite meaning words	Determine literal meaning of words and phrases	Use words to complete meaningful sentences
	L.3.5.c	Identify words that describe personal experiences.	Identify feeling states in self	Understand and identify common feeling words	Understand and identify feeling words	Identify feeling words for personal state	Draw the internal and external character traits
Using Writing to Communicate	W.3.2.a	Select a topic and write about it including one fact or detail	Seek about objects	Displays interest in objects	Can select a topic from familiar choices	Write about a topic by providing facts and details	Independently select a topic and produces relevant facts and details
	W.3.4	With guidance and support produce writing that expresses more than one idea	Directs attention to objects and people	Displays interest in objects and people	Produces writing that expresses one idea	Produces writing that expresses more than one idea	Independently produce writing with multiple ideas

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
M.C2.2	M.3.G.2	Recognize wholeness, parts/wholes of given unit	Partition shapes into 2, 3, or 4 parts	Model equal part, partition circles, and rectangles	Partition shapes into equal parts	Recognize 1/2, 1/3, 1/4, and 1/10 on an area model
M.C3.1	M.3.MD.1	Pay attention; recognize different	Recognize measurable attributes	Recognize hours and minutes on a digital clock	Tell time to the nearest hour on a digital clock	Tell time to half/quarter hour on a digital clock
M.C3.2	M.3.MD.3	Arrange objects in pairs; recognize attribute values	Classify and order objects by attribute	Recognize and interpret the framework of bar and picture graphs	Use bar and picture graphs to answer questions	Interpret information in bar or picture graphs
M.C3.1	M.3.MD.4	Recognize attribute values	Compare the lengths of two objects	Measure length or distance using informal units	Measure the length of an object in inches or feet	Compare the lengths of two or more objects
M.C1.1	M.3.NBT.2	Recognize a set; recognize separateness	Recognize unit; recognize ten as a group of 10 objects	Compose numbers by recognizing tens and ones	Know place value for ones and tens	Relate rounding and place value; compare numbers
M.C1.1	M.3.NBT.3	Identify before or after	Explain number sequence pattern	Count verbally or using objects to 30	Skip count by 10s	Skip count by 10's using expanded addition and/or money
M.C1.1	M.3.NF.1-3	Recognize "same"	Recognize wholeness and separateness	Divide shapes into distinct parts	Recognize parts of whole/unit; know unit fraction	Recognize fraction, whole, and one-half
M.C4.1	M.3.OA.1-2	Recognize set, subset, and separateness	Combine sets; demonstrate the concept of addition	Represent repeated addition with an equation or model	Solve repeated addition problems	Display understanding of multiplication
M.C1.3	M.3.OA.4	Recognize separate objects and objects in a set	Combine/partition sets; know addition subtraction	Know addition, subtraction, and equal signs	Determine unknown in addition/subtraction equation	Solve addition and subtraction problems with an unknown.
M.C4.1	M.3.OA.8	Combine two or more sets and partition sets	Show concepts of addition and subtraction	Find the unknown in an addition/subtraction equation	Solve addition and subtraction word problems	Solve 2-step addition or subtraction word problems

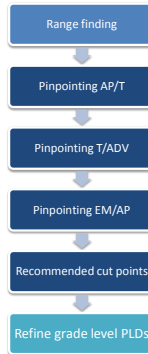
Pinpointing Procedure

- Consider all profiles in each folder
 - Different from range finding folders
- Decide which performance level best describes each profile
 - Independent evaluation (round 1)
 - Group discussion
 - Independent final ratings (round 2)



The Week's Agenda

- Training
- Do the cycle for 2-3 grades/courses
- Standard setting evaluation & wrap-up
- Template review (as time permits)



ROLES AND RESPONSIBILITIES



Panelists

- Bring your expertise to the process
- Make judgments and complete ratings using established procedures and criteria



97



Panelists

- Engage in constructive group discussions
 - Share your opinions and explain your decisions
 - Listen to others and understand their opinions
 - Shared understandings bring some consensus, but you are not “required” to agree with others



98



Panelists

- One of you at each table is a table leader
 - Help ensure materials are properly organized when turned in
 - May assist facilitator with timekeeping, keeping discussion on track



99



Facilitators

- Guide your table through discussions and rating procedures
- Record and present results
- Answer your questions
- They are not expert consultants, co-panelists, etc.



100



Room Staff

- Room leaders
- Room assistants
- Organizers/runners



101



Other Staff

- Psychometricians
 - Ensure data collected meet the needs of the process
 - Prepare for steps that come after this panel
- Leadership
 - Observe
 - Provide feedback to staff
 - Address panelists as needed



102



Observers

- State agency observers
 - Learn about the DLM standard setting process
 - Help the agency interpret the recommendations as part of internal review and decision making process



103



Observers

- Technical Advisory Committee (TAC)
 - Evaluate the process
 - Fidelity, quality
 - Report back to the other TAC members
 - Provide summary statement of evaluation to state partners



104



POLICY PERFORMANCE LEVEL DESCRIPTORS



105



Purpose of Policy-Level PLDs

- Provide clear high-level distinctions between one level of performance and another
- Identify what is “good enough” for mastery at each level
- Used for accountability purposes
 - may include program and school evaluation, educator effectiveness, etc.
- Used to set cut points



106



DLM Performance Levels

1. The student demonstrates **emerging** understanding of and ability to apply content knowledge and skills represented by the Essential Elements. (EM)
2. The student’s understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements is **approaching the target**. (AT)
3. The student’s understanding of and ability to apply content knowledge and skills represented by the Essential Elements is **at target**. (T)
4. The student demonstrates **advanced** understanding of and ability to apply targeted content knowledge and skills represented by the Essential Elements. (ADV)



107



Clear interpretation of PLD descriptors relies on consensus of best professional judgment in your panels



108



Discussion about Policy PLDs

What does this mean?

*The student demonstrates ***** understanding of and ability to apply content knowledge and skills represented by the Essential Elements.*



109



Discussion about Policy PLDs

What do these terms mean?

- Emerging
- Approaching the target
- At target
- Advanced



110



Considerations for Using PLDs

- Refer to any notes you may have taken during discussions of what the PLD descriptors mean.
- Remember: You are determining the number of linkage levels that are necessary to be mastered to reach each performance level.
- It is critical to have a common understanding of the PLDs.
 - You will be seeing profiles with the same number of LLs but with different Essential Elements mastered.



111



Preview of Part 4 Training

In separate rooms:

- Walking through the process with more detail
- Getting to know the materials and specific procedures
- Hands-on activity
- Discussion & debrief



112



BREAK



113



TRAINING: PART 4



114



Purpose

- Familiarize you with materials
- Prepare you for range finding



115



Resource Materials

- Blank profile
- Blueprint
- EE node tables / mini-maps
- Node booklet



116



Blank Profile

Area	Essential Element	Level Mastery				
		Initial Precursor	Distal Precursor	Proximal Precursor	Target	Successor
EA.C.1	EA.RL.3.1	Attend to object characteristics	Identify familiar people, objects, places, or events	Answer who and what questions about details in familiar text	Answer who and what questions about story details	Answer wh- questions about story details
EA.C.1	EA.RL.3.2	Seek absent objects	Identify familiar people, objects, places, or events	Associate details with events in a familiar story	Associate details with events in diverse stories	Recount diverse stories with key details
EA.C.1	EA.RL.3.3	Identify feeling states in text	Identify feeling words	Identify character feelings in a familiar story	Identify character feelings	Relate character feelings to actions
EA.C.2	EA.RL.3.4	Attend to object characteristics	Understand words for absent objects and people	Identify real-world uses of words	Identify words or phrases to complete a literal sentence	Identify the meaning of an unambiguous word
EA.C.1	EA.RL.3.5	Express interest in book sharing	Differentiate between text and pictures	Identify the beginning and end of a familiar story	Identify beginning, middle, and end of a familiar story	Identify the beginning and end of a story
EA.C.3	EA.RL.3.6	Attend to object characteristics	Identify familiar people, objects, places, or events	Associate details and events	Identify similar story elements in two stories	Identify similarities of plots in two stories
EA.C.1	EA.RL.3.1	Attend to object characteristics	Identify familiar people, objects, places, or events	Identify concrete details in an informational text	Understand simple questions about concrete details	Identify words related to explicit information
EA.C.1	EA.RL.3.2	Seek absent objects	Attend to object characteristics	Identify illustrations for a familiar text	Identify a concrete detail in an informational text	Identify explicit details in informational texts
EA.C.1	EA.RL.3.3	Identify a forward sequence in a familiar routine	Identify actions in familiar routines	Identify events in a familiar informational text	Identify which event comes first	Identify temporal information or events
EA.C.2	EA.RL.3.4	Attend to object characteristics	Understand words for absent objects and people	Identify the meaning of words	Identify words or phrases to complete a literal sentence	Identify the meaning of an unambiguous word
EA.C.1	EA.RL.3.5	Seek absent objects	Identify familiar people, objects, places, or events	Identify illustrations that go with a text	Use basic text features to find information	Use specific text features to locate information

Blueprint (IM)

Grade 3: Available Essential Elements and minimum expectation for each student's assessment

Claim	Conceptual Area	EE	Description
1	Students demonstrate increasingly complex understanding of number sense. Choose two EEs from Claim 1 in different conceptual areas, i.e., one EE in C.1.1 and one EE in C.1.3.		
	M.C.1.1	3.NBT.2	Demonstrate understanding of place value to tens.
		3.NBT.3	Count by tens using models such as objects, base ten blocks, or money.
		3.NF.1-3	Differentiate a fractional part from a whole.
	M.C.1.3	3.OA.4	Solve addition and subtraction problems when result is unknown, limited to operands and results within 20.
2	Students demonstrate increasingly complex spatial reasoning and understanding of geometric principles. All students are assessed on the EE in Claim 2.		
	M.C.2.2	3.G.2	Recognize that shapes can be partitioned into equal areas.
3	Students demonstrate increasingly complex understanding of measurement, data, and analytic procedures. Choose two EEs from Claim 3.		
	M.C.3.1	M.C.3.1	Tell time to the hour on a digital clock.
		3.MD.4	Measure length of objects using standard tools, such as rulers, yardsticks, and meter sticks
		M.C.3.2	3.MD.3
4	Students solve increasingly complex mathematical problems, making productive use of algebra and functions. Choose one EE from Claim 4.		
	M.C.4.1	3.OA.1-2	Use repeated addition to find the total number of objects and determine the sum.
		3.OA.8	Solve one-step real-world problems using addition or subtraction within 20.
		M.C.4.2	3.OA.9

Blueprint (YE/EOI)

Grade 3: Essential Elements Assessed

Conceptual Area	EE	DESCRIPTION
EA.C.1.1	EE.RL.3.1	Answer who and what questions to demonstrate understanding of details in a text.
	EE.RL.3.2	Associate details with events in stories from diverse cultures.
	EE.RL.3.3	Identify the feelings of characters in a story.
	EE.RL.3.5	Determine the beginning, middle, and end of a familiar story with a logical order.
	EE.RL.3.1	Answer who and what questions to demonstrate understanding of details in a text.
	EE.RL.3.2	Identify details in a text.
EA.C.1.2	EE.RL.3.3	Order two events from a text as "first" and "next".
	EE.RL.3.4	Determine words and phrases that complete literal sentences in a text.
	EE.RL.3.4	Determine words and phrases that complete literal sentences in a text.
	EE.RL.3.8	Identify two related points the author makes in an informational text.
EA.C.1.3	EE.L.3.5.a	Determine the literal meaning of words and phrases in context.
	EE.L.3.5.c	Identify words that describe personal emotional states.
	EE.RL.3.9	Identify common elements in two stories in a series.
EA.C.2.1	EE.RL.3.9	Identify similarities between two texts on the same topic.
	EE.W.3.2.a	Select a topic and write about it including one fact or detail.
	EE.W.3.4	With guidance and support produce writing that expresses more than one idea.

How Can The Blueprint Help?

Both models:

- Full wording of the EE
- Easy to see breadth of assessment

Integrated model:

- Understand available EEs and constraints on choice, expected coverage



120



EE/Node Table



ESSENTIAL ELEMENT, NODES, AND MINI-MAP

MATH: 3RD GRADE

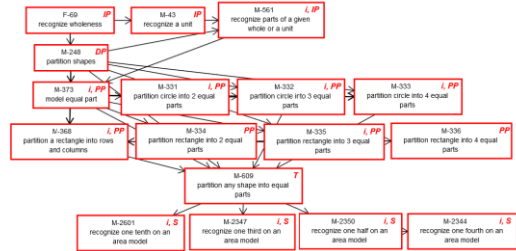
M.EE.3.G.2

Grade-Level Standards	DLM Essential Element	Node Linkage Progression
M. 3.G.2 Partition shapes into parts with equal areas. Express the area of each part as a unit fraction of the whole. For example, partition a shape into 4 parts with equal area, and describe the area of each part as $\frac{1}{4}$ of the area of	M.EE.3.G.2 Recognize that shapes can be partitioned into equal areas.	Successor Node: <ul style="list-style-type: none"> Recognize one tenth on an area model Recognize one third on an area model Recognize one half on an area model Recognize one fourth on an area model Target Nodes: <ul style="list-style-type: none"> Partition any shape into equal parts Proximal Precursor: <ul style="list-style-type: none"> Model equal part Partition circle into 2 equal parts Partition circle into 3 equal parts Partition circle into 4 equal parts

Mini-Maps

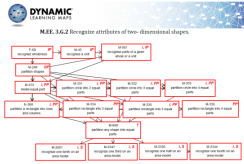


M.EE. 3.G.2 Recognize attributes of two-dimensional shapes.



How Can Node Tables and Mini-Maps Help?

- More fine-grained information than linkage level descriptor
- See relationships among nodes within and across linkage levels



Node Booklet

Node	Node Name	Node Description	Node Observation
F-1	Understands that objects have a function	Comprehends that all objects have some function or action typically associated with it (object action)	The student can demonstrate an understanding of an object's function through interaction with the object in an expected way.
F-2	recognize same	Identify or name same as objects that are identical. For ELA, identify sameness within familiar contexts.	The student is able to demonstrate recognition that two objects are identical.
F-11	Can demonstrate understanding of action words.	Can perform requested actions on objects. ("Kick it. Throw it.")	When presented with familiar objects and given a prompt to complete an action, the student is able to use the object and perform an action related to the object.
F-22	Can detect specific members within a broader category.	When applied with a member of a category, can determine if the member belongs in the category.	When presented with two different objects, the student is able to determine which of the two objects a third object is related to.
F-23	Can make judgments about the meaning (of words)	Can determine when two words have the same, similar, or different meanings or whether meanings (of a single word) are the same or different.	The student is able to determine whether two words are similar or different.

How Can The Node Booklet Help?

- Deeper understanding of the behavior that is assessed (items are aligned to nodes)

Node Name	Node Description	Node Observation
F-1	Understands that objects have a function	Comprehends that all objects have some function or action typically associated with it (object action)
F-2	recognize same	Identify or name same as objects that are identical. For ELA, identify sameness within familiar contexts.
F-11	Can demonstrate understanding of action words.	Can perform requested actions on objects. ("Kick it. Throw it.")
F-22	Can detect specific members within a broader category.	When applied with a member of a category, can determine if the member belongs in the category.
F-23	Can make judgments about the meaning (of words)	Can determine when two words have the same, similar, or different meanings or whether meanings (of a single word) are the same or different.

Suggested Process

1. Analyze LL descriptors
2. Review EEs/mini-maps to understand relationships and find the nodes
3. Go to node booklet to understand nodes better
4. Identify LLs where you would like to see testlets

Ready to Rate?

After reviewing materials:

Can you imagine the kind of assessment items that would measure what is at each linkage level and how a student would have to respond in order to demonstrate mastery of that linkage level?



127



PRACTICE ACTIVITY



128



Activity

- Mock range finding
- Extra discussion on the meaning of PLDs
- Extra discussion on the rating process



129



2. Range Finding

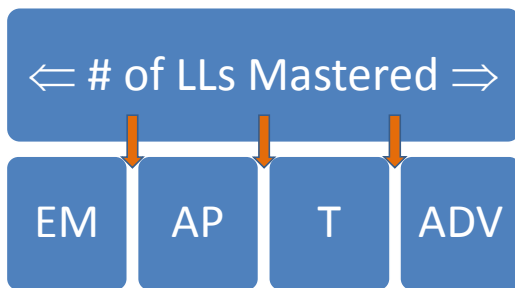
- What is the approximate LL# for each of the cut points between adjacent linkage levels?
 - EM/AP
 - AP/T
 - T/ADV



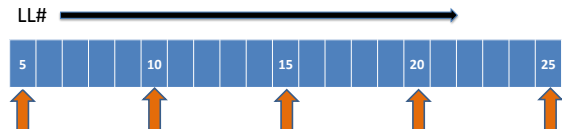
130



Identify Cut Points



Range Finding



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



133



Training Activity

1. Orientation to blank profile and resources
 - Review to level of comfort
2. Discuss policy PLDs in light of what you have seen
3. Independent evaluation (mock range finding)
4. Discuss the rating process
5. Whole room debrief
6. See the way ratings will be handled



134



Training Folder

Six profiles total:

- Three LL#s, in order
 - 10, 20, 30
- Two profiles at each LL#
- From your first assigned grade/course
 - Use that blank profile



135



Range Finding Rating Form

DLM Standard Setting
Rating Form – Range Finding

Panelist ID: _____ Table ID: _____ Room: IM YE Subject: ELA Math Grade/Course: _____

Profile ID	# Lls	First Rating				Final Rating			
		EM	AP	T	ADV	EM	AP	T	ADV
1									
2									
3									
4									
5									

Only one "X" per row
Must choose only one level – no sitting on the fence!



136



Let's Start!

- Your facilitator will guide you through the steps



137



Debrief



138



Day Two



Topics

- Pinpointing
- Drafting grade level PLDs
- Last step: explaining your rationale



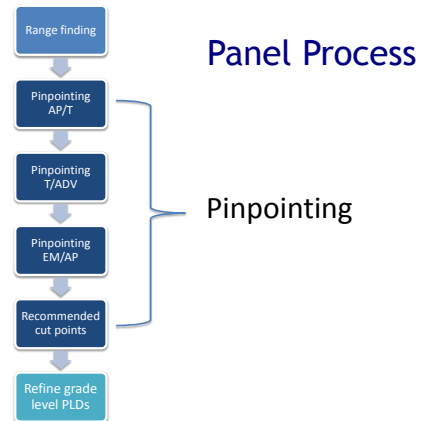
140



PINPOINTING



141



Pinpointing

- Honing in on the specific LL# that represents the best recommendation of the threshold between two adjacent levels
- Starts with range finding results



143



Pinpointing



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



145



Pinpointing Procedure

Step 1: Approaching the Target / At Target (AP/T)

Step 2: At Target / Advanced (T/ADV)

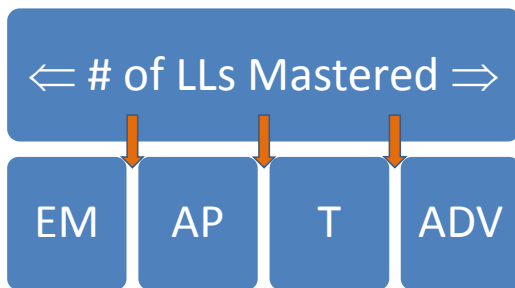
Step 3: Emerging / Approaching the Target (EM / AP)



146



Identify Cut Points



Pinpointing Folder

- Seven LL#, in order
- Three profiles at each LL#
 - Will likely include repeats at one LL from range finding



148



Getting Ready

- Review resource materials so you are ready to rate
 - EEs, mini-maps, nodes
 - Testlets
 - PLDs (and your notes about them)
 - Panelist hints (and your own notes)



149



Ready to Rate?

After reviewing materials:

Can you imagine the kind of assessment items that would measure what is at each linkage level and how a student would have to respond in order to demonstrate mastery of that linkage level?



150



Pinpointing Rating Form

DLM Standard Setting
Pinpointing Form: AP/T

Panelist ID: _____ Table ID: _____ Room: IM YE Subject: ELA Math Grade/Course: _____

Profile ID	# Lls	First Rating				Final Rating			
		EM	AP	T	ADV	EM	AP	T	ADV
1									
2									
3									
4									
5									

151



Pinpointing Procedure

- Consider each profile in folder
- Decide which performance level best describes each profile
 - Independent evaluation (round 1)
 - Group discussion
 - Independent final ratings (round 2)

152



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities

153



Finer Distinctions

- You should really feel confident that a profile shows the **minimum amount of mastery needed to be at the upper performance level**. Otherwise, put it at the lower level.
- Hint: Note in your blank profile the LLs you think are very important for a student to be at a certain level

154



One New Step: During Group Discussion

- Write each LL descriptor that is very important and that exemplifies that performance level on a sticky note
 - EE code and level OR
 - Short description
- Facilitator will add to chart paper

155



SUBJECT/GRADE-SPECIFIC PERFORMANCE LEVEL DESCRIPTORS

156



Policy-Level Performance Level Descriptors

- Are used to set cut points
- Are used for accountability purposes

The student demonstrates <i>emerging</i> understanding of an ability to apply content knowledge and skills represented by the EEs.
The student's understanding of and ability to apply content knowledge and skills represented by the EEs is <i>approaching the target</i> .
The student's understanding of and ability to apply content knowledge and skills represented by the EEs is <i>at target</i> .
The student demonstrates <i>advanced</i> understanding of and ability to apply targeted content knowledge and skills represented by the EEs.



157



Purpose of Subject/Grade-Specific Performance Level Descriptors (PLDs)

- To communicate to educators and parents about the *types of performances that are necessary for students to make progress toward and beyond grade-level expectations.*
- To provide useful information to educators and IEP teams for *curricular and instructional planning.*



158



What are Subject/Grade-Specific PLDs?

- They are *high level summaries* of the nodes/knowledge, skills, and abilities (KSAs) mastered in the range of profiles identified for each level of proficiency.
- They classify student performances into categories that describe the KSAs that students in the category are *typically* able to demonstrate.



159



Format of Subject/Grade-Specific PLD

1. A few sentences that describe the standards addressed at the subject/grade level (from blueprint)
2. A bulleted list of KSAs that are typically mastered at each performance level (from pinpointing)



160



PLD Example: Grade 7 ELA At Target

1. Summary of the claims/conceptual areas that are the focus of the subject/grade

Students performing at the *At Target* level demonstrate basic knowledge and skills in vocabulary and concept development; comprehension skills and response to text; writing as a process; and writing forms, audiences, and purposes with a moderate level of performance.



161



PLD Example (cont.)

2. List of important skills that students typically demonstrate in this category

- Identify the meaning of words, given choices
- Identify contextual clues for word meaning
- Locate words in a dictionary
- Answer questions about text, such as drawing conclusions or identifying evidence to support given conclusions
- Sequence multi-step directions
- Match traits to types of literature
- Generate details about a topic
- Write a topic sentence
- Edit and revise sentences



162



Process for developing S/G-PLDs

- Lay groundwork during pinpointing when classifying profiles and identifying important skills in each category
- Finalize after all recommended cut points are identified through pinpointing
- Will do for each grade/course



163



THE FINAL STEP



164



Capture Your Rationale

Purpose of this activity:

- Understand the general rationale that goes along with your final judgments and the draft grade-level PLDs
- Allows us to share with states and TAC your reasoning - helps them interpret the results



165



Capturing Your Rationale

- Blank, unlined paper
- Write your name, panelist ID, and the grade/course you are rating
- Explain why you placed the cut points where they are and not somewhere else
 - Describe your reasons
- Explain for state education agency audience



166



Day Three



Topics

- Update on where we are
- A few reminders about the purposes and processes of rating
- Refining our procedures



168



Update: Timeline

- Learning curve is leveling off
 - A few things left to address at tables
- Need to adjust pace for the last two days
 - Finish grade/course #2 by end of day Wednesday
 - Strategies



169



Reminders about Ratings

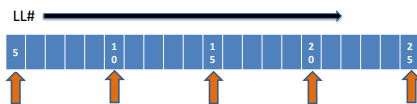
- Relationship between range finding and pinpointing
- Essential question and things to consider when rating



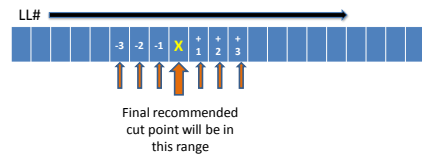
170



Range Finding



Pinpointing



The Essential Question

Which performance level best describes this profile?

- Use your best professional judgment
- Consider all students with significant cognitive disabilities



173



Pitfalls

- Profiles are examples from the universe
 - And the goal is to classify the universe
 - Profiles with the same # of LLs should be classified in the same category in your round 2 ratings
- Rating without anchors
- Data don't reflect what the process should yield



174



Expected Data Pattern - Rangefinding

Profile ID	# of LLs Mastered	Group Frequencies of Round 2 Final Ratings				Total
		Emerging	Approaching	Target	Advanced	
13	5	4				4
14	5	4				4
28	10	4				4
29	10	3	1			4
43	15	2	2			4
44	15	2	2			4
58	20		4			4
59	20		3	1		4
73	25		2	2		4
74	25		1	3		4
88	30		1	2	1	4
89	30		1	2	1	4
103	35			2	2	4
104	35			2	2	4

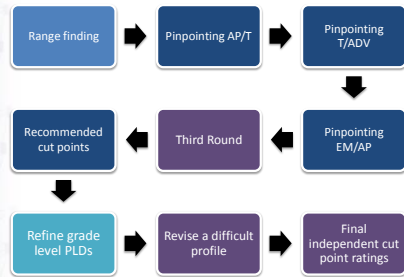
Expected Data Pattern - Pinpointing

Step	Profile ID	# of LLs Mastered	Group Frequencies of Round 2 Final Ratings			Total
			Emerging	Approaching	Target	
STEP 1	61	21		4		4
STEP 1	62	21		4		4
STEP 1	63	21		4		4
STEP 1	64	22		3	1	4
STEP 1	65	22		3	1	4
STEP 1	66	22		3	1	4
STEP 1	67	23		3	1	4
STEP 1	68	23		3	1	4
STEP 1	69	23		4		4
STEP 1	70	24		3	1	4
STEP 1	71	24		2	2	4
STEP 1	72	24		3	1	4
STEP 1	73	25		2	2	4
STEP 1	74	25		3	1	4
STEP 1	75	25		2	2	4
STEP 1	76	26		1	3	4
STEP 1	77	26		1	3	4
STEP 1	78	26		1	3	4
STEP 1	79	27		1	3	4
STEP 1	80	27		1	3	4
STEP 1	81	27		1	3	4

REFINING PROCEDURES



177



178



Third Pinpointing Round

- Do only when absolutely necessary
- As required by psychometrician
- May be for one or more cut points
- Typically samples at 4 LLs instead of 7



179



Revise a Difficult Profile

- Table selects one pinpointing profile that was at lower of two levels and that was hard to rate
 - Odd pattern of mastery?
- As a group, edit the profile to show what it would take to get it to the next level
- Turn in edited profile & short explanation



180



Final Independent Ratings

- Review table results for all three cut points
- No additional discussion
- Each panelist:
 - Indicates comfort with the number
 - Provides a final rating of what the number should be



181



A Small Revision

Grade level PLDs:

- Facilitator will take notes & project
- Most work happens during pinpointing discussion
- At end of all three pinpointing steps, just take things off the fence

Rationale:

- Delete this activity



182





DYNAMIC™
LEARNING MAPS

Dynamic Learning Maps Standard Setting Event

June 15 – 18, 2015

*Hilton Kansas City Airport
8801 NW 112th Street
Kansas City, MO 64153*

AGENDA

Sunday, June 14, 2015	
Throughout the day	Out-of-town attendees arrive in Kansas City <i>*If arriving by air, please use the courtesy phone in the baggage claim area to summon the Hilton's 24-hour complimentary shuttle.</i>

Monday, June 15, 2015 Shawnee Ballroom	
8:00 – 8:30 a.m.	Registration
8:30 – 8:40 a.m.	Welcome and introductions
8:40 – 9:15 a.m.	Training Part 1 <ul style="list-style-type: none"> • Security and confidentiality • Refresher on DLM system design
9:15 – 9:45 a.m.	Training Part 2: Student Results and Mastery Profiles
9:45 – 10:15 a.m.	Training Part 3: Overview of Standard Setting <ul style="list-style-type: none"> • Purpose • Processes • Roles and responsibilities • Performance level descriptors (PLDs)
10:15 – 10:30 a.m.	<i>Break</i>
10:30 – 12:00 p.m.	Training Part 4: Standard Setting Procedures <ul style="list-style-type: none"> • Materials • Procedures • Practice activity • Discussion
12:00 – 12:45 p.m.	<i>Lunch</i>
12:45 – 1:15 p.m.	Training Debrief
1:15 – 3:00 p.m.	Range finding #1 <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete their independent round 1 ratings
3:00 – 3:15 p.m.	<i>Break</i>

3:15 – 4:30 p.m.	Range finding #1 (cont.) <ul style="list-style-type: none"> • Tables discuss round 1 ratings • Panelists complete their independent round 2 ratings • Final range finding results compiled and shared with table
4:30 – 4:45 p.m.	Turn in materials
4:45 p.m.	<i>Adjourn</i>

Tuesday, June 16, 2015 Shawnee Ballroom	
8:00 – 8:15 a.m.	Check-in
8:15 – 8:30 a.m.	Announcements and instructions
8:30 – 8:50 a.m.	Refresher training on pinpointing procedure
8:50 – 10:00 a.m.	Pinpointing #1: AP/T <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings
10:00 – 10:15 a.m.	<i>Break</i>
10:15 – 11:00 a.m.	Pinpointing #1: AP/T (cont.) <ul style="list-style-type: none"> • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for AP/T cut point
11:00 – 12:00 p.m.	Pinpointing #1: T/ADV <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for T/ADV cut point
12:00 – 12:45 p.m.	<i>Lunch</i>
12:45 – 1:00 p.m.	Announcements
1:00 – 1:30 p.m.	Pinpointing #1: T/ADV (cont.)
1:30 – 2:30 p.m.	Pinpointing #1: EM/AP <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for EM/AP cut point
2:30 – 3:00 p.m.	Grade/Course PLDs <ul style="list-style-type: none"> • Tables review draft lists of PLDs for each performance level and make adjustments, as desired
3:00 – 3:15 p.m.	<i>Break</i>
3:15 – 4:40 p.m.	Range finding #2 (grade/course #2) <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings

	<ul style="list-style-type: none"> Panelists complete their independent round 1 ratings
4:40 – 4:45 p.m.	Turn in materials
4:45 p.m.	<i>Adjourn</i>

Wednesday, June 17, 2015 Shawnee Ballroom	
8:00 – 8:15 a.m.	Check-in
8:15 – 8:30 a.m.	Announcements and instructions
8:30 – 9:30 a.m.	Range finding #2 (cont.) <ul style="list-style-type: none"> Tables discuss round 1 ratings Panelists complete their independent round 2 ratings Final range finding results compiled and shared with table
9:30 – 10:15 a.m.	Pinpointing #2: AP/T <ul style="list-style-type: none"> Tables review materials and prepare for independent ratings Panelists complete independent round 1 ratings Tables review round 1 ratings and discuss as a group Panelists identify key skills that emerged during discussion Panelists independently complete round 2 ratings Final ratings submitted for that cut point
10:15 – 10:30 a.m.	<i>Break</i>
10:30 – 11:00 a.m.	Pinpointing #2: AP/T (cont.)
11:00 – 12:00 p.m.	Pinpointing #2: T/ADV <ul style="list-style-type: none"> Tables review materials and prepare for independent ratings Panelists complete independent round 1 ratings Tables review round 1 ratings and discuss as a group Panelists identify key skills that emerged during discussion Panelists independently complete round 2 ratings Final ratings submitted for that cut point
12:00 – 12:45 p.m.	<i>Lunch</i>
12:45 – 1:00 p.m.	Announcements
1:00 – 2:00 p.m.	Pinpointing #2: EM/AP <ul style="list-style-type: none"> Tables review materials and prepare for independent ratings Panelists complete independent round 1 ratings Tables review round 1 ratings and discuss as a group Panelists identify key skills that emerged during discussion Panelists independently complete round 2 ratings Final ratings submitted for that cut point
2:00 – 3:00 p.m.	Grade/Course PLDs <ul style="list-style-type: none"> Tables review draft lists of PLDs for each performance level and make adjustments, as desired
3:00 – 3:15 p.m.	<i>Break</i>
3:15 – 4:40 p.m.	Range finding #3 (grade/course #3) <ul style="list-style-type: none"> Tables review materials and prepare for independent ratings

	<ul style="list-style-type: none"> • Panelists complete their independent round 1 ratings • Tables discuss round 1 ratings • Panelists complete their independent round 2 ratings • Final range finding results compiled and shared with table
4:40 – 4:45 p.m.	Turn in materials
4:45 p.m.	<i>Adjourn</i>

Thursday, June 18, 2015 Shawnee Ballroom	
8:00 – 8:15 a.m.	Check-in
8:15 – 8:30 a.m.	Announcements and instructions
8:30 – 10:00 a.m.	Pinpointing #3: AP/T <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for that cut point
10:00 – 10:15 a.m.	<i>Break</i>
10:15 – 11:15 a.m.	Pinpointing #3: T/ADV <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for that cut point
11:15 – 12:00 p.m.	Pinpointing #3: EM/AP <ul style="list-style-type: none"> • Tables review materials and prepare for independent ratings • Panelists complete independent round 1 ratings • Tables review round 1 ratings and discuss as a group • Panelists identify key skills that emerged during discussion • Panelists independently complete round 2 ratings • Final ratings submitted for that cut point
12:00 – 12:45 p.m.	<i>Lunch</i>
12:45 – 1:00 p.m.	Announcements
1:00 – 1:45 p.m.	Grade/Course PLDs <ul style="list-style-type: none"> • Tables review draft lists of PLDs for each performance level and make adjustments, as desired
1:45 – 2:15 p.m.	Standard Setting Wrap-up <ul style="list-style-type: none"> • Panelists complete evaluations of standard setting process • Final instructions and reminders • Turn in materials
2:15 – 3:30 p.m.	Template review activity
3:30 p.m.	<i>Adjourn – Attendees Depart</i>

Standard Setting – Detailed Procedures

RANGE FINDING

An overview of each step is providing in the blue boxes. The text that follows provides greater detail for steps to follow.

<p>1. Orientation to blank profile: Panelists review profiles and use mini-maps, node tables, and node booklets Panelists request to view sample testlets (per EE/linkage level)</p>	<p>Materials: Resource notebook Blank profile (in notebook) Mini-maps and node tables <i>Facilitators:</i> <i>List of CB IDs for testlets to preview</i> <i>Cheat sheet of repeating nodes</i></p>	<p>Reminders: Make sure panelists are familiar with the materials and that they feel comfortable asking to see more testlets, ask more questions, etc.</p>
--	--	--

1. Remind panelists that the goal at this stage is to **become familiar with and comfortable with the linkage level descriptors**. We want every panelist to be able to imagine the **kind of assessment items** that would measure what is at this linkage level, and **how a student would have to respond** in order to answer correctly.
2. Remind them about the resources available to help them become comfortable with the linkage levels described in their current grade/subject profile:
 - a. **Mini-maps** show the specific nodes within each linkage level and identify the node by number
 - b. **Node booklet** provides node descriptions and observations
 - c. Use the **facilitator cheat sheet of repeating LLs** to help panelists identify identical wording in more than one place.
3. Remind them about strategies for reviewing the **blank profile**.
 - a. At this stage, they only have to review this one profile. They are looking at the text in the boxes, since there is no shading yet.
 - b. If they want to pull apart the stapled profile, they may do that.
 - c. They may want to highlight or underline key ideas, or write notes to themselves in the margin to help them remember their interpretations
 - d. If there are any EEs/LLs where they want to see a sample testlet, they should make a list of those.

4. After everyone has had time to review the blank profile:
 - a. **ASK: Which linkage levels do you have questions about? Which would you like to discuss?**
 - i. Facilitate discussion within the panel about their interpretations of the LLs. If there are different opinions of what a LL means, suggest that they add it to the list of testlets they would like to view.
 - b. Working down the list of EEs, ask panelists to identify the LLs they would like to see testlets for.
 - i. Use the **CB ID spreadsheet** to look up the numbers. Mark the CB ID on facilitator’s master copy of the blank profile. If there is more than one testlet available, make sure you choose the general form (not BVI, not braille).
 - ii. Open CB and go to testlet preview to show the testlets. Answer questions about the testlets. (Remember that TE items may not preview correctly in CB.)
 - c. If there are lingering questions about the LLs that you cannot answer, ask the runner to locate the room lead so that person can come to the table and assist or identify a content team member to come help.
 - d. Confirm that panelists feel they are familiar with and comfortable with the linkage level descriptors.
 - i. **ASK: Can you imagine the kind of assessment items that would measure what is at each linkage level, and how a student would have to respond in order to demonstrate mastery of that linkage level?**

<p>2. Independent evaluation: Panelists receive their range finding folder. Facilitator reviews (re-explains) the rating task. Panelists independently evaluate the profiles in the range finding folder and identify the performance level that describes each profile. They record the decision for each exemplar profile on the rating sheet.</p>	<p>Materials: Same as previous step PLUS: Range finding folder (1 per panelist) Range finding rating sheet</p>	<p>Recommendations: Use paraphrase of recommended language to introduce the task Make sure panelists understand the task (what ratings they are making, and how) Monitor for: independent work, understanding of how to record choices</p>
---	---	---

Preparation

1. Pass out the range **finding folders** and **rating forms**.
2. Instruct panelists to write their panelist ID number on the upper right corner of their **range finding folder** (as it is oriented before they would open it to read the contents). Their panelist ID number is their table number plus a unique number (e.g., 14-1).
3. Point out the ordering of profiles in their folders and how to identify them (by LL# and profile ID).
4. Have panelists fill in the last four digits of the portfolio IDs and the LL# on their **range finding rating form**. This should be done in the same order as the portfolios appear in their folder. (Note: their order should also match what is in the **facilitator workbook**.)
5. **INTEGRATED MODEL ONLY: Look up the number of EEs/linkage levels for the grade and subject on the Maximum number of Essential Elements and Linkage Levels table. Remind the panelists how many Essential Elements are expected to be covered in that grade. Refer them to the blueprint for more information.**
6. Introduce the activity.
 - a. Explain the purpose of the rating task: to label each profile with the performance level descriptor that best describes it.
 - b. **SAY the key question to consider: *Using your best professional judgment and considering all students with significant cognitive disabilities, which performance level best describes this profile?***
 - c. Reminders to the panelists:
 - i. These profiles are **examples** of how students might master the number of linkage levels.
 - ii. You'll rate the **whole profile**, not separately evaluate the EEs or linkage levels
 - iii. Refer to the PLD handout and your notes to help guide your ratings.
7. Remind them how to complete the range finding rating form:
 - a. Only fill in the "first rating" part of the form on this round.
 - b. Only mark one X per row.

- c. Make sure to mark the row that corresponds with the profile you are looking at.
 - 8. Other reminders:
 - a. They are welcome to go through the profiles in whatever way they wish.
 - b. They can take profiles apart if they need to.
 - c. They can change their minds about any profile and adjust a rating any time up until the round is finished.
 - d. **There should be no discussion during this phase.**
 - 9. When they are finished, they should check to make sure they have only one X on every row and that each X is clearly marked (i.e., if they changed their mind while working, their final answer is still clear). They should also make sure no rows are blank.
- ➔ As they are working, monitor for:
- independent work
 - understanding of how to record choices (glance at rater sheets to make sure there is just one rating per row and all are in the first rating section)
- ➔ When panel work is going as intended, you may skip ahead to open the [facilitator workbook](#) and follow the “set-up” step in “Round 1 results”.

<p>3. Round 1 results: Facilitator polls the panel via a raise of hands to find the number of panelists who rated each profile in each of the performance levels. The facilitator records the recommendations in the facilitator workbook.</p>	<p>Materials: Same as previous step PLUS: Facilitator workbook (projected)</p>	<p>Recommendations: Ratings must be entered with 100% accuracy.</p>
--	--	--

Set-Up

1. Open the [facilitator workbook](#). Follow the reminder to “save as” with your initials at the end. Save the new file in the same dropbox folder.
2. Enter the number of panelists in the prompt box when it appears.

Recording

3. Identify two panelists who will watch the hand counts and what is recorded in the facilitator workbook.
4. Mention each profile by ID number. Ask how many panelists rated it as:
 - a. emerging,
 - b. approaching target,
 - c. at target, or
 - d. advanced

For example: “Raise your hand if you categorized profile 1009 as emerging.”

5. Count the number of hands raised in each category for every profile and record in the corresponding cell of the worksheet.
 - a. These entries go in columns D-G (Round 1 ratings).
 - b. If nobody raises their hand for a particular category, leave the cell blank.
 - c. The Total column will remain highlighted in red until it equals the number of panelists entered in Step 1 by facilitator.
6. Ask watchers to confirm that the numbers entered for each performance level match the number of hands that were raised.

<p>4. Table discussion of round 1 results: A summary of the table's initial ratings is projected for the group to review. Facilitator leads discussion of the original ratings.</p>	<p>Materials: Same as previous step</p>	<p>Recommendations: Avoid making judgments about the ratings they chose. Use neutral language. Ask open-ended questions that allow panelists to explain their thought process. Attend to group dynamics. Discussion should focus on what exemplifies a performance level and differentiates it from another. Make sure every panelist has a chance to speak.</p>
---	---	--

1. Remind panelists of the purpose of this discussion: to explore the thinking behind their first round ratings.
 1. They will have a chance to share their reasons and to hear about other panelists' reasons for their ratings.
 2. Remind them there are no right or wrong answers.
 3. Eventually, the goal is for areas where the group may be split between two performance levels to be focused around just one LL#.
2. Ensure that the panelists understand the definitions of the PLDs and how the profiles are examples of ways a student might perform at that total LL level.
 1. They should review any notes they made on their blank profiles as needed.
3. Instruct them **NOT to erase or mark out any of their first round ratings during this step.**
 1. They should not mark on their rating sheets at all. If they want to make notes to themselves, they should use their notebook or blank profile.
 2. There will be a second round later when they will be able to change their minds.

Discussion Hints for This Step

Discussion should focus on what **exemplifies a performance level** and what **differentiates one profile from another**.

- How does this linkage level exemplify the performance level you gave it?
- Are there any profiles where you were really on the fence at first? How did you come to your final decision?
- How did this profile change your understanding of the performance level?
- Here's a place where the ratings were evenly split [Or, where two profiles with the same LL# got different results]. Let's take a closer look.
- How did you approach the process of rating these profiles?

4. Use the frequency distributions (columns D-G) and the results under the box titled "Do we need to discuss before Round 2 Final ratings?" (columns K-N) to guide group discussion. See the "Discussion Hints for This Step" below.
 1. Discuss as a group any profile where the frequency distribution (columns D-G):
 - i. Has counts that cover three of the four categories
 - ii. Shows a split panel (nearly half in each of two groups)
 - iii. Has a different general pattern than the other profile with the same LL#.
 2. Discuss as a group any profile associated with a green "YES" in columns K-N titled "Do we need to discuss before Round 2 Final ratings?". Remind the panelists this is just a general flag to help the panel see where the group seems to have some disagreement.
 3. Discuss as a group any other profile the panelists wish to discuss.
5. By the end of the discussion, some panelists may be leaning toward changing their ratings and others may not. While unanimous agreement is not a requirement, you should hear some evidence that raters are considering changing their ratings.
 1. Remind them that the goal at this stage is to hone in on one LL# where panelists disagree about which category it belongs in (for each of the three cuts/four categories).
 2. Do **not** ask them to vote at this stage. Just monitor conversation and especially listen for convergence (i.e., panelists have shared their opinions and others have heard them and to some extent agree with the opinion). They will have a chance to make their ratings again during the final round.
 3. If you have concerns about the discussion at your table or if questions are raised that you are not comfortable answering, ask your runner to bring the room lead to the table. If discussion is leading to greater likelihood of convergence, wrap up when everyone has had a chance to discuss as they like.

<p>5. Panelists review/revise ratings: Panelists are given the opportunity to adjust their independent ratings. They enter their final (round 2) ratings for all profiles.</p>	<p>Materials: Return to range finding folder Range finding rating sheet</p>	<p>Recommendations: Make sure panelists understand the task (what ratings they are making, and how) Monitor for: independent work, understanding of how to record choices</p>
--	---	---

1. Explain the purpose of the rating task: to label each profile with the performance level descriptor that best describes it.
 - a. **SAY the key question to consider: *Using your best professional judgment and considering all students with significant cognitive disabilities, which performance level best describes this profile?***
 - b. Remember that these profiles are examples of how students might master the number of linkage levels.
 - c. Consider the discussion we just had, but remember that you do not have to reach a unanimous decision. You should still complete your ratings based on your own professional judgment.
 - d. Use the PLDs to guide your decisions.
 2. Remind them how to complete the range finding rating form:
 - a. Only fill in the “final rating” part of the form on this round.
 - b. Only mark one X per row.
 - c. Provide the rating for every single row – even if they are not changing their rating from round one.
 - d. Make sure they are marking the row that corresponds with the profile they are looking at.
 3. Tell panelists they are welcome to go through the profiles in whatever way they wish.
 - a. They can change their minds about any profile and adjust a rating any time up until the round is finished.
 - b. **They should not be discussing anything during this phase.**
 4. When they are finished, they should check to make sure they have only one X on every row and have not skipped any rows.
 5. As they finish, ask each panelist to reassemble any profiles they have taken apart and return the profiles to their **range finding folder**. They may keep their **blank profile** with their notes on it.
- ➔ As they are working, monitor for:
- independent work
 - understanding of how to record choices (glance at rater sheets to make sure there is just one rating per row and all are in the first rating column)

<p>6. Round 2 ratings submitted: Panelists show their votes via a raise of hands, like in previous round. Round 2 ratings are entered into the excel sheet by the facilitator and shared with the group. Runner notifies room assistant that final sheet is ready for psychometrician check.</p>	<p>Materials: Projected: Facilitator workbook</p>	<p>Recommendations: Ratings must be entered with 100% accuracy.</p>
--	---	--

Set-Up

1. Return to the [facilitator workbook](#).

Recording

2. Identify two panelists who will watch the hand counts and what is recorded in the facilitator workbook.
3. Mention each profile by ID number. Ask how many panelists rated it as:
 - a. emerging,
 - b. approaching target,
 - c. at target, or
 - d. advanced

For example: “Raise your hand if you categorized profile 1009 as emerging.”

4. Count the number of hands raised in each category for every profile and record in the corresponding cell of the worksheet.
 - a. Record these frequencies in columns S-V (Round 2 final ratings).
 - b. The Total column will remain highlighted in red until it equals the number of panelists entered previously by facilitator.
 - c. **They should not mark on their forms at all during this phase.**
5. Ask watchers to confirm that the numbers entered for each performance level column match the number of hands that were raised.
6. Ask runner to notify psychometrician that table is ready for review.
[Note: This step may not be needed after the first grade is completed by the panel.]

<p>7. Facilitator finalizes range finding results: After psychometrician review, the facilitator uses the macro to populate pinpointing values. Facilitator fills out pinpointing materials form and gives to runner.</p>	<p>Materials: Facilitator workbook Pinpointing materials form</p>	<p>Recommendations: N/A</p>
---	---	-----------------------------

1. Psychometrician reviews data in the spreadsheet and confirms facilitator is ready to run the macro. [This step may not be necessary after the first time you use the workbook. The psychometrician will tell you if you need to continue to ask for reviews.]
 2. Facilitator completes Pinpointing Materials Form and gives it to **runner (see note below)**.
 3. Display the pinpointing results. Out loud for the group, summarize the cut point ranges using the orange cells in columns AG – AI that will be analyzed during the pinpointing exercise.
 - a. “Based on logistic regression results, the cut point between approaching target and target is somewhere between ___ and ___.”
 - b. “Based on logistic regression results, the cut point between target and advanced is somewhere between ___ and ___.”
 - c. “Based on logistic regression results, the cut point between emerging and approaching target is somewhere between ___ and ___.”
 4. Tell the group that based on these results, they will find the final recommended cut point within these ranges.
 5. Facilitator uses the macro to populate pinpointing values.
 - a. Click on the button titled “Get Pinpointing Results”
 - i. Once you complete this step, you CANNOT change the Range finding results!
 - ii. Type “Yes” if you are certain you are ready to move onto Pinpointing.
- ➔ If the end of this step is not timed well with a break built into the agenda, give panelists a 15-minute break.
- ➔ If there is additional wait time for pinpointing files to be prepared, have panelists start on the review of materials/resources and their blank profile with their first cut point in mind. Start thinking about what it would take to reach the “at target” level.
6. **Runner** coordinates with room assistant to build pinpointing files:
 - a. Prepare, check, and deliver folders for AP/T pinpointing.
 - b. Prepare and check folders for T/A and EM/AP pinpointings. Set aside and label these groups in temporary holding space.

PINPOINTING

This process is completed in its entirety and in the following order for each of the following cut points:

1. Approaching Target and At Target (AP/T) → **STEP 1**
2. At Target and Advanced (T/ADV) → **STEP 2**
3. Emerging and Approaching Target (EM/AP) → **STEP 3**

<p>8. Materials distribution and preparation: Folders are distributed to the panel containing the profiles for the seven levels including and around the value obtained during range finding</p>	<p>Materials: Pinpointing folder for target (1 folder, per panelist) Resource book Pinpointing rating form for that step Chart paper for that step</p>	<p>Recommendations: Make sure panelists are familiar with the materials</p>
<p>9. Re-familiarize with profiles (as needed): Panelists review profiles as needed and use mini-maps, node tables, and node booklets to reach their desired level of familiarity with the linkage level descriptors. Panelists may request to view additional sample testlets (per EE/linkage level)</p>	<p>Materials: Same as previous step</p>	<p>Recommendations: Make sure panelists feel comfortable asking to see more testlets, ask more questions, etc.</p>

1. **Runner** delivers **pinpointing folders** and hangs 2 sheets of **chart paper** with dividing line between the four levels.
2. Instruct panelists to write their panelist ID number in the upper right corner of their folder (as it is oriented before they would open it to read the contents).
3. Instruct panelists to fill in the information at the top of their **Pinpointing Form**.
4. Instruct panelists to go through the profiles in their folder, **in the order in which they were provided**. On their rating form, write in:
 - a. Last four digits of profile ID
 - b. Number of linkage levels mastered for that profile
5. Confirm that all panelists have data filled in for as many rows as there should be (3 per LL# x 7 profiles each = 21 rows).

6. Before beginning the pinpointing phase, ask panelists if there are any more LLs on which they need discussion or time to refresh their memory using the materials in their resource books.
7. Tell panelists this is their chance to re-familiarize themselves with the contents of the linkage levels for this grade/subject. During the pinpointing process, we still want every panelist to be able to imagine the kind of assessment items that would measure what is at this linkage level, and how a student would have to respond in order to answer correctly. They may wish to:
 - a. Review mini-maps
 - b. Review node booklets
 - c. View testlets
 - d. Discuss a linkage level descriptor with other panelists (NOTE: they should NOT be discussing their ratings at this point)
8. When all panelists confirm they are ready to move on, go to the next step.

<p>10. Independent evaluation: Panelists independently evaluate the profiles in each folder and sort the profiles into two categories – those that are in the higher performance level and those that are not. Panelists complete their rating sheet accordingly and turn in their ratings.</p>	<p>Materials: Same as previous step</p>	<p>Recommendations: Use paraphrase of recommended language to introduce the task Make sure panelists understand the task (what ratings they are making, and how) Monitor for: independent work, understanding of how to record choices</p>
---	---	--

1. Introduce the task.
 - a. Explain the purpose of the rating task: to label each profile with the performance level descriptor that best describes it.
 - b. **SAY the key question to consider: *Using your best professional judgment and considering all students with significant cognitive disabilities, which performance level best describes this profile?***
 - c. Reminders for panelists:
 - i. These profiles are examples of how students might master the number of linkage levels.
 - ii. Keep the PLDs handy. Use them to help guide your ratings.
 - iii. If you want to rate a profile differently than you rated a similar profile during the range finding step, that is okay. But because your pinpointing profiles were selected based on your range finding results, the outcome of your pinpointing activity should fall somewhere within the range of LL#s provided in this folder.
 - iv. **If you are stuck on the fence between two possible ratings, you should really feel confident that a profile shows the minimum amount of mastery needed to be at the upper performance level. Otherwise, it should be at the lower level.**
 - v. **If there are certain linkage level descriptors that you believe are very important for students to have achieved in order to be at a certain performance level, you may want to make a special mark by those on your blank profile.**
2. Remind them how to complete the **pinpointing rating form**:
 - a. Only fill in the “first rating” part of the form on this round.
 - b. Only mark one X per row.
 - c. Make sure they are marking the row that corresponds with the profile they are looking at.
3. Tell panelists they are welcome to go through the profiles in whatever way they wish.
 - a. They can take profiles apart but should be careful about keeping pages of profiles correctly grouped.
 - b. They can change their minds about any profile and adjust a rating any time up until the round is finished.
 - c. **They should not be discussing their work with others during this phase.**
4. When they are finished, they should check to make sure they have only one X on every row and that no rows have been skipped.

- ➔ As they are working, monitor for:
 - independent work
 - understanding of how to record choices (glance at rater sheets to make sure there is just one rating per row, all are in the unshaded columns, and all are in the first rating column)
- ➔ When panel work is going as intended, you may skip ahead to open the [facilitator workbook](#) and follow the “set-up” step in “Round 1 results”.

<p>11. Round 1 results: Facilitator polls the panel via a raise of hands to find the number of panelists who rated each profile in each of the performance levels. The facilitator records the recommendations in the facilitator workbook.</p>	<p>Materials: Same as previous step PLUS: Projected: Facilitator workbook</p>	<p>Recommendations: Ratings must be entered with 100% accuracy.</p>
---	---	--

Set-Up

1. Return to the pinpointing sheet in the [facilitator workbook](#).
 - a. If you reach the pinpointing step on a different day or closed the file after range finding, follow the same prompts to “save as” but save the final the exact same way you did the previous time.
2. Enter the profile ID numbers, in order, in Column B. Make sure they are with the correct “STEP” (Column A) and LL# (column C).

Recording

3. Identify two panelists who will watch the hand counts and what is recorded in the [facilitator workbook](#). (This responsibility should rotate among panelists from round to round.)
4. Mention each profile by ID number. Ask how many panelists rated it as:
 - a. [category below the cut point],
 - b. [category above the cut point],

For example: “Raise your hand if you categorized profile 1009 as emerging.”

5. Count the number of hands raised in each category for every profile and record in the corresponding cell of the worksheet.
 - a. These entries go in columns D-G (Group Frequencies of Round 1 Ratings). **Make sure you enter results for the correct profile ID and LL#. See table below.**

Cut Point	Step	Columns for round 1
AP/T	1	E & F

T/ADV	2	F & G
EM/AP	3	D & E

- b. The Total column will remain highlighted in red until it equals the number of panelists entered by the facilitator at the beginning of the range finding phase.
6. Ask watchers to confirm that the numbers entered for each performance level column match the number of hands that were raised.

<p>12. Table discussion of round 1 results: A summary of the table's initial ratings is presented to the panelists at the table.</p>	<p>Materials: Same as previous step Chart paper</p>	<p>Recommendations: Avoid making judgments about the ratings they chose. Use neutral language. Ask open-ended questions that allow panelists to explain their thought process Attend to group dynamics Discussion should focus on what exemplifies a performance level and differentiates it from another Make sure every panelist has a chance to speak</p>
--	---	--

1. Remind panelists of the following:
 1. The purpose of this discussion is to explore the thinking behind their first round ratings. They will have a chance to share their reasons and to hear about other panelists' reasons for their ratings.
 2. They may disagree, but we are looking for consensus building by actively sharing our own rationales and listening carefully to others.
 3. Based on the range finding activity, their recommended cut point for this cut will be within the range of LLs in their pinpointing folder.
2. Ensure that the panelists understand the definitions of the PLDs and how the profiles are examples of ways a student might perform at that total LL level. They should review any notes they made on their blank profiles as needed.
3. Instruct them **NOT to erase or mark out any of their first round ratings during this step**. They should not mark on their rating sheets at all. If they want to make notes to themselves, they should use their notebook or blank profile sheet.

Discussion Hints for This Step

Discussion should focus on what **exemplifies a performance level** and what **differentiates one profile from another**.

- How does this linkage level exemplify the performance level you gave it?
- Are there any profiles where you were really on the fence at first? How did you come to your final decision?
- How did this profile change your understanding of the performance level?
- Here's a place where the ratings were evenly split [Or, where two profiles with the same LL# got different results]. Let's take a closer look.
- How are these two profiles different (or the same)?

4. Use the frequency distributions to guide group discussion where the frequency distribution:
 1. **Shows a split panel (nearly half in each of two groups)**
 2. **Has a different general pattern than the other profiles with the same LL#.**
 3. **Shows an unexpected pattern for adjacent LL#s (for example, people tended to rate one with higher LL# as lower performance level).**
5. By the end of the discussion, some panelists may be leaning toward changing their ratings and others may not. While unanimous agreement is not a requirement, you should hear some evidence that raters are considering changing their ratings.
 1. Remind them that the goal at this stage is for any disagreement about which category profiles belong in to be in the vicinity of adjacent LL#s. (We shouldn't see split votes at the upper and lower ends of the pinpointing range and agreement in the middle.)
 2. **If they are stuck on the fence between two possible ratings, they should really feel confident that a profile shows the minimum amount of mastery needed to be at the upper performance level. Otherwise, it should be at the lower level.**
 3. Do **not** ask them to vote at this stage. Just monitor conversation and especially listen for convergence (i.e., panelists have shared their opinions and others have heard them and to some extent agree with the opinion). They will have a chance to make their ratings again during the final round.
 4. If you have concerns about the discussion at your table or if questions are raised that you are not comfortable answering, ask your runner to bring the room lead to the table. If discussion is leading to greater likelihood of convergence, wrap up when everyone has had a chance to discuss as they like.
6. **Grade level PLDs: Before moving on to the next step, ask panelists to consider whether there are any linkage level descriptors that they believe are very important for a student to have in order to be categorized into one of the categories being considered at this phase.**
 1. **Ask them to write each LL descriptor or concept that is very important and that exemplifies that performance level one on a sticky note.**

- i. Use EE code and linkage level OR write short description of the concept.
2. They should hand it to you and tell you which of the two categories it belongs in.
3. Place the sticky note on the corresponding side of the line on the chart paper.
4. If there is some agreement (not even full consensus) that the skill is important but disagreement about where it belongs, put it right on the line.

➔ If you prefer to do this activity directly in a word document with the four performance levels clearly marked, you can do that instead of the sticky note version. Save the document in dropbox.

➔ If you do the sticky note version, while your table is engaged in a future independent step (ratings, reviewing materials) enter the sticky notes in your word document so you can project.

<p>13. Round 2 ratings: Panelists independently complete their final ratings.</p>	<p>Materials: Same as previous step.</p>	<p>Recommendations: Make sure panelists understand the task (what ratings they are making, and how) Monitor for: independent work, fatigue with process (quick ratings to get it over with)</p>
---	--	---

6. Explain the purpose of the rating task: to label each profile with the performance level descriptor that best describes it.
 - a. **SAY the key question to consider: *Using your best professional judgment and considering all students with significant cognitive disabilities, which performance level best describes this profile?***
 - b. Remember that these profiles are examples of how students might master the number of linkage levels.
 - c. Consider the discussion we just had, but remember that you do not have to reach a unanimous decision. You should still complete your ratings based on your own professional judgment.
 - d. Use the PLDs to guide your decisions.
 7. Remind them how to complete the range finding rating form:
 - a. Only fill in the “final rating” part of the form on this round.
 - b. Only mark one X per row.
 - c. Provide the rating for every single row – even if they are not changing their answer.
 - d. Make sure they are marking the row that corresponds with the profile they are looking at.
 8. Tell panelists they are welcome to go through the profiles in whatever way they wish.
 - a. They can change their minds about any profile and adjust a rating any time up until the round is finished.
 9. When they are finished, they should check to make sure they have only one X on every row.
 10. As they finish, ask each panelist to reassemble any **profiles** they have taken apart and return the profiles to their folder. They should put their **blank profile** with their notes on it back into their **resource notebook**.
- As they are working, monitor for:
- independent work
 - understanding of how to record choices (glance at rater sheets to make sure there is just one rating per row and all are in the first rating column)
- If doing the first or second pinpointing, notify runner that you are nearly finished with this round of pinpointing and will be needing the next materials soon.

<p>14. Round 2 ratings submitted: Panelists show their votes via a raise of hands, like in previous steps. Revised ratings are entered into the facilitator workbook and shared with the group.</p>	<p>Facilitator workbook</p>	<p>Ratings must be entered with 100% accuracy.</p>
---	-----------------------------	--

1. Using same process as before, raise hands and count again for Round 2 Final ratings. Complete the columns that correspond to Round 2 ratings.
 - a. The Total column will remain highlighted in red until it equals the number of panelists entered in Step 1 by facilitator.
 - b. **Make sure you enter results for the correct profile ID and LL#. See table below.**

Cut Point	Step	Columns for round 2
AP/T	1	Z & AA
T/ADV	2	AA & AB
EM/AP	3	Y & Z

- ➔ If this step is not timed well with a break built into the agenda, give panelists a 15-minute break.
- ➔ If there is additional wait time for the next pinpointing files to be prepared, here are two options:
 - **Ask if they want to adjust the location or content of any of the sticky notes**
 - Have panelists start on the review of materials/resources and their blank profile with their first cut point in mind. Start thinking about what it would take to reach the higher of the two performance levels that will be considered at the next pinpointing step.
- c. Continue to next pinpointing phase (i.e., next cut point) OR proceed to grade level PLD phase.
NOTE: Once Round 2 Final ratings are complete, disconnect or turn off projector.

Collect **pinpointing rating forms** and **folders**. Turn in to the **runner**.

SPECIAL PROCEDURE (if needed): Round 3 ratings Panelists complete discussion of round two ratings and submit a third set of ratings.	Same as previous pinpointing steps PLUS third round pinpointing form (pink sheet)	Same as previous pinpointing steps, although discussion goals may be tailored to the source of concern about ratings.
--	---	---

If your panel does not successfully identify a cut point during a pinpointing round, your psychometrician may ask you to do a third round of pinpointing. You will receive special instructions about how to handle this round.

1. Pay attention to the patterns of data from round 2 that contributed to the need to go to round 3.
2. Carefully guide group discussion during this phase, especially in these areas:
 - a. illogical patterns (e.g., lower LL profiles being rated at the higher performance level and vice versa)
 - b. signs that their ratings would put the cut point outside the range used in this step of pinpointing
 - c. Inconsistent ratings of profiles within the same LL#
3. Have panelists complete their third round of pinpointing in the “final round” columns on the pink pinpointing sheet. There are no additional rounds of discussion or rating.

RECOMMENDED CUT POINTS

Complete this step only after all three pinpointing steps are complete for a grade/course.

15. Announce the table’s recommended cut points: Facilitator shares final results for the grade/subject with the panel.	Facilitator workbook	N/A
--	----------------------	-----

1. Display the final recommendation results in columns AM – AO of the **facilitator workbook** pinpointing sheet. Out loud for the group, summarize the cut points using the yellow cells in the “predicted” row.
 - a. Column AM: “Based on logistic regression results, a student would have to master a minimum of ___ linkage levels in order to be at or above the **approaching target** level.”
 - b. Column AN: “Based on logistic regression results, a student would have to master a minimum of ___ linkage levels in order to be at or above the **at target** level.”
 - c. Column AO: “Based on logistic regression results, a student would have to master a minimum of ___ linkage levels in order to be at or above the **advanced** level.”
- ➔ If they ask how many students would achieve at those levels, remind them that the analysis of impact data will come at a later stage and will be evaluated by the Technical Advisory Committee and the state education agencies.

GRADE LEVEL PLDs

16. Update PLD brainstormed list.	Materials: Projected word document with notes from flip chart paper (all three pinpointing phases)	Recommendations: Focus on skills that are important milestones and have the greatest consensus
-----------------------------------	---	---

Once the group has completed the discussion of their independent ratings in pinpointing the panel will review the nodes/KSAs included in the performance level categories in the projected word document table and refine the list.

Open the **PLD word document** from your table’s dropbox folder. Locate the correct grade and cut point.

As time permits, if they placed anything on the fence during an earlier round of discussion, ask them to move that skill to one side or the other based on the final ratings.

THE FINAL STEPS

<p>17. Discussion and revision of a difficult profile: Panelists refer back to one self-selected profile with unusual evidence of mastery and discuss what would make a difference in the rating for that profile.</p>	<p>Materials: Step 3 (EM/AP) pinpointing folders (or other step upon request) Clean copy of their self-selected profile (one for the whole table) Blank paper</p>	<p>Recommendations: None</p>
--	---	----------------------------------

1. Introduce the activity: The purpose of this activity is to have panelists explain how one of the pinpointing profiles could have been rated at a higher performance level if it had a different pattern of mastery.
2. Have the group identify one **profile** they would like to use for this activity.
 - a. It could be from any of the three steps of pinpointing, but the final pinpointing ratings should have placed it in the lower of the two levels or the vote should have been split.
 - b. The best choice is probably the one that had the same LL# as others but looked sufficiently different that the group tended to rate it differently.
 - c. Their selected profile probably generated a lot of discussion.
3. Ask the **runner** to retrieve the necessary pinpointing folders. Locate a clean copy of that profile.
4. Have the group discuss what additional LLs would have to be mastered in order to achieve at the next highest performance level.
5. Where there is consensus (but not necessarily 100% agreement) about what those additional levels would be:
 - a. Highlight those cells on the **clean copy** (i.e., edit the profile to show what the group believes would put it at the next highest performance level)
 - b. On **blank paper**, write the table ID, subject, and grade and a short explanation of why the changes led the group to think the edited profile now reaches the next performance level.

<p>18. Final independent ratings: Panelists review their table’s recommended cut points for the grade. Independently, they evaluate those recommendations and each make a final recommendation.</p>	<p>Materials: Evaluation of Standard Setting Results Form</p>	<p>Recommendations: None</p>
---	---	----------------------------------

1. Pass out the **Evaluation of Standard Setting Results** forms. Have panelists fill in the header information.
2. Point out the three cut points that are recommended by their table based on the previous pinpointing round. Have panelists fill in these values. (The psychometrician will enter these values on one copy of the form for you.)
3. With all the benefit of previous conversations but no further discussion, each panelist:
 - a. Indicates whether or not s/he is comfortable with the table’s recommended cut point.
 - b. Provides a final independent rating for what that cut point should be.

Check each form for completeness when it is turned in.

Facilitator Hints

Key responsibilities:

- Build rapport while also showing respect for the panelists, their states, their students, and the assessment system
- Manage timelines and group dynamics
- Listen actively
- Ensure panelists follow intended procedures while rating
- Make sure everyone is on the same page before a task starts

During discussions

- Avoid making judgments about the ratings they chose. Use neutral language.
- Ask open-ended questions that allow panelists to explain their thought process.
- Attend to group dynamics. Do not interrupt panelists. Encourage panelists to listen to one another and not interrupt one another.
- Make sure every panelist has a chance to speak and that different perspectives (e.g., general and special education, classroom teacher and other professional) are represented.
- Minimize the impact of dominant panelist(s).
- Make sure discussion matches the purpose of the phase you are in. Guide it back to that purpose if it gets off track.

Problems and Possible Solutions – General Group Dynamics

Problem	Response
Table talk	<ul style="list-style-type: none"> • Remember, you should only be recording ratings based on your own opinions right now. • We will have a chance to compare notes later.
Quiet panelists	<ul style="list-style-type: none"> • [Name], we haven't heard from you yet. Could you tell us about your ratings (or thought process)?
Outliers	<ul style="list-style-type: none"> • I notice your rating is a bit different from the others. Can you tell us what led you to that decision? And what additional information would you like from the members of this panel regarding their rationale for their ratings?
Personalizing	<ul style="list-style-type: none"> • Let's remember these ratings should be about students in the DLM population in general – not just the individual students we know. • It is great that this process is making some connections for you in terms of instruction. But for now, let's get back to the standard setting task.
Approval seeking	<ul style="list-style-type: none"> • Great question. That's why we brought educators like you together. Your professional judgment is what helps determine the recommended outcomes.
Dominance	<ul style="list-style-type: none"> • Good thought. Would someone like to respond to/add your idea on that point? • That is an interesting perspective based on your role as a (content specialist). What do the (call on different role – special educators) at the table think?
Wariness of conflict	<ul style="list-style-type: none"> • Remember, we don't have to agree but we need to understand where our differences are.

Problems and Possible Solutions – Making Ratings

Besides the *Panelist Hints* page, consider these:

Problem	Response
Slow raters	Dependent on which phase you are in: <ul style="list-style-type: none"> It may be hard to make a decision, but remember, we will be have more chances to refine our ratings. OR These final phases can be difficult as one profile may not look very different than the adjacent profile. Remember, the Technical Advisory Committee and the states will be reviewing these recommendations.
Fast raters	<ul style="list-style-type: none"> Are there additional resources you would like to see before making a decision? (offer to show testlets, etc).
Verb obsessed	<ul style="list-style-type: none"> Be careful not to rely too much on verbs to create rules for yourself. Complexity of the skills at each linkage level comes from a combination of the content <i>and</i> the cognitive process expected.
This doesn't look like my kid	<ul style="list-style-type: none"> Remember that the profiles come from a variety of real students across DLM states. Others are simulated based on real students. With only a few examples in your folder, you might not see your own student's performance. When you make your ratings, you consider all students with significant cognitive disabilities, not just single students.
How can these two be at the same LL#?	<ul style="list-style-type: none"> Those profiles show two very different ways of getting to the same overall number of linkage levels mastered. Students in this population have very different skill sets. Just rate each profile based on the key question and remember the minimum threshold to get to the higher level.
These kids can't do that.	<ul style="list-style-type: none"> Remember, this is a new alternate assessment aligned to challenging academic standards. This assessment <i>does</i> reflect high expectations for students with significant cognitive disabilities. The DLM consortium states approved these expectations. We want all students with significant cognitive disabilities to have the opportunity to strive for these goals. The students in this population are pretty diverse, and there are some students who can do these things.
There's a mistake in this profyle!	<ul style="list-style-type: none"> Thanks for catching that! We want to make sure it is correct for score reporting. Please note the problem and the profile grade/subject and ID on an index card so I can turn it in.
What do YOU think?	<ul style="list-style-type: none"> The results of this panel need to represent expectations from the partner states. My role is to facilitate your process, not to influence the outcome.
What about all these empty lines in this profile?	<ul style="list-style-type: none"> There are many reasons why lines could be blank. A student might not have tested on that EE. Or might have tested, but not demonstrated mastery. Focus on the skills the student mastered when making your ratings – not the ones that are blank.
Should some EEs be weighted more than others?	<ul style="list-style-type: none"> The DLM consortium states approved test blueprints to cover all of the Essential Elements on the assessment, as we've described. You might think that certain skills are extremely important and others are less important. There is an opportunity for you to share those opinions during this panel process. But there is no quantitative "weighting" of EEs so that some count more than others. Your ratings should reflect your best professional judgment.

Problem	Response
Is the target linkage level the same as “at target”?	<ul style="list-style-type: none"> • The target linkage level for the Essential Element indicates the grade level expectation for that EE. The “at target” performance level is a more general description of the student’s performance across all EEs on the assessment. The panel’s job is to make a recommendation about how many skills must be mastered in order to be “at target”. A student does not necessarily have to master the target linkage level on every single EE in order to be at the “at target” level.
I just can’t decide!	<ul style="list-style-type: none"> • Just put it in the category that is the best fit, based on your professional judgment. You should really feel confident that a profile shows the minimum amount of mastery needed to be at the upper performance level. Otherwise, put it at the lower level. • [For all except pinpointing round 2]: That’s okay. There is more than one chance to rate a profile.

Panelist Hints: Making Your Ratings

Key question to consider:

Using your best professional judgment and considering all students with significant cognitive disabilities, which performance level best describes this profile?

Hints:

1. **Make sure you feel ready to start rating before you rate.** Spend time reviewing the resources and sample testlets so the process of evaluating each profile is easier.
2. **Consider the skills for which the student demonstrated mastery** (green shading). Shaded and unshaded skills may help you think about a profile, but do not focus on why a linkage level is not shaded. No shading means there was no evidence during the assessment that the student mastered that skill this year. You are deciding on the performance level that best describes the profile based on what was mastered.
3. **Use the Performance Level Descriptors (PLDs) to help you make your ratings.** Refer to any notes you may have taken during discussions of what the PLD descriptors mean.
4. **When you're on the fence, think about what it takes to hop the fence.** If you are undecided between two possible ratings, you should really feel confident that a profile shows the minimum amount of mastery needed to be at the upper performance level. Otherwise, it should be at the lower level.

Write your own additional hints and reminders here:

Appendix H: Example Rating Forms for Range-Finding and Pinpointing

DLM Standard Setting Rating Form – Range Finding

Panelist ID: _____ Table ID: _____ Room: IM YE Subject: ELA Math Grade/Course: _____

	Profile ID	# LLs	First Rating				Final Rating			
			EM	AP	T	ADV	EM	AP	T	ADV
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										
23										
24										
25										
26										
27										

EM = Emerging AP = Approaching Target T = At Target ADV = Advanced

**DLM Standard Setting
Pinpointing Form: AP/T**

Panelist ID: _____ Table ID: _____ Room: IM YE Subject: ELA Math Grade/Course: _____

	Profile ID	# LLS	First Rating				Final Rating			
			EM	AP	T	ADV	EM	AP	T	ADV
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										
11										
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										

EM = Emerging

AP = Approaching Target

T = At Target

ADV = Advanced

Appendix I: Independent Evaluation and Meeting Evaluation Forms

Evaluation of Standard Setting Results

Panelist ID: _____ Table ID: _____ Room: IM YE Subject: ELA Math Grade/Course: _____

Cut Point	Table's recommended LL# (after pinpointing):	Are you comfortable with this recommendation?	What is your final, independent, recommended LL# for this cut point? <i>(If you agree with the table recommendation, copy that value here.)</i>
EM/AP		YES NO	
AP/T		YES NO	
T/ADV		YES NO	

Dynamic Learning Maps Standard Setting Panelist Questionnaire
June 2015

Please consider the statements below and place an "X" in a box to indicate the level of agreement or disagreement you have with each statement. A 4-point rating scale ranging from strongly disagree to strongly agree is provided. Please mark only one of the options for each statement.

	Strongly Disagree	Disagree	Agree	Strongly Agree
1. The overall goals of the standard setting panel meeting were clear.				
2. The panel meeting was well-organized.				
3. The training and practice exercises provided the information I needed to complete my tasks.				
4. It was clear what knowledge, skill, or ability a student would need to demonstrate to achieve a certain profile.				
5. I considered the <i>performance level descriptors</i> when I rated each profile.				
6. I considered the <i>assessment items</i> when I rated each profile.				
7. I considered the <i>other panelists' opinions</i> when I rated each profile.				
8. I considered <i>my experience in the field</i> when I rated each profile.				
9. I understood how to rate each profile.				
10. I had enough time to complete the tasks.				
11. I felt confident when rating the profiles.				
12. The procedure for recommending cut points was free from bias.				
13. Overall, I was satisfied with the ratings made by panelists in my group.				
14. The impact data was useful when making my final ratings.				
15. I would defend the group's <i>At Target</i> decisions against criticism that they are <i>too high</i> .				
16. I would defend the group's <i>At Target</i> decisions against criticism that they are <i>too low</i> .				
17. I would defend the group's <i>Advanced</i> decisions against criticism that they are <i>too high</i> .				
18. I would defend the group's <i>Advanced</i> decisions against criticism that they are <i>too low</i> .				

	Strongly Disagree	Disagree	Agree	Strongly Agree
19. I would defend the group's <i>Approaching Target</i> decisions against criticism that they are <i>too high</i> .				
20. I would defend the group's <i>Approaching Target</i> decisions against criticism that they are <i>too low</i> .				
21. I am confident that the meeting produced valid cut score recommendations.				
22. Overall, I believe my opinions were considered and valued by the group.				
23. Overall, my group's discussions were open and honest.				

Please consider the statements below and place an "X" in a box to indicate the level of agreement or disagreement you have with each statement.

	Strongly Disagree	Disagree	Agree	Strongly Agree	Not applicable
24. Participating in the process increased my understanding of DLM.					
25. Overall, I valued the panel meeting as a professional development experience.					
26. This experience will help me plan and provide instruction for my students with significant cognitive disabilities.					
27. This experience will help me use DLM more effectively.					

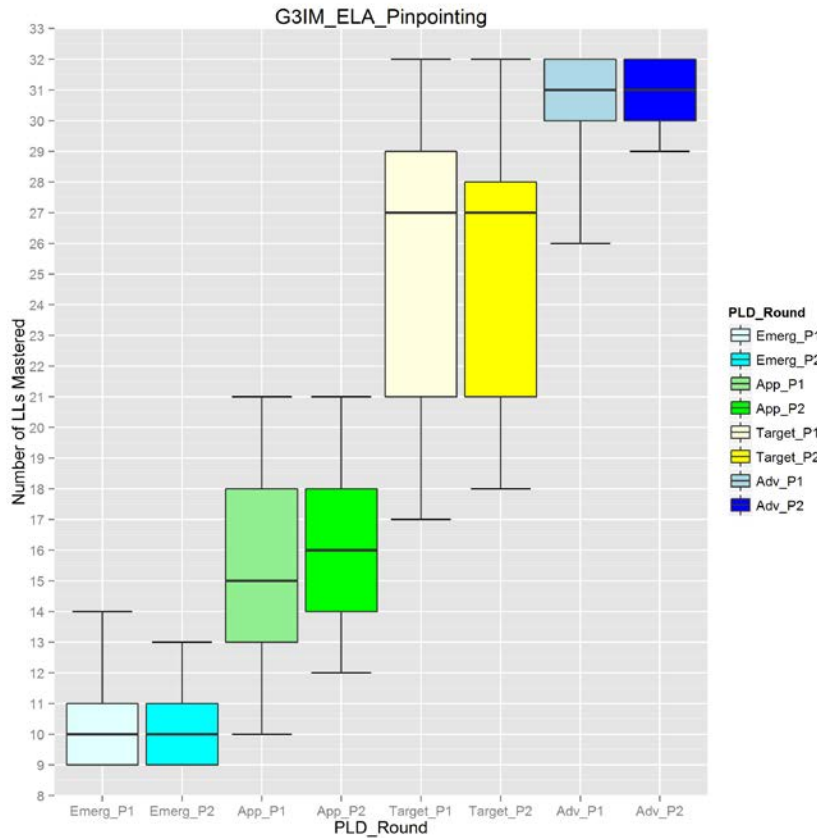
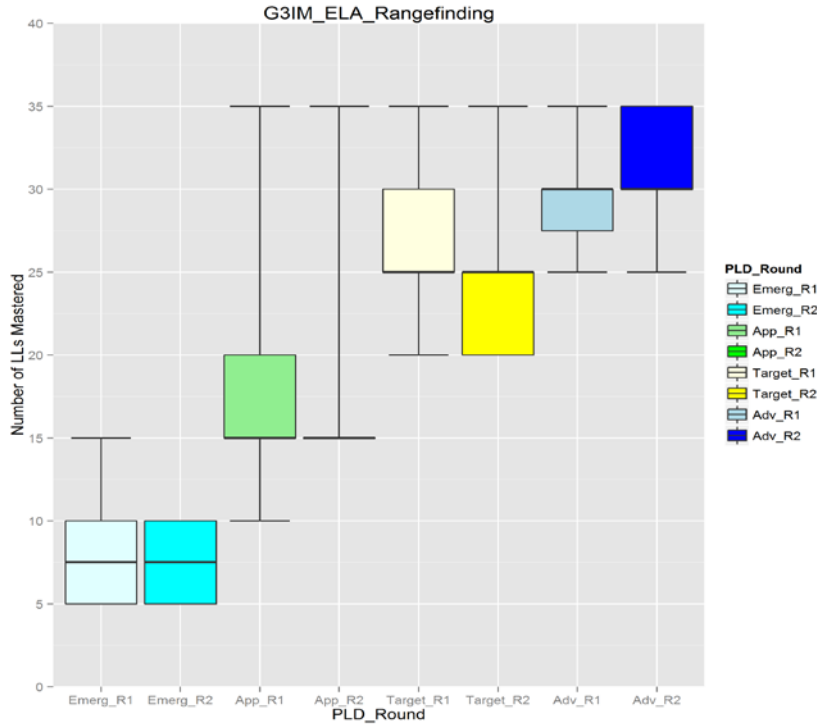
In the space below, please feel free to:

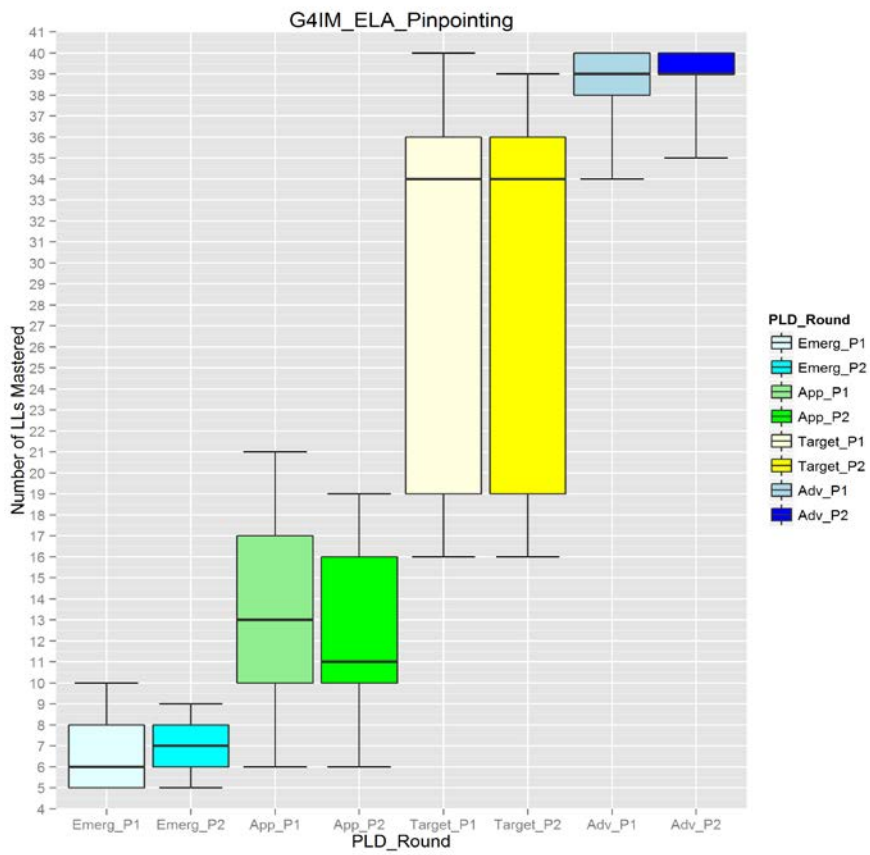
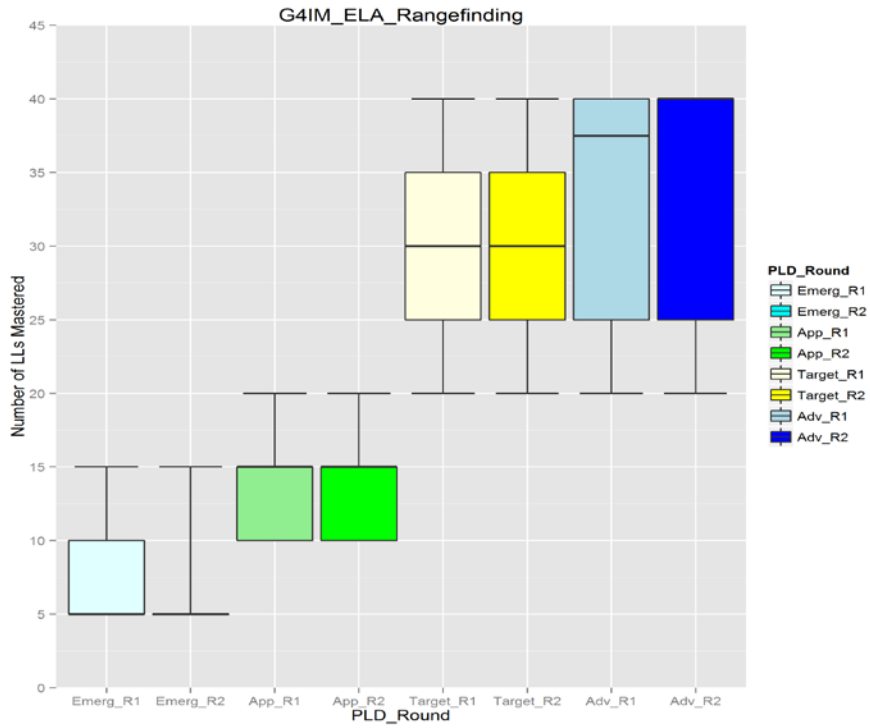
- Add comments regarding any of the responses to the questions above
- Make suggestions to improve future standard setting workshops
- Tell us what you liked and/or did not like about the workshop

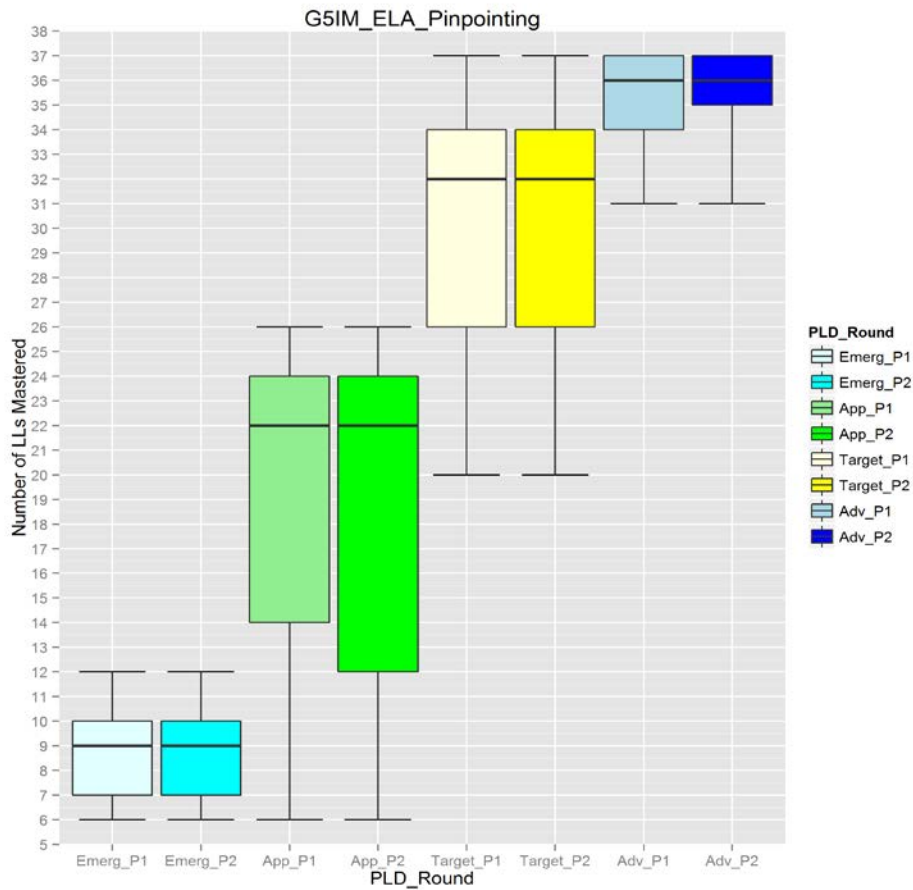
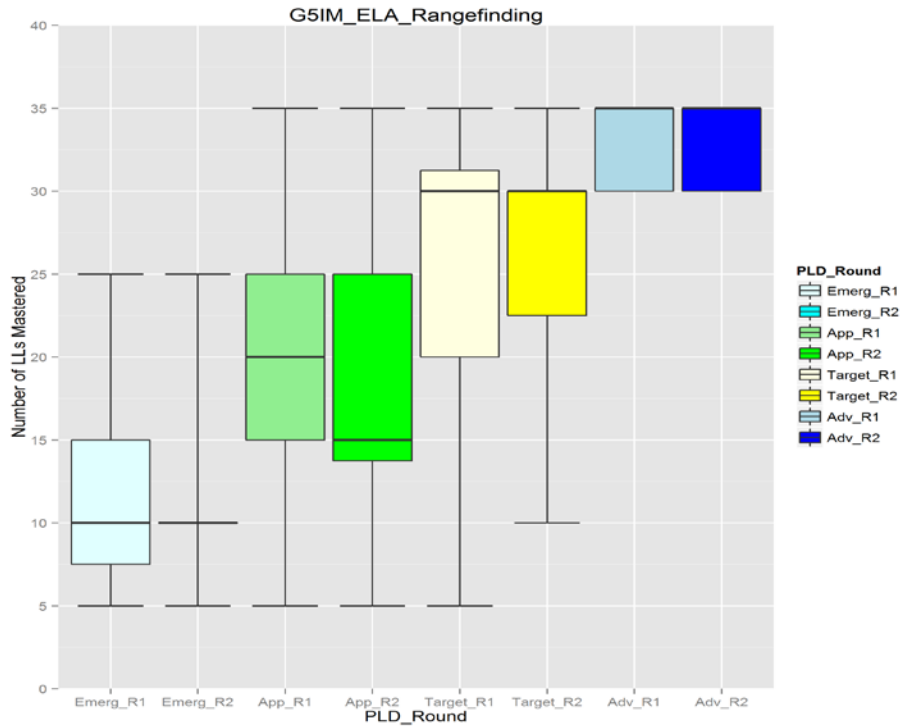
Appendix J: Original and Revised Standard Setting Procedures

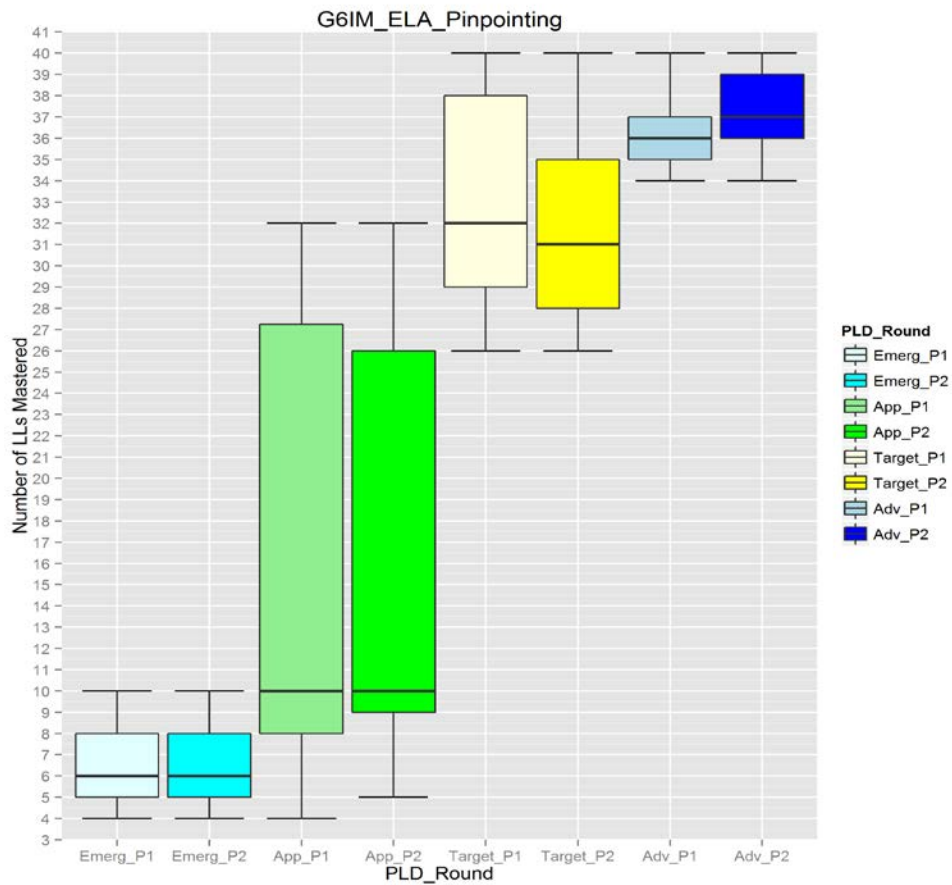
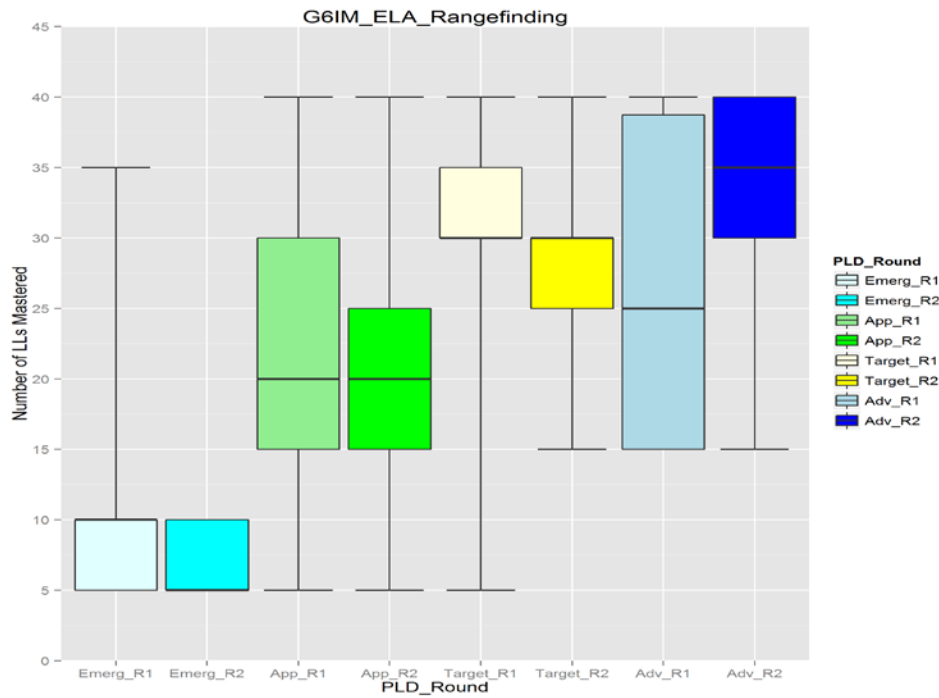
Procedure	Action	Rationale
Range-finding	None	N/A – Implemented as intended
Pinpointing	None	Procedures were implemented as intended. In a limited number of cases, panels required a third round with profiles at four linkage levels instead of seven.
Panel review of impact data	Removed	Impact data was not available on the anticipated timeline in order to be ready on day two. This decision was made the day before the panel meeting started.
Adjacent grade comparison	Removed	This step was originally included for panelists to self-smooth results by using impact data throughout the week. Without impact data, this step was no longer possible.
Grade/subject-specific PLDs	Modified	After the procedure was revised, panels still captured key ideas about content that differentiated one performance level from another. However, the procedure was simplified due to time constraints. Also, its location in the sequence of procedures was changed so panelists could work on this task while waiting for range-finding profiles to be delivered.
Capture rationale	Added & Removed	Added on day two as a data source to support panel recommendations when it became clear that impact data would not be available. Removed on day three when it became evident the task was time consuming and the results overlapped substantially with the grade/subject PLD notes rather than producing the more general rationale expected.
Revise a difficult profile	Added	Addressed panelists' challenges with rating profiles that had uneven patterns of mastery across Essential Elements. Collected as a secondary source of evidence about what amount of mastery differentiates two adjacent performance levels.
Final independent evaluation of panel pinpointing results	Added	Substitute for the rationale-capturing step that was removed. Also served as an evaluation of the panel process.

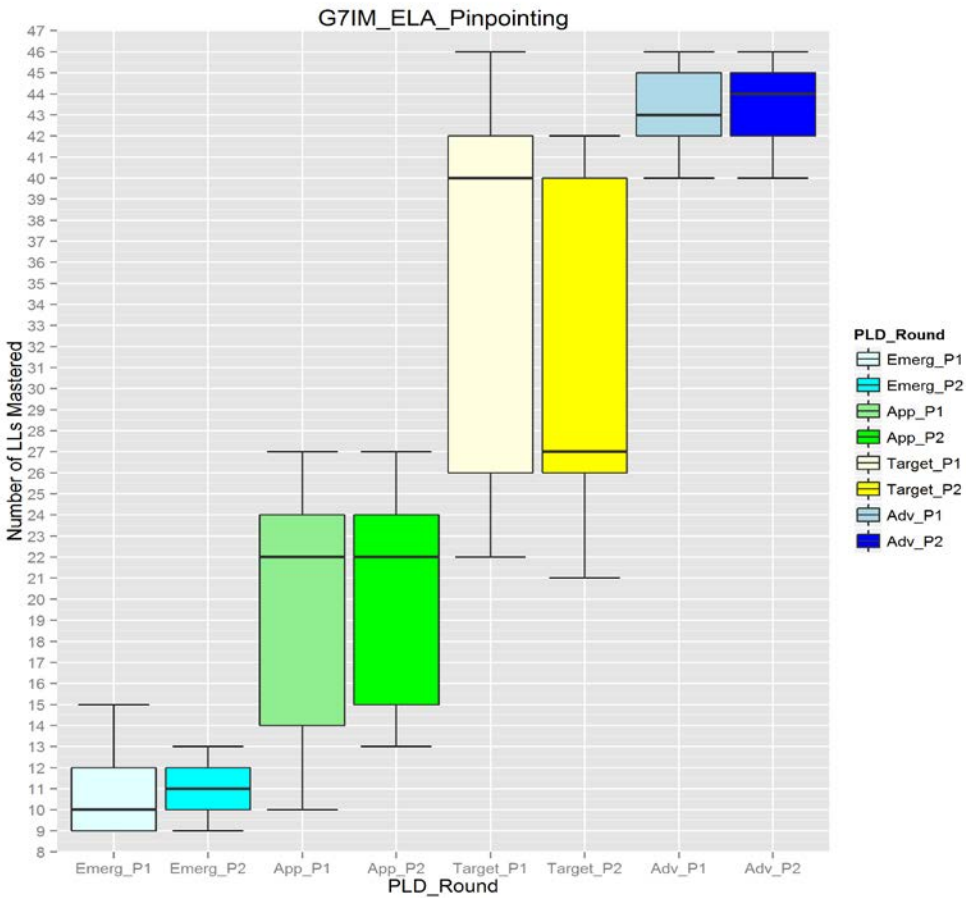
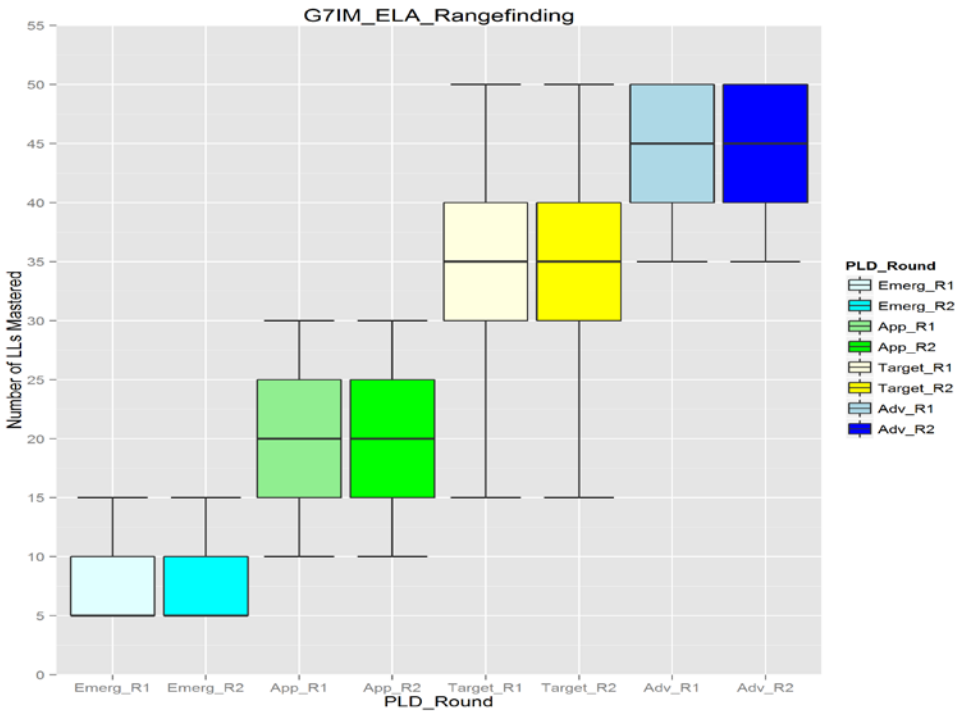
Appendix K: Convergence Plots for Range-Finding and Pinpointing Ratings

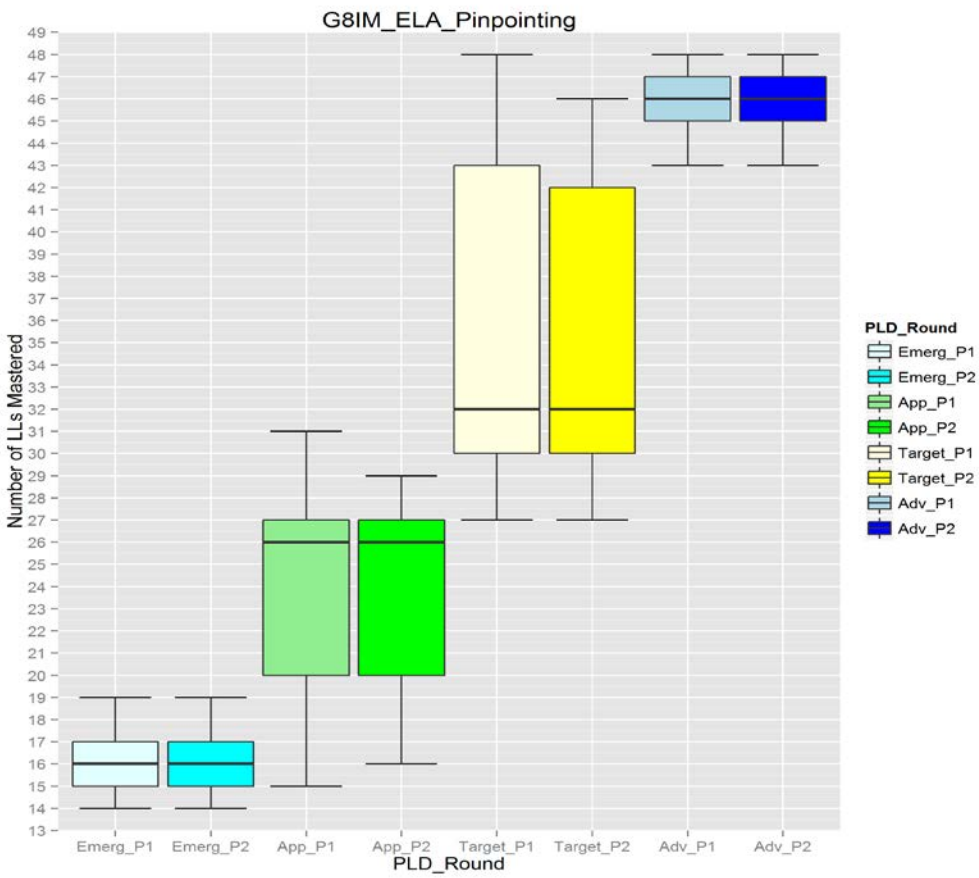
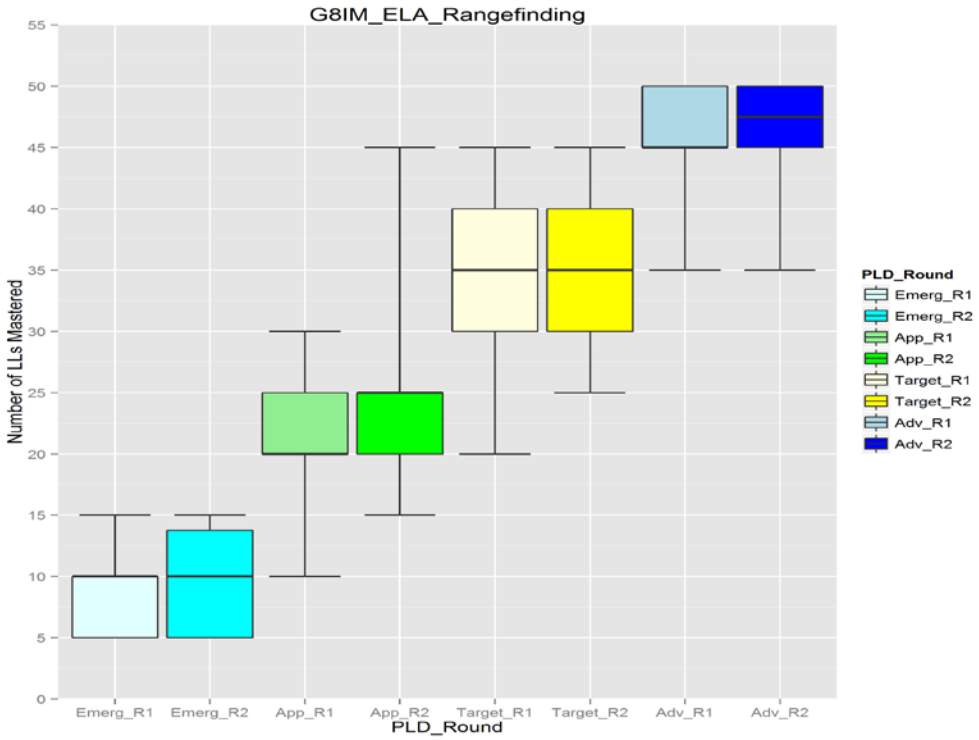


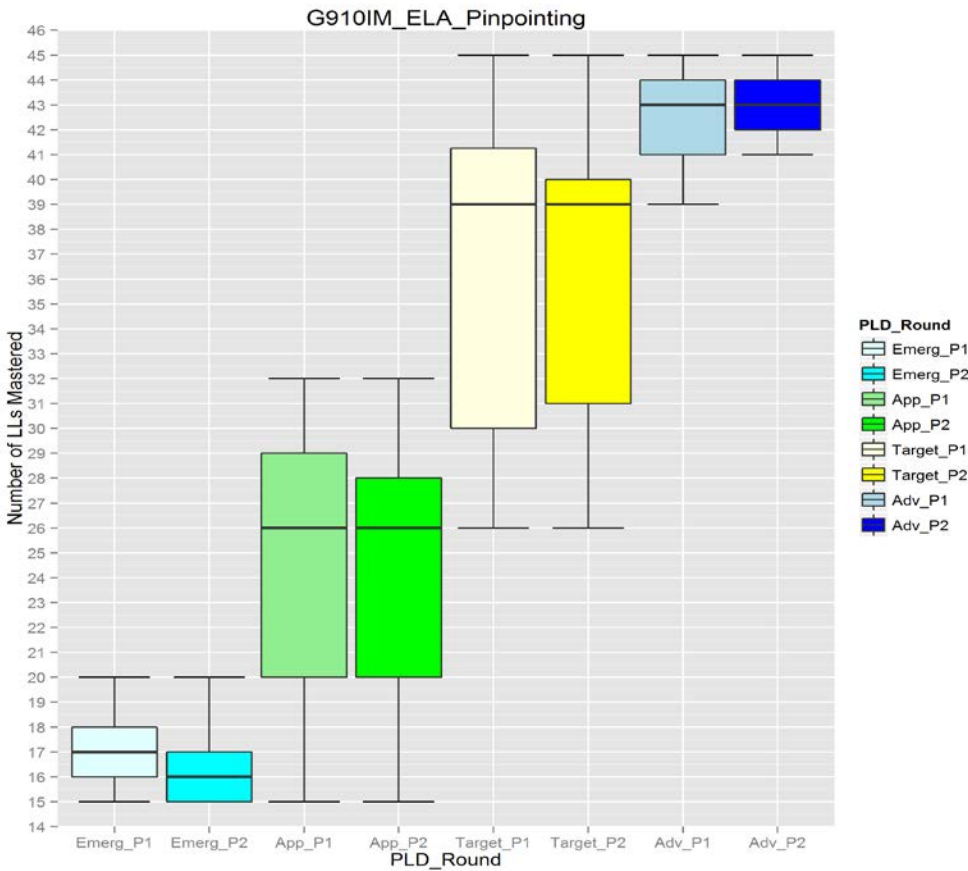
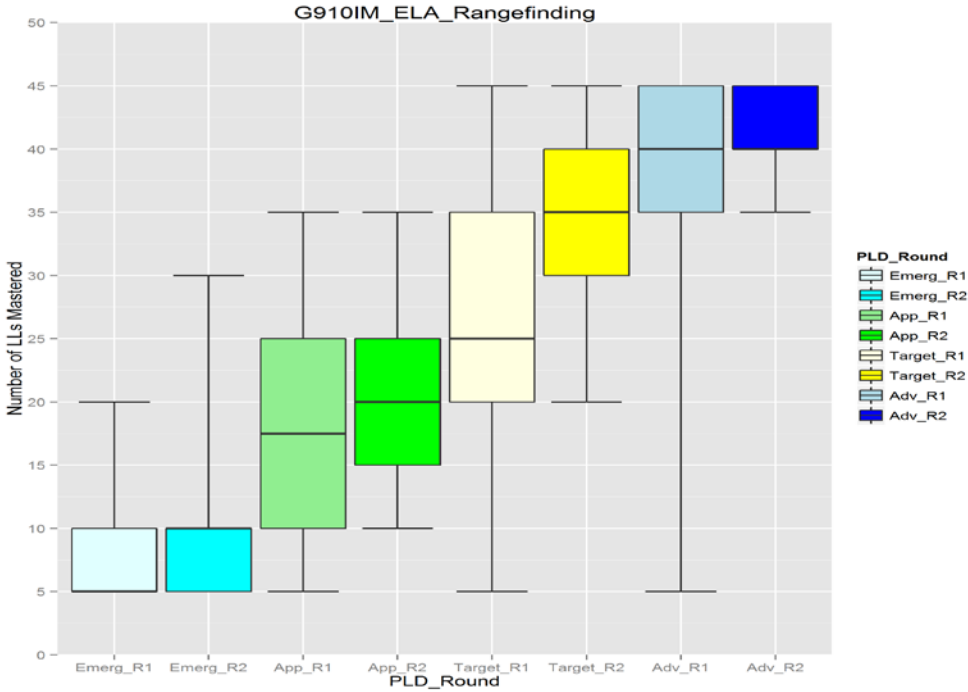


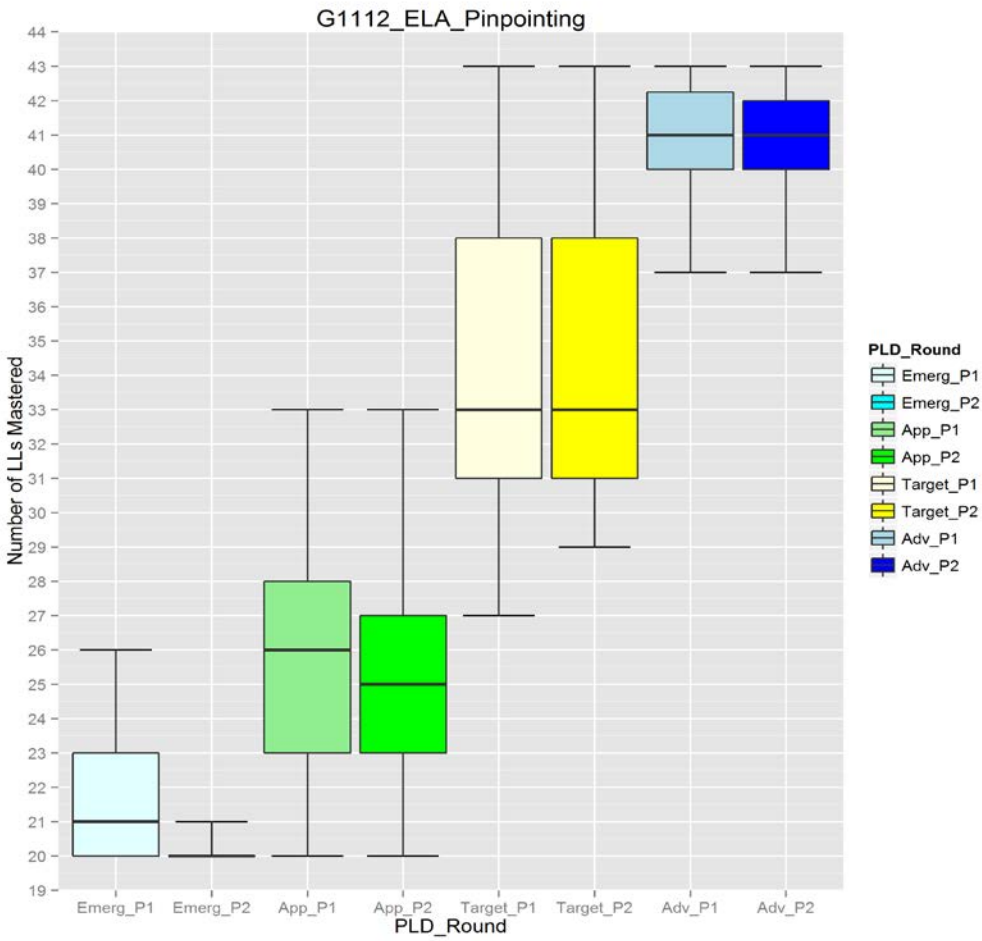
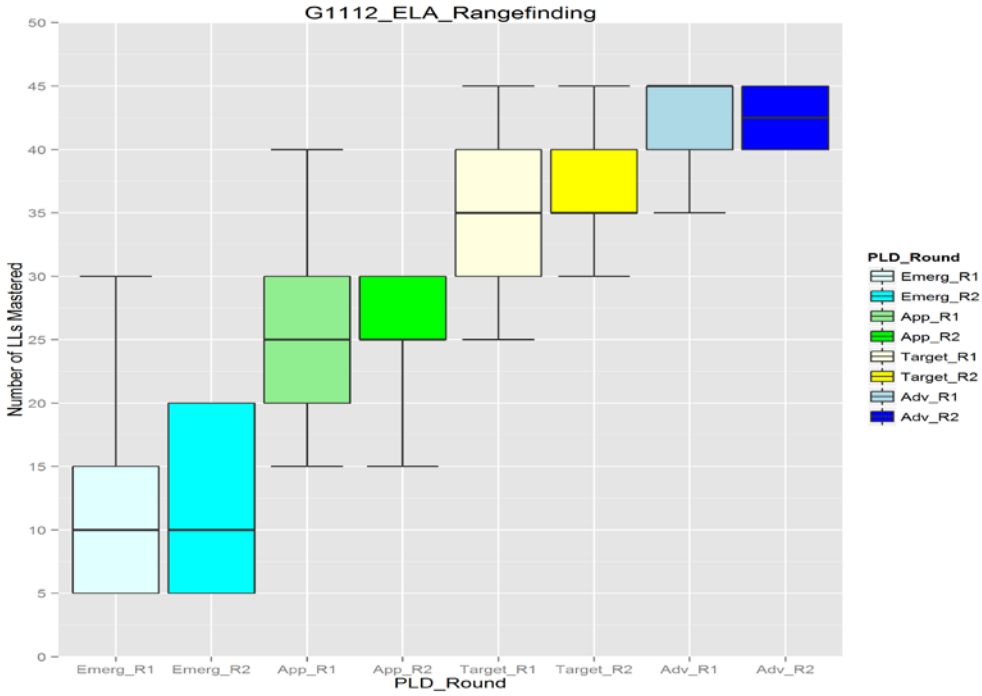


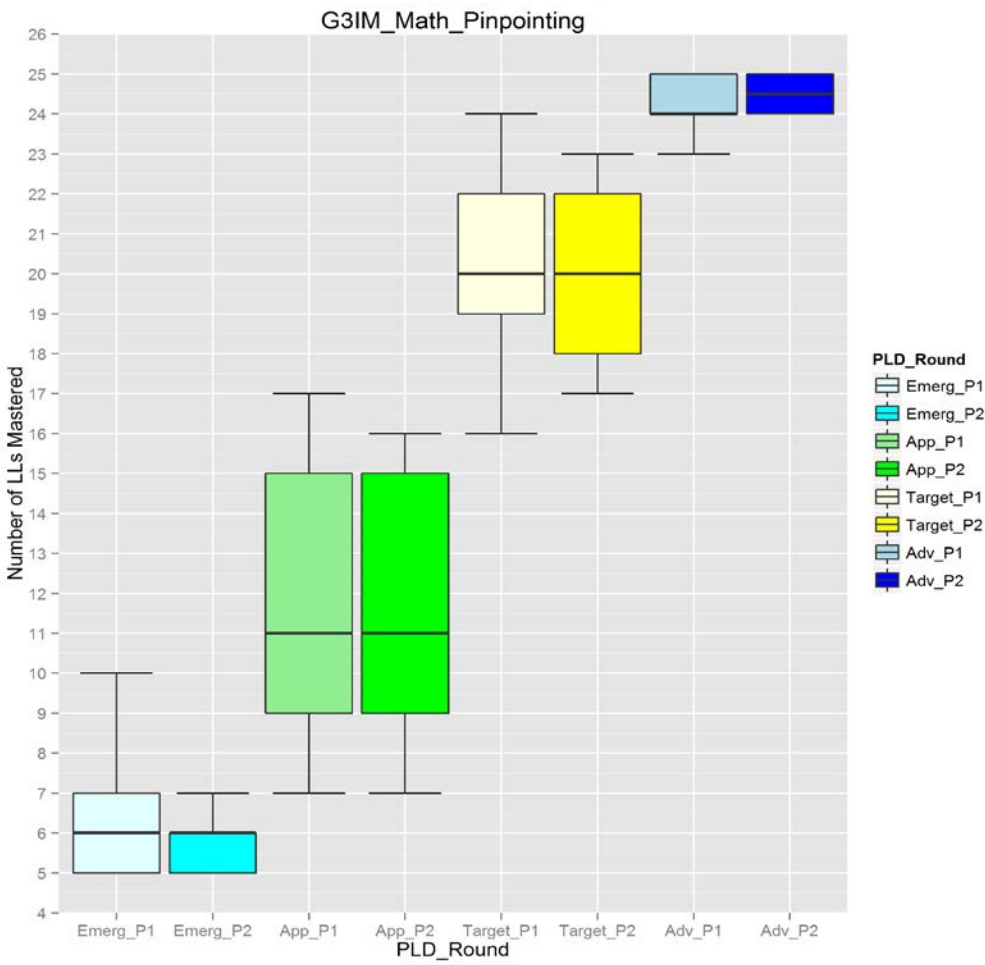
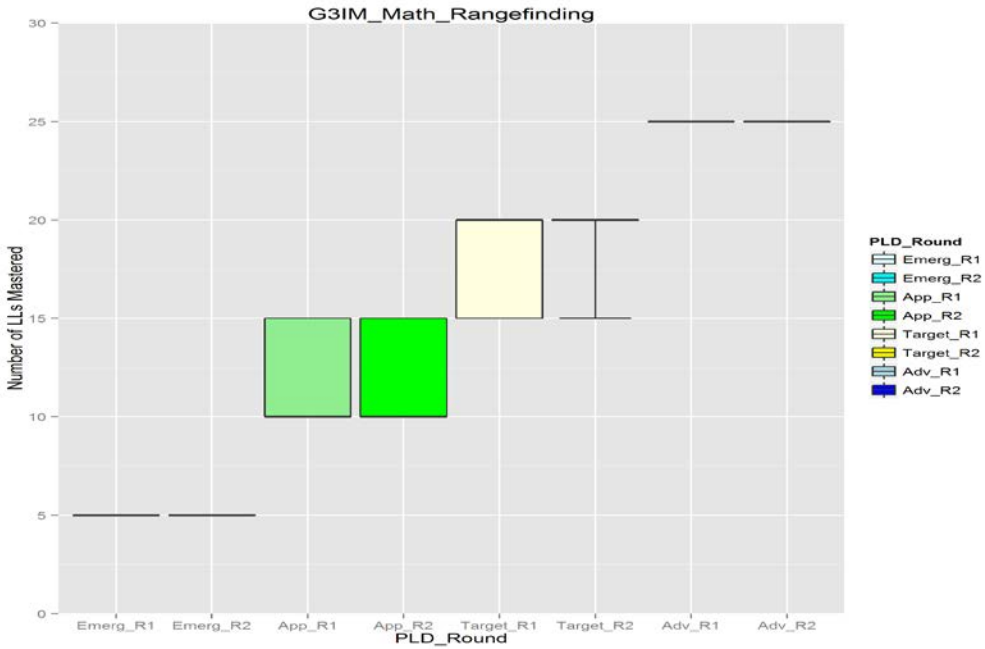


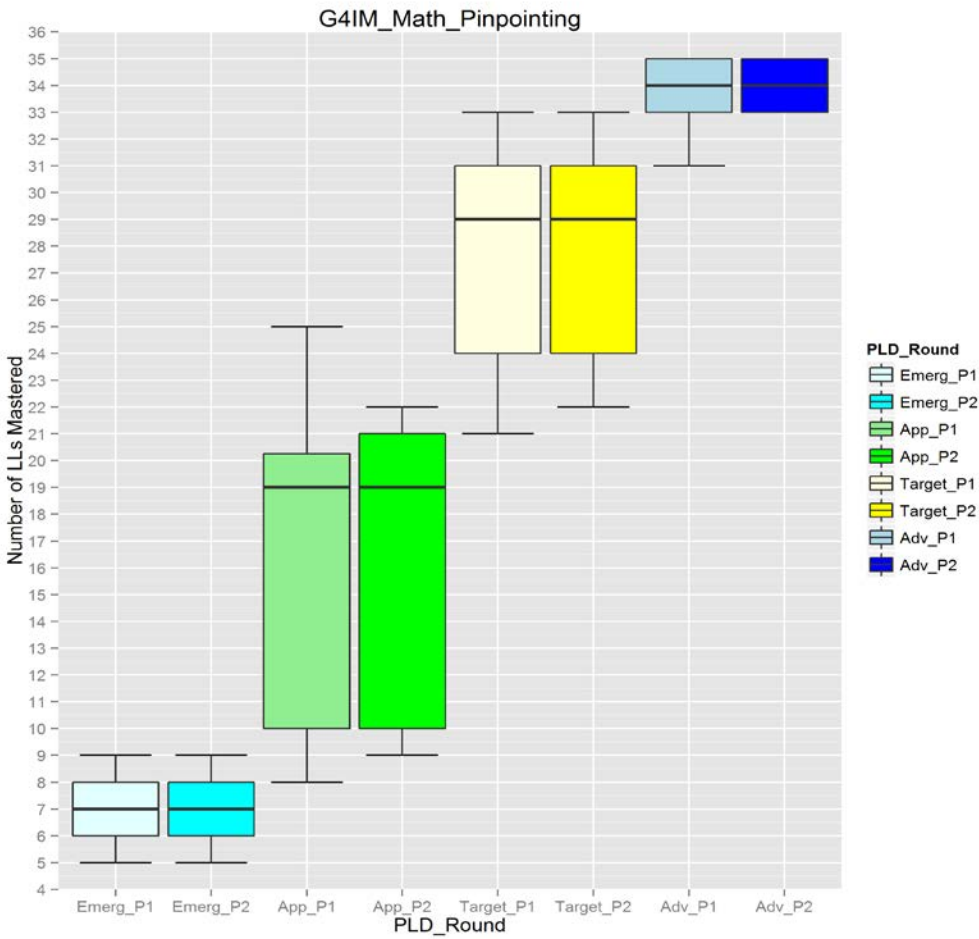
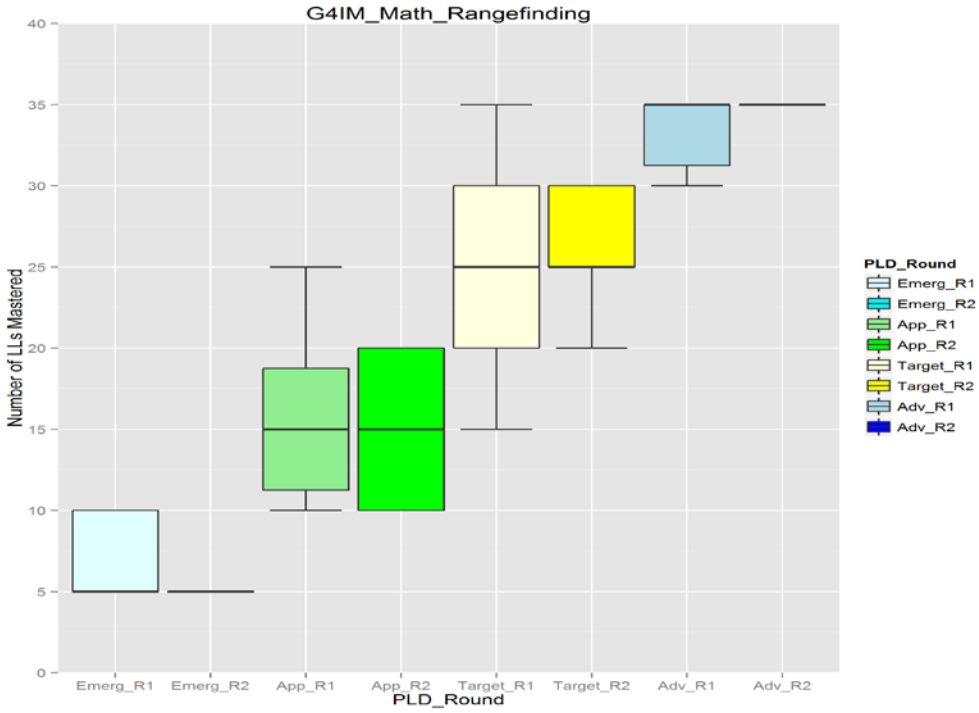


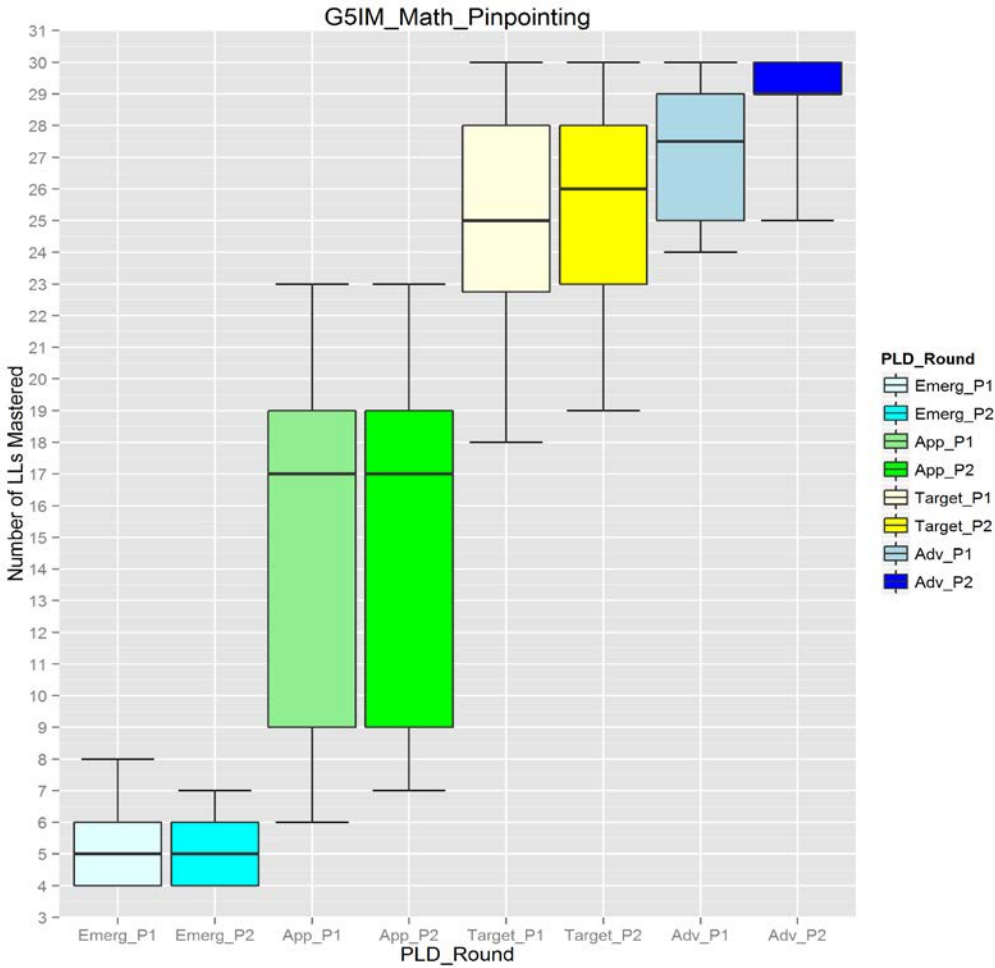
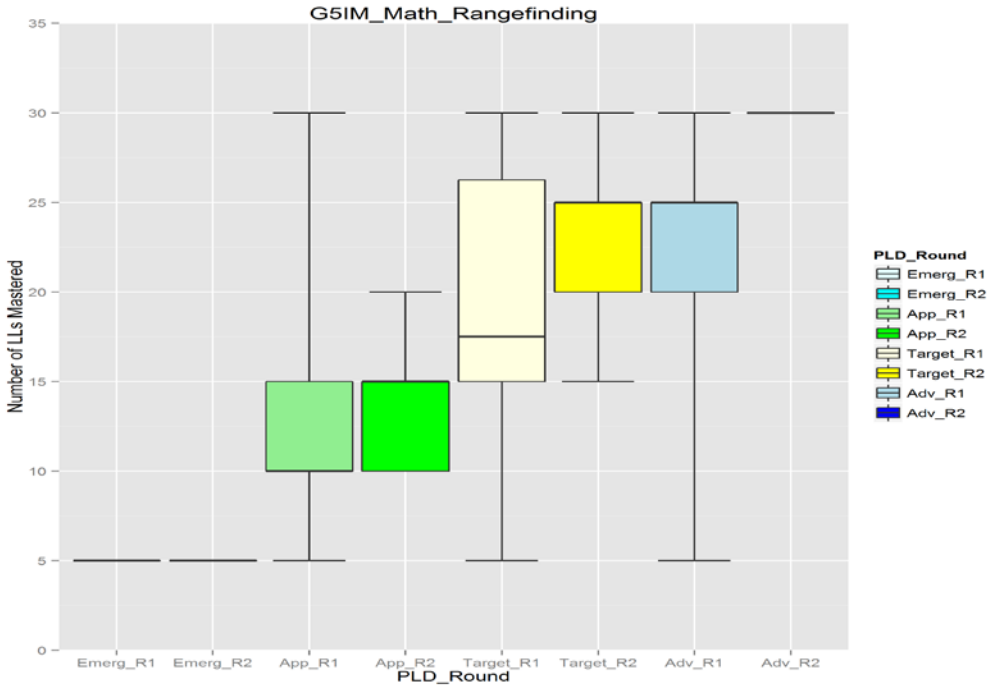


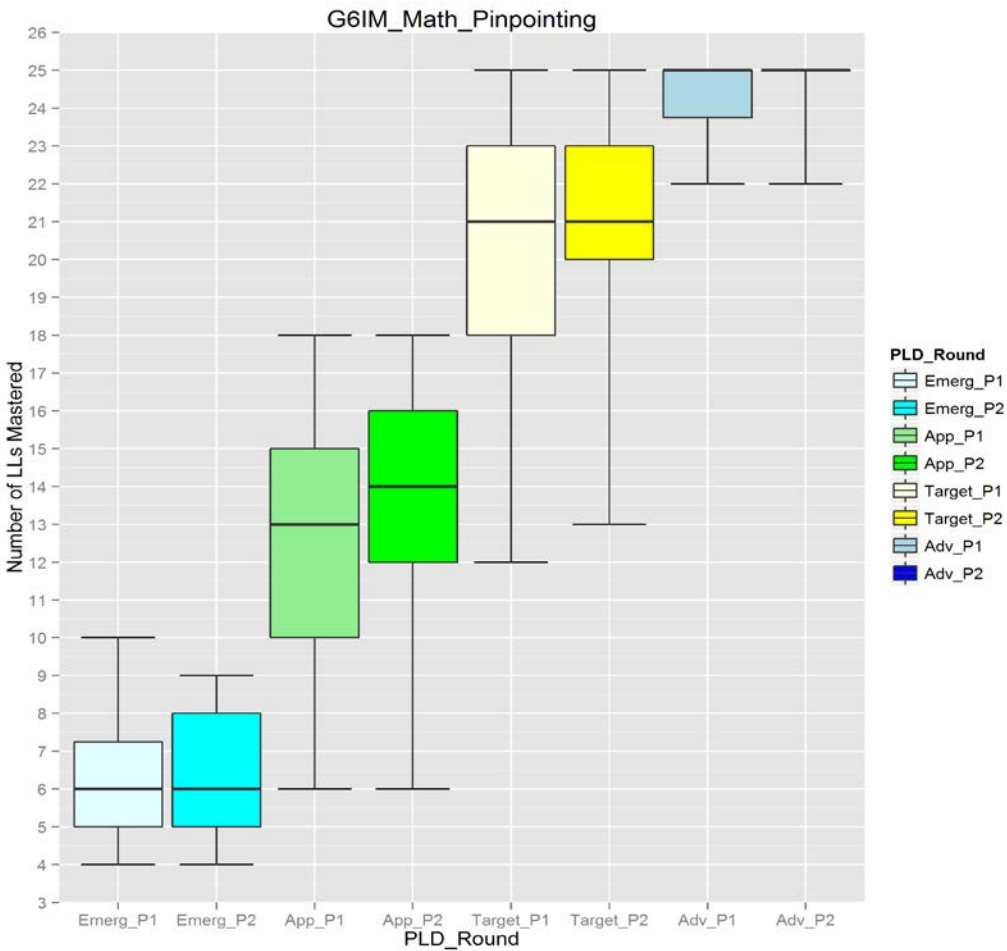
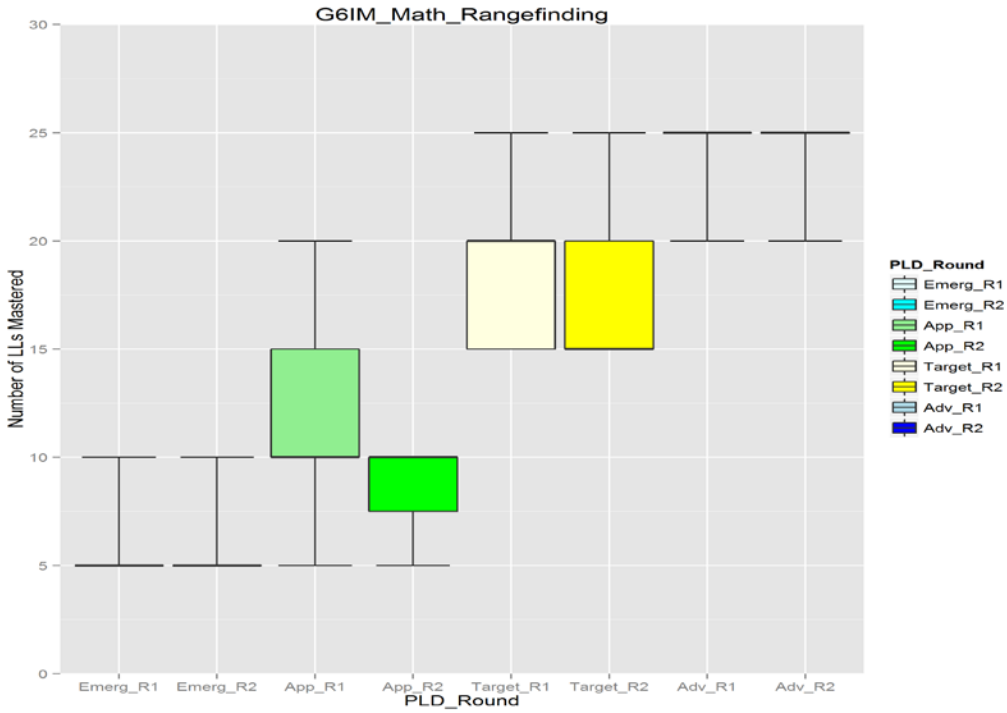


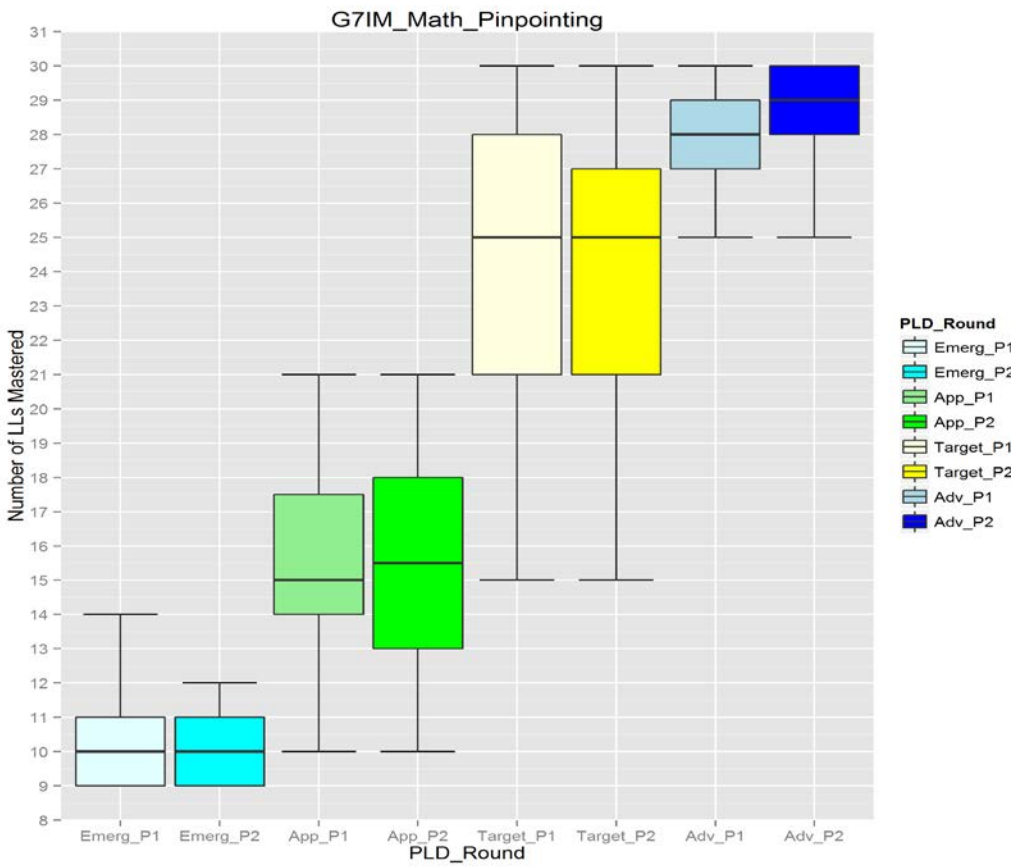
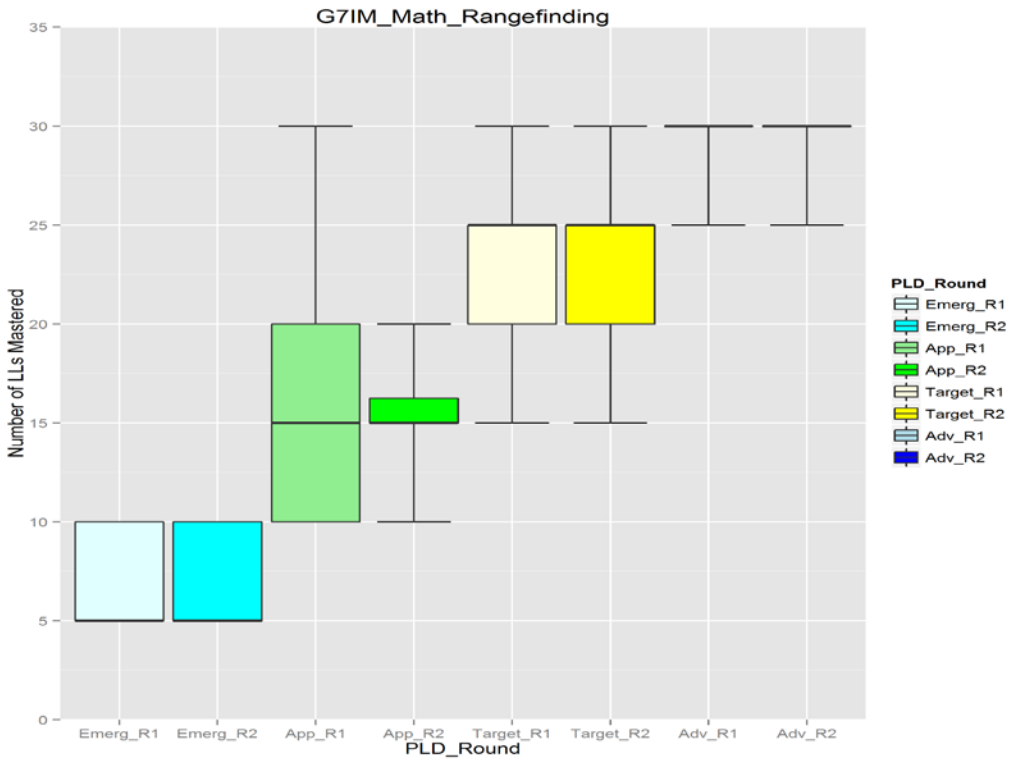


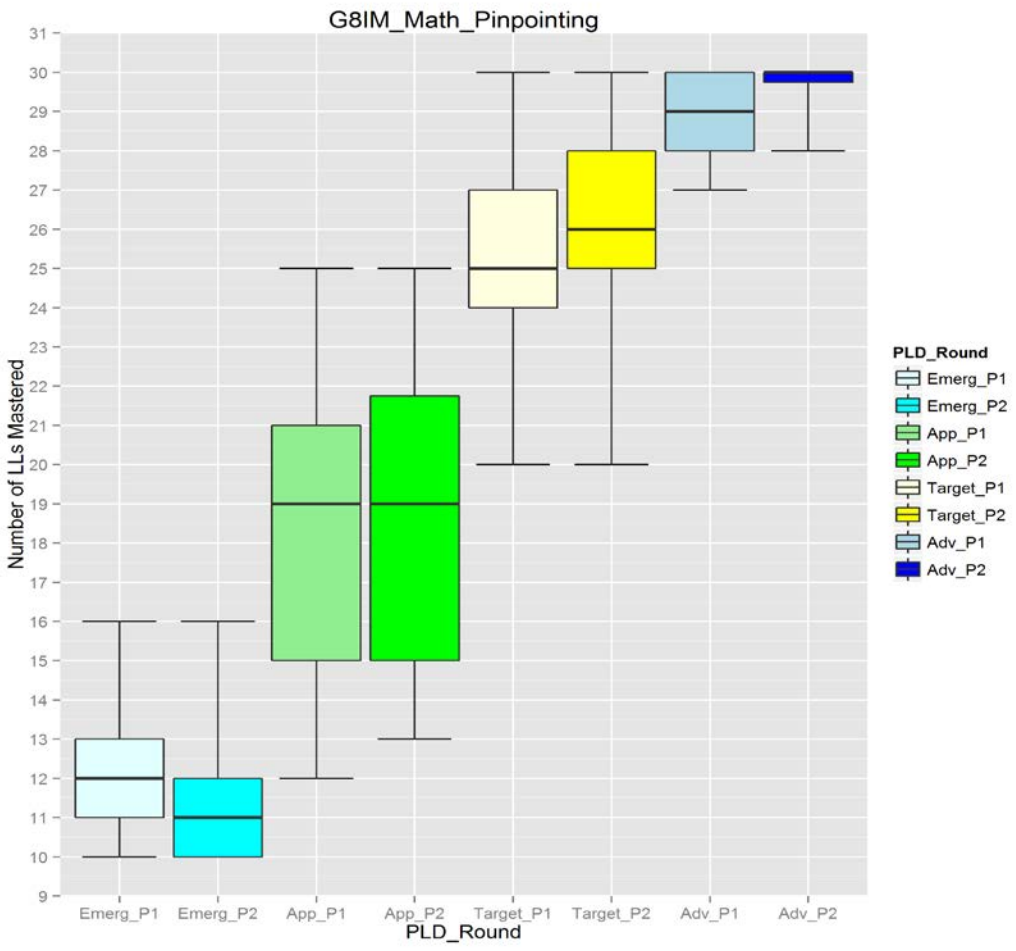
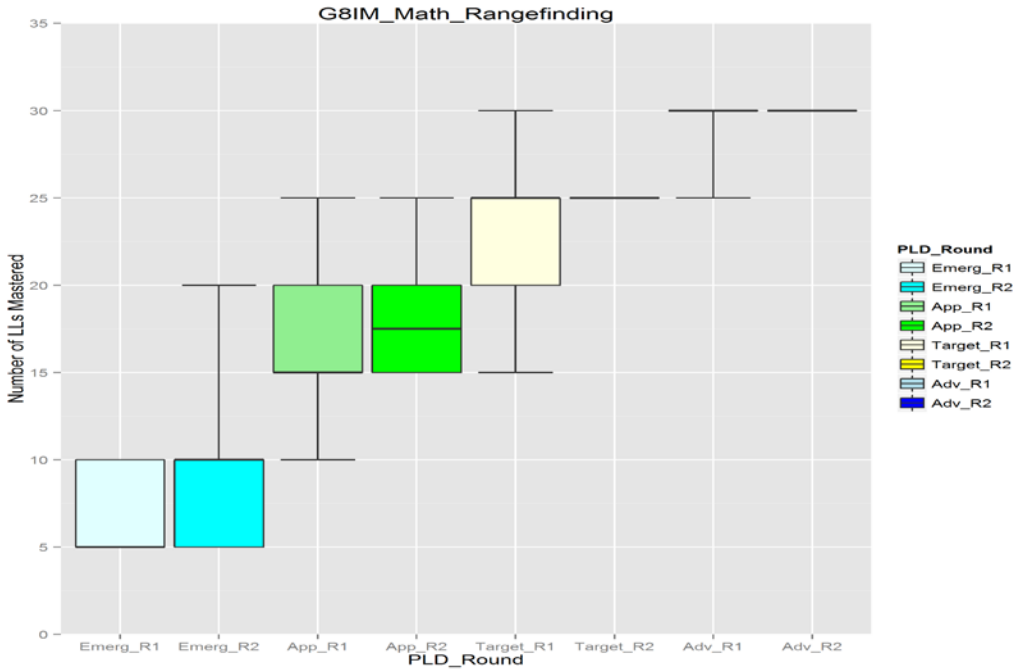


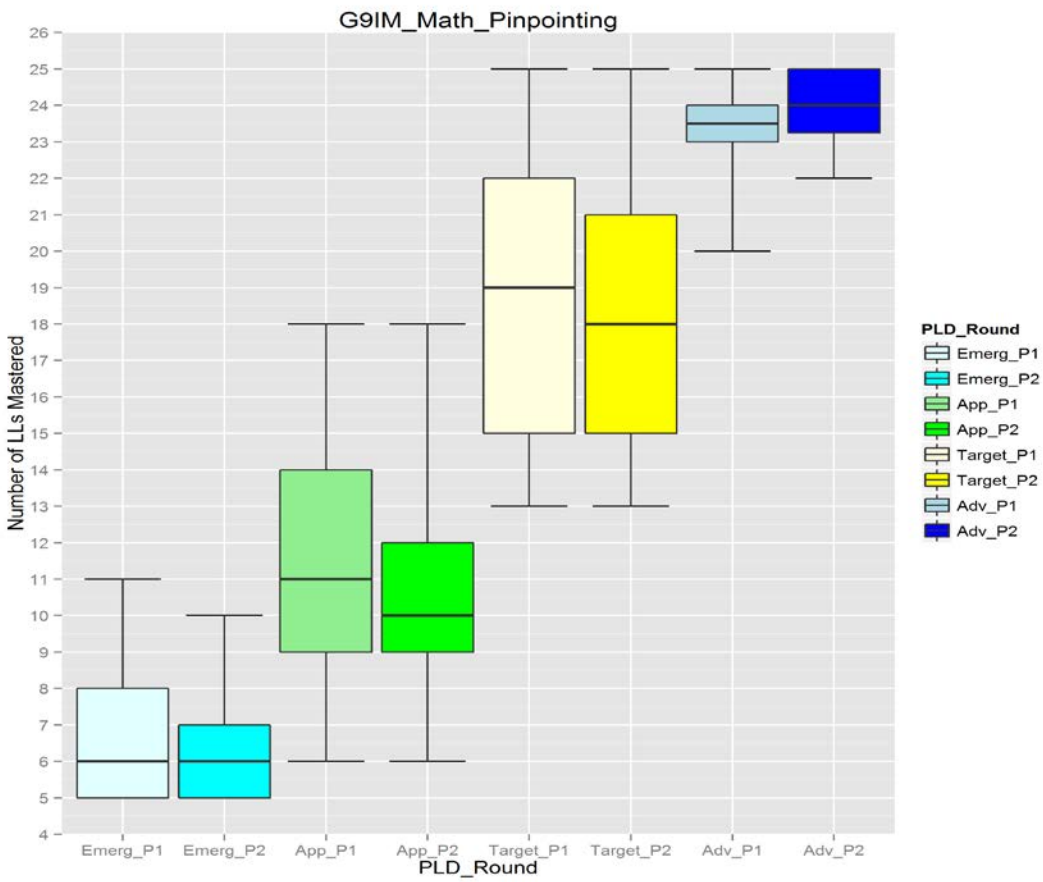
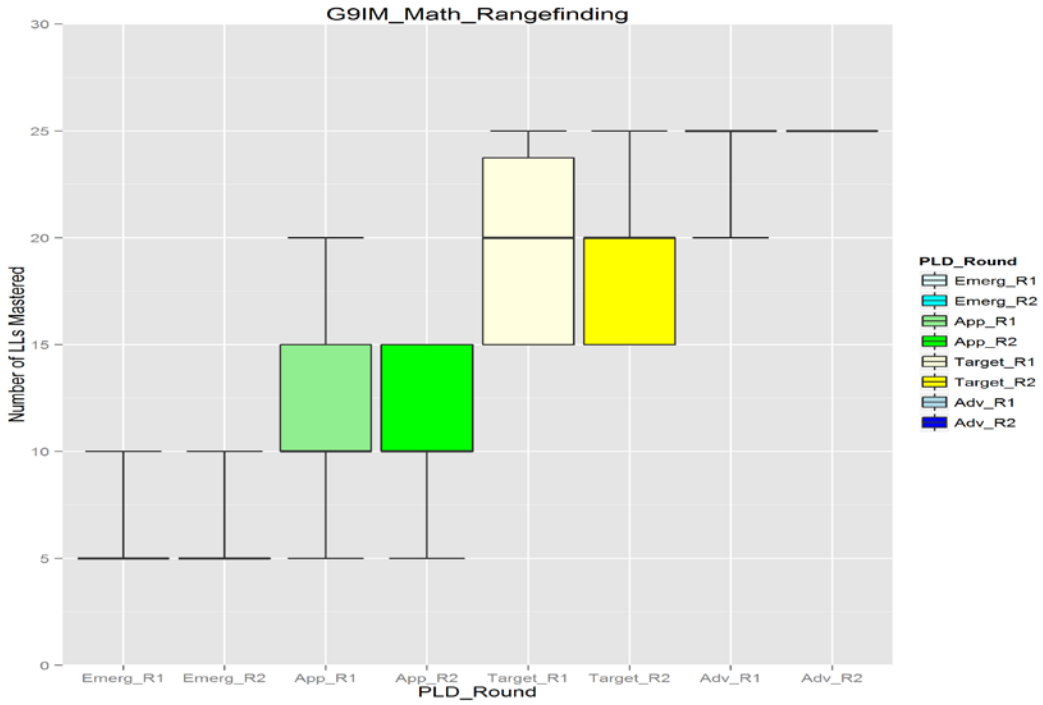


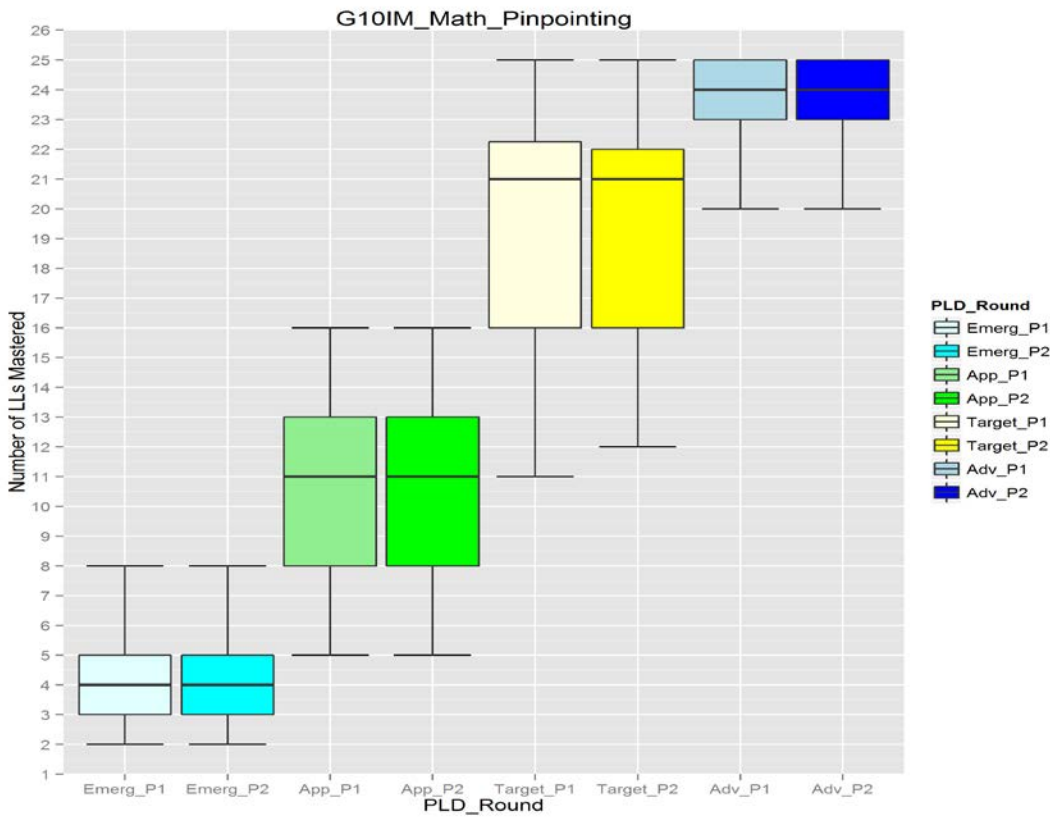
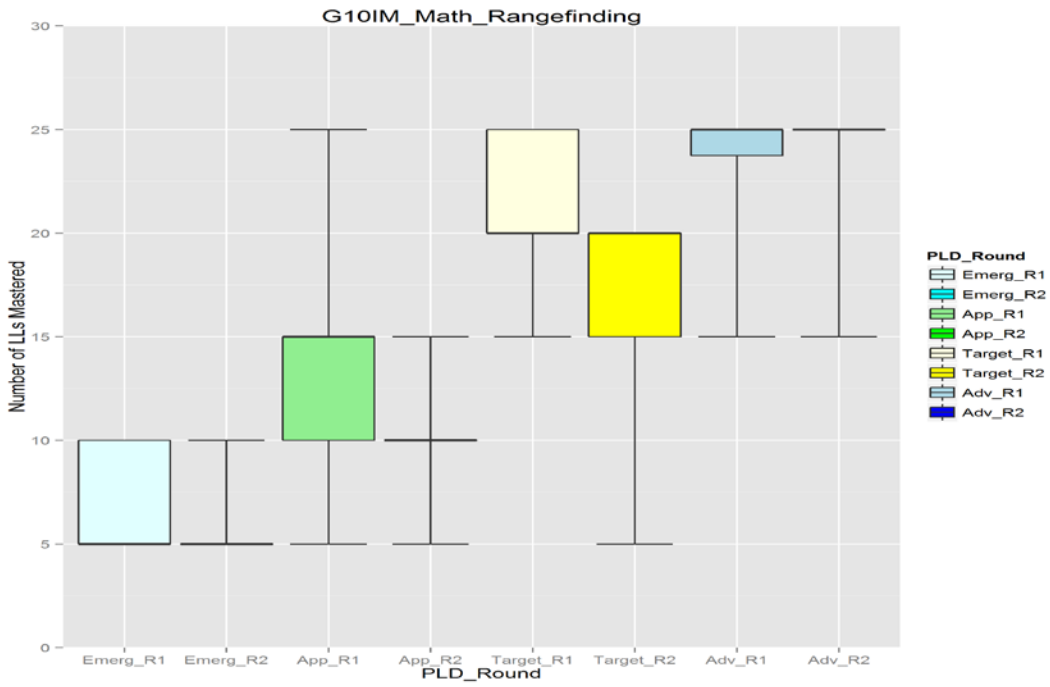


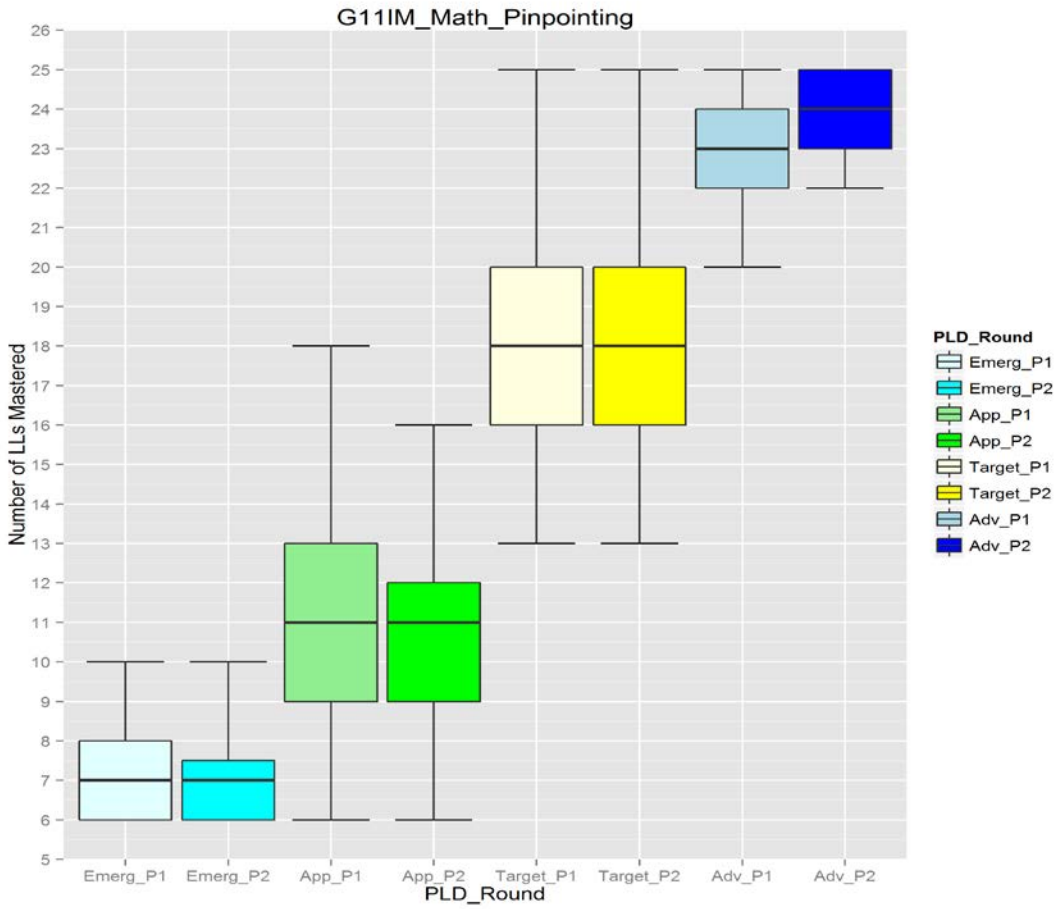
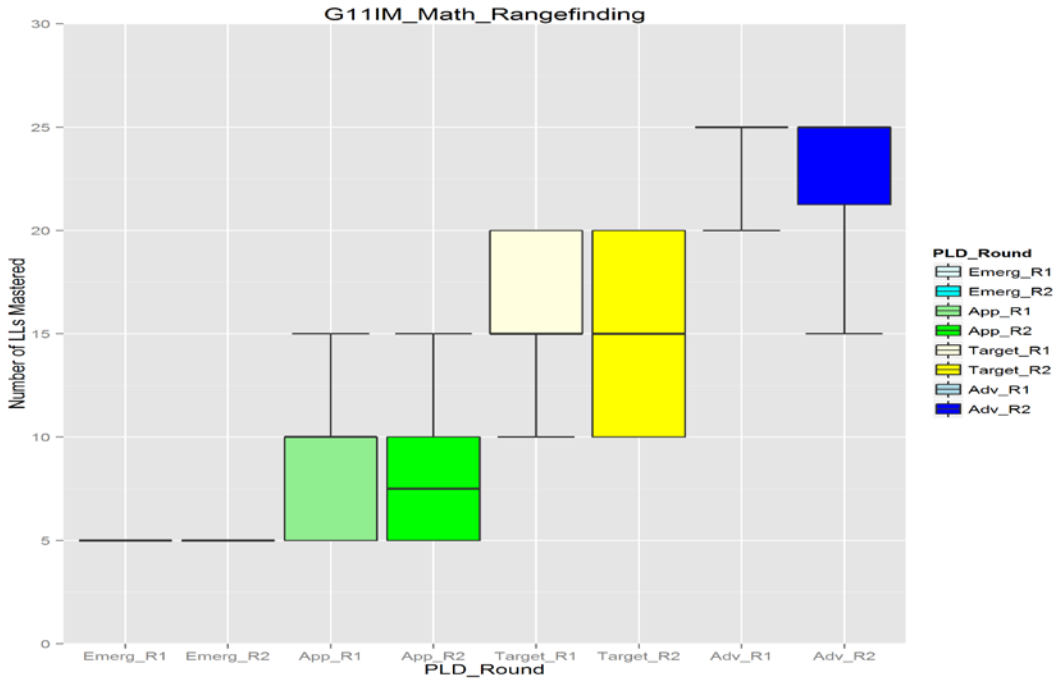












Appendix L: TAC Resolution on DLM Standard Setting

MEMORANDUM

To: DLM Staff and Participating States

From: Edward Roeber, Chair
DLM Technical Advisory Committee

Date: October 21, 2015

Subject: TAC Overview and Commentary on the DLM Standard Setting Process

The DLM Technical Advisory Committee Chair was in attendance during the entire June 15-18, 2015 meeting that was conducted to set standards on the DLM assessments. These observations, summarized below, were shared with both the full TAC as well as the state members at their Governance meeting held in early July 2015, shortly after the standard setting meeting. Additional information on the results of the standard setting process and outcomes were also presented by DLM staff to the TAC and reviewed by the TAC.

Overview of the Standard Setting Process

1. Fourteen panels of special educators were convened by DLM staff to set standards for the year-end (YE) and the integrated model (IM) DLM assessments in grades 3-8 as well as in high school courses (ELA and mathematics).
2. Each panel had between four and nine members on it.
3. Each panel set standards for two courses or three grades. For the grade 3-8 assessments, panels worked on either grades 3-5 or grades 6-8, starting with grade 5 and grade 6, to assure smooth transition between the grade 5 and 6 assessments. The panels worked outwards from those to grades 4 and 3 or grades 7 and 8.
4. A modified body of work standard setting method was used to set standards. Panelists were instructed to use the answer to this question: "What performance level best describes this profile?"
5. The performance levels to be established were emerging (EM), approaching the target (AT), at target (T), and advanced (ADV).
6. Considerable pre-meeting training (2.5 hours) was required of all panelists, who then completed a self-assessment at the conclusion of the training indicating their comfort level with understanding the DLM model and listing any questions that they still had about the model or standard setting procedures.
7. The standard setting event was thoroughly scripted. The process of training the facilitators who led the work at each table included a full-scale tryout of the standard setting process (i.e., actually setting standards based on sets of the materials that actually would be used). This was more than just a review of the agenda or list of activities.

8. Panelists were shown actual student profiles or reports of student results with actual linkage levels (LLs) shown. Panelists were told that cut points would be set by the number of LLs mastered. The potential number of LLs ranged from 40 to 100, depending on the assessment.
9. On student profiles, the LLs for assessments not administered (that were below the linkage levels shown as mastered—lower linkage levels for which students had a high probability of being able to master—were left blank, as were LLs on which students were assessed but did not show mastery (80% or higher performance).
10. The steps in the process of setting standards included two rounds of range-finding, with 20 student profiles that ranged from 5 to 60 or beyond, selected at intervals of 5 points (e.g., 5, 10, 15, etc.). This was followed by two rounds of pinpointing with 21 profiles for each round, which focused on the areas of greatest disagreement within the overall range.
11. In the first round of range-finding, the scores entered served to trigger an indication in a separate table as to whether the level of agreement or disagreement warranted further discussion. Panelists were instructed to focus on these, as well as any other ratings that they wanted to discuss. Once the first round ratings had been discussed, the panelists were instructed to enter their second round ratings. This resulted in a calculation of a suggested cut point, as well as potential high and low points (plus or minus 3 points from the suggested cut points) for use in pinpointing.
12. Pinpointing followed a similar pattern. Two rounds were used. Most of the panels rated the profiles in such a manner that intended cut points emerged by the second round of pinpointing.
13. In a few cases, the pinpointing results (the cuts suggested by the panel) fell outside the original suggested range from range-finding, likely due to panelists rethinking what constituted achieving the higher of the two levels they were rating. This required addition of more profiles for rating in a third round of pinpointing to more accurately determine where the cut points should be set.
14. Panelists' ratings for range-finding and pinpointing were entered into a pre-defined spreadsheet that contained the student profile number and profile scores (that corresponded to paper profiles prepared for each panelist in a group). The spreadsheet was projected on a wall for all panelists to see as ratings were announced.
15. Impact results were not used during standard setting. DLM saw patterns in results that needed to be verified and, in some cases, corrected, before the data was rerun (and re-verified after runs that would take several days). These issues were corrected in reading by the time the panel meeting started but continued in writing.
16. After standards were set, panelists were asked to write a few sentences to describe each of the performance levels as a precursor to grade level PLDs, specifically, the KSAs addressed at each performance level. Panelists were asked to describe students who were solidly at each of two levels, or on the fence between the two levels, for each pair of levels, during this downtime.
17. By the conclusion of the meeting, each panelist was asked to indicate his or her agreement with each cut point set for each assessment. Panelists were also asked to complete an overall survey about the standard setting experience and their satisfaction with the performance levels that were set. The panelists indicated substantial overall support for the performance levels that they set on both the individual cut score surveys and the overall survey.
18. Once performance data could be calculated (after the meeting), the DLM staff reviewed the cut points that were set by the panelists, suggested where smoothing should occur, and then reviewed the proposed changes with the TAC. The TAC first reviewed changes to the cut points before they were presented to the states for their concurrence.

Commentary

1. The actual standard setting event was carefully scripted. The training of the 14 facilitators who led the work at each table included a full-scale tryout of the standard setting process (i.e., actually setting standards based on sets of the materials that would be used at the event). This unusual yet useful step permitted all facilitators to understand the steps in standard setting and permitted all panels to receive the same instructions at each step in the process for each grade/course for which standards were set.
2. There were daily debriefs with the facilitators, which permitted any needed mid-course corrections to be made to the process or instructions. This served to keep the standard setting on schedule.
3. In the student profiles, cells were blank for LLs when there was no evidence of mastery. The lack of evidence could have been due to the student not being tested on that LL or to having been tested and not performed well. Some panelists evaluated those empty cells compared to adjacent mastered LLs and believed the student should have mastered the blank cells. Some of the panelists did not know whether to count the non-mastered LLs or not.
4. The use of logistic regression to identify pinpointing samples mostly worked, but not in every case. In one case, the suggested cut was indeterminate because the panelists all agreed on where the cut point between T and ADV should be set. This is an artifact of the statistical process used.
5. By the end of the second day and through the third, panelists seemed to show greater agreement in range-finding and pinpointing. More agreement, faster discussion, and some resolution of differences occurred. Panelists seemed better able to deal with unusual profiles.
6. The standard setting meeting was carried out well, the staff were helpful to the panelists, and the panelists worked hard to set standards. The panelists strongly supported the levels that they had set and were very supportive of the processes they used to set standards.
7. The independent evaluations of panelist-set cut points, shown in Tables 1-8 attached, summarize the evaluations of panelists. Table 1 through Table 3 indicate panelists' overwhelming support for each performance level that they had established. Table 5 and Table 6 show equally positive overall satisfaction with the standard setting process and all standards that were set. Panelists were very pleased with their work and many wanted to continue to be involved with the DLM project in the future.
8. Following the meeting, the DLM staff took several steps to examine and articulate the results across grade levels. This summary was prepared by the DLM staff for inclusion in the DLM technical report on standard setting:

To mitigate issues related to considering a system with cut points, many testing programs have borrowed strength by considering impact data in the grade at question **and** contiguous grades. The logic is that, under most circumstances (especially when there is no significant shift in demographics), students in bordering grades should have similar distributions within performance levels. Dramatically different distributions are likely due to sampling error and not differences in true cut points.

Therefore, statistical adjustments were made to the panel-recommended cut points in an effort to smooth distributions within the system of cut points being considered. Adjustments were applied to the panel-recommended cut points, separately for the year-end and integrated models.

The following steps were used as part of this method. The method was applied to each subject within each grade level and within each model (YE and IM). No adjustments were made for EOI courses because both the standards assessed and students taking these assessments were assumed to be very different from one course to another.

- A. Create a frequency distribution of the number of linkage levels mastered (from low to high). This step was done separately for each grade and subject. The number of possible linkage levels varies considerably from one grade to another.
- B. Calculate cumulative proportions from low to high.
- C. Perform a probit transformation (z-score associated with the cumulative proportion of students) for each number of linkage levels mastered. Because at the top of the distribution (proportion equal to 1), a finite z-score cannot be calculated. To perform subsequent calculations, z-scores were defaulted to 3.5.
- D. Find the z-score associated with the raw cut point of interest (for example, Approaching-Target) for each grade level.
- E. Create a weighted rolling average of z-scores for the cut-point of interest, using a weight of 0.4 for the grade of interest, 0.2 for contiguous grades, and 0.1 for the next grades. At the end (grades 3, 4, 11, 12), there cannot be a symmetric set of five grade levels involved in the rolling average. For ELA, in the integrated model, the end of the grade continuum was 9-10 and 11-12.
- F. Using the table of probit-transformed cumulative proportions, look up the raw number of linkage levels mastered for which the z-score is closest to the weighted rolling average of z-scores.

Note, an alternative approach might have been to look up the first z-score that exceeded the rolling weighted average z-score. This might have been seen more in keeping with a definition of a cut point being the lowest score that puts a student in the higher category, however doing so would systematically decrease the proportion of students in the higher category over the system of cut-points.

9. Without impact data, it was not possible to look at the consistency of the cut points across grades within a content area and within a DLM model. This was noted as an activity that would occur after the standard setting meeting using the state partners as the policy review committee. This represents a missed opportunity since the strategy of starting with two panels working first on adjacent grades, checking in with each other to look at cross-grade consistency, and then working outwards from there could have provided a means so that post-meeting articulation or smoothing activities might not have been necessary.

Conclusion

At the October 1, 2015, DLM TAC meeting, the TAC's discussion indicated support for adequacy of the procedures, quality of panelist judgments, and overall adherence of the standard setting process and outcomes to professional standards. The DLM TAC recommended its support for the standard setting process as articulated by the DLM staff and approved by the project's governance states. This occurred after the review of the process used to set standards, the review of the data from panelists about agreement with each performance level that had been set, as well as evidence panelist satisfaction on the overall standard setting process, along with the steps taken by the DLM staff to articulate standards across grades within each content area.

On October 21, the DLM TAC took formal action on this report of standard setting. Phoebe Winter moved acceptance of the report on standard setting, as well as the standard setting process as articulated by DLM staff. Jamal Abede seconded the motion. The motion was adopted unanimously by the DLM TAC.

Independent Evaluations of Panel-Recommended Cut Points

A short questionnaire was created to evaluate panelists' comfort with the final panel-recommended cut points. Once pinpointing was complete for the grade level, questionnaires were distributed to each panelist. The panelists each rated their comfort level with the three cut points their group set for each grade or course. Panelists responded with whether or not they agreed with the group-determined cut point (yes or no) and indicated what they would independently suggest be applied for each cut point. Table 1 summarizes the panelist responses to this questionnaire. Note that the percent included in the table is based on all three cut points; each panelist rated their comfort three times: once for the Emerging/Approaching cut, once for the Approaching/Target cut, and once for the Target/Advanced cut.

Table 1: Panelist Comfort with Group-Recommended IM Cut Points

Content Area	Number of Panelists	% Cuts Panelists Comfortable With
ELA		
3	8	100.0
4	8	100.0
5	8	79.2
6	7	95.2
7	7	100.0
8	7	100.0
9–10	7	95.2
11–12	7	100.0
Math		
3	7	100.0
4	7	100.0
5	7	100.0
6	8	100.0
7	8	100.0
8	8	100.0
9	8	100.0
10	8	100.0
11	8	91.7

Across all panelists, panels, grades/courses, and cut points ($N=861$), 95.9% of panelists ($n = 826$) indicated that they were comfortable with the group-recommended cut point. Only 4.1% of responses ($n = 35$) indicated a discomfort with a group-recommended cut.

Complete panelist agreement with the recommended cut point was found in 99 out of 120 cuts (82.5%) across models and subjects/grades/courses. Complete independent panelist comfort with all three recommended cut points was found for 25 out of 40 cut point panels (62.5%). Most recommendations for a change to the cut point were for just one of the three cut points for a given panel, and most often, the recommended changes differed from the initial recommendation by only a single point.

To further evaluate panelist comfort with the recommended cut points, the median panelist-recommended cut point was compared to the group-recommended cut point. In all instances but two, the median and group-recommended cut points were the same. Table 2 summarizes the two instances where the median differed from the group-recommended value. In both instances, the median panelist recommendation was lower than the group-recommended cut point for the cut between the **approaching** and **target** performance levels.

Table 2: Panels with Median Independent Cut Point Different From Group-Recommended Cut Point

Content Area/Model	Grade	Performance Cut Point	Group Recommended Cut Point	Median Panelist Independent Cut Point
ELA IM	5	AP/T	26	23

Panelist Evaluations of the Meeting

Panelists completed panel-meeting evaluations at the conclusion of the standard setting meeting. The evaluation included self-evaluation of readiness to rate and understanding of the tasks, plus evaluation of outcomes. Panelists rated their responses to the 27 questions on a Likert scale choosing either “Strongly Disagree” (SD), “Disagree” (D), “Agree” (A), or “Strongly Agree” (SA). For the last three questions, “Not applicable” was an additional option. A summary of responses is presented in Table 3.

Table 3: Percentages of IM Panelist Responses to Evaluation Items

Question	SD	D	A	SA
1. The overall goals of the standard setting panel meeting were clear.	0	0	27	73
2. The panel meeting was well organized.	0	0	24	76
3. The training and practice exercises provided the information I needed to complete my tasks.	0	7	40	53
4. It was clear what knowledge, skill, or ability a student would need to demonstrate to achieve a certain profile.	0	2	57	41
5. I considered the performance level descriptors when I rated each profile.	0	0	32	68
6. I considered the assessment items when I rated each profile.	0	0	44	56
7. I considered the other panelists' opinions when I rated each profile.	0	2	36	61
8. I considered my experience in the field when I rated each profile.	0	0	27	73
9. I understood how to rate each profile.	0	0	47	53
10. I had enough time to complete the tasks.	0	0	24	76
11. I felt confident when rating the profiles.	0	2	47	51
12. The procedure for recommending cut points was free from bias.	0	0	40	60
13. Overall, I was satisfied with the ratings made by panelists in my group.	0	0	20	80
14. I would defend the group's At Target decisions against criticism that they are too high.	0	0	49	51
15. I would defend the group's At Target decisions against criticism that they are too low.	0	0	42	58
16. I would defend the group's Advanced decisions against criticism that they are too high.	0	0	47	53
17. I would defend the group's Advanced decisions against criticism that they are too low.	0	0	42	58
18. I would defend the group's Approaching Target decisions against criticism that they are too high.	0	0	50	50
19. I would defend the group's Approaching Target decisions against criticism that they are too low.	0	0	42	58
20. I am confident that the meeting produced valid cut score recommendations.	0	0	36	64
21. Overall, I believe my opinions were considered and valued by the group.	0	0	20	80
22. Overall, my group's discussions were open and honest.	0	0	13	87
23. Participating in the process increased my understanding of DLM.	0	0	4	96
24. Overall, I valued the panel meeting as a professional development experience.	0	0	2	98
25. This experience will help me plan and provide instruction for my students with significant cognitive disabilities.	0	0	17	83
26. This experience will help me use DLM more effectively.	0	0	16	84