ACCESS for ELLs 2.0 Assessment Proficiency Level Scores

Standard Setting Project Report

H. Gary Cook, Ph.D.,

David MacGregor, Ph.D.

WIDA Research (WIDA) and the Center for Applied Linguistics (CAL)

Contents

Executive Summary **1**
Purpose of the Standard Setting Study 2
Process 3
Approval of the study design 3
Design of the standard setting study 3
Expert review summary 3
Panelist recruitment, selection and participation 3
Approval of panelist recommendations 4
Communication campaign 4
Section I: Phase 1 Report **4**
Introduction 4
Methods 5
Standard setting methods 5
Panelists 5
Materials 6
Process 7
Results 11
Domain results 11
Vertical articulation 15
Training evaluation summary 18
Section II: Phase 2 Report **20**
Introduction 20
Developing proposals for cut scores 20
Choosing items for Listening and Reading Ordered Item Booklets 21
Choosing Speaking and Writing portfolios 22
Methods 23
Standard setting methods 23
Panelists 23
Materials 23
Process 24
Results 26
Domain results 26
Smoothing procedures 31
Creation of composites 38
Training evaluation summary 40
Study evaluation summary **41**

Tables

Table 1. *Phase 1 Standards Setting Groups by Grade* 7
Table 2. *Listening Round 3 Median Cut Scores by Group* 11
Table 3. *Reading Round 3 Median Cut Scores by Group* 12
Table 4. *Speaking Round 2 or 3 Median Cut Scores by Group* 13
Table 5. *Writing Round 3 Median Cut Scores by Group* *14*
Table 6. *Smoothed Phase 1 Recommended Cut Scores* 16
Table 7. *Phase 1 Training Evaluation Responses (55 Responses)* 19
Table 8. *Proposed Cut Scores: Listening* 20
Table 9. *Proposed Cut Scores: Reading* 21
Table 10. *Proposed Cut Scores: Writing* 21
Table 11. *Number of Items Selected per Cluster* 22
Table 12. *Panelists' Initial Speaking Cut Scores across Grades (Unsmoothed)* 27
Table 13. *Panelists' Initial Writing Cut Scores across Grades (Unsmoothed)* 28
Table 14. *Panelists' Initial Listening Cut Scores across Grades (Unsmoothed)* 30
Table 15 *Panelists' Initial Reading Cut Scores across Grades (Unsmoothed)* 31
Table 16. *Quadratic Function Fitted Cut Scores for Speaking across Grades (Final Recommended Cuts)* 32
Table 17. *Quadratic Function Fitted Cut Scores for Writing across Grades (Final Recommended Cuts)* 34
Table 18. *Interaction Model for Listening* 35
Table 19. *Interaction Model for Reading (Final Recommended Cuts)* 36
Table 20. *Amended Interaction Model for Listening (Final Recommended Cuts)* 37
Table 21. *Final Cut Scores: Oral* 38
Table 22. *Final Cut Scores: Literacy* 39
Table 23. *Final Cut Scores: Comprehension* 39
Table 24. *Final Cut Scores: Overall* 40
Table 25. *Phase 2 Training Evaluation Responses (53 Responses)* 40
Table B1. *Panelist Participation by WIDA State* *45*

Table B2. *Panelist Familiarity with the WIDA ELD Standards, the Six WIDA Proficiency Levels, and the WIDA ACCESS for ELLs Assessment* 46
Table B3. *Panelist Teaching Certification or Specialization, Additional Certifications, and Highest Degree Obtained* 47
Table B4. *Panelists who Provided Instruction in These Language Services in Their Current Position or Job Title* 48
Table B5. *Panelist General Work Experience* 48
Table B6. *Panelist Demographic Information* 49
Table F1. *Number of Years Currently Teaching Grades 1–12* 57
Table F2. *Distribution of Panelists to Groups and Tables in the Phase 1 Standard Setting Meeting* 58
Table F3. *Distribution of Panelists to Groups and Tables in the Phase 2 Standard Setting Meeting* 59
Table U1. *Phase 1 and Phase 2 Panelist Average Opinion Rating on the Adequacy of the Seven Components of the Standard Setting Study for Reading, Listening, Writing, and Speaking (4 = Excellent; 3 = Good, 2 = Fair, 1 = Poor)* 120
Table U2. *Panelist Average Rating of Their Confidence in the Ability of the Group's Recommended Cut Scores to Adequately and Appropriately Classify Students as Borderline English Proficient in Reading, Listening, Writing, and Speaking (5 = Very High , 4 = High; 3 = Medium, 2 = Low, 1 = Very Low)* 123

Figures

Figure 1. Grade 1 Listening test item from an Ordered Item Booklet in the web-based display. 9
Figure 2. Grades 4-5 Writing item from a Student Portfolio in the web-based display 10
Figure 3. Panelists' Round 3 Listening cut scores across raters by grade. 12
Figure 4. Panelists' Round 3 Reading cut scores across raters by grade 13
Figure 5. Panelists' Round 3 Speaking cut scores across raters by grade. 14
Figure 6. Panelists' Round 3 Writing cut scores across raters by grade 15
Figure 7. Listening Phase 1 smoothed and unsmoothed cut scores by grade 16
Figure 8. Reading Phase 1 smoothed and unsmoothed cut scores by grade 17
Figure 9. Speaking Phase 1 smoothed and unsmoothed cut scores by grade 17
Figure 10. Speaking Phase 1 smoothed and unsmoothed cut scores by grade 18
Figure 11. Panelists' Speaking cut scores across grades 27
Figure 12. Panelists' Writing cut scores across grades 28
Figure 13. Panelists' Listening cut scores across grades 29
Figure 14 Panelists' Reading cut scores across grades 30
Figure 15. Quadratic function fitted curves for Speaking across grades 32
Figure 16. Quadratic function fitted curves for Writing for Grades 1–3 (paper). 33
Figure 17. Quadratic function fitted curves for Writing for Grades 4–12 (paper and online). 34
Figure 18. Interaction model fitted curves, with panelist values (•) and exit cuts (X) for Listening 35
Figure 19. Interaction model fitted curves, with panelist values (•) and exit cuts (X) for Reading 36
Figure 20. Amended interaction model fitted curves (for Cut 2.0, Grade 1 and Grades 10–12 cuts are set to panelist means), with panelist values (•) and exit cuts (X) for Listening 37

Appendices

Appendix A: Review of the Plan for the Standard Setting Study by Dr. Marianne Perie 42

Appendix B: Recruitment and Selection of Panelists 45

Appendix C: Email Sent to WIDA Member State Education Agencies requesting Panelist Nominations 50

Appendix D: ACCESS 2.0 Standard Setting Study Nominee Qualifications Survey 51

Appendix E: Standard Setting Communication Campaign Tools and Resources 56

Appendix F: Assignment of Panelists to Groups and Tables 57

Appendix G: ACCESS 2.0 Standards Setting Phase 1 Meeting Agenda 60

Appendix H: ACCESS 2.0 Standards Setting Phase 1 Table Leader Training Agenda 61

Appendix I: ACCESS 2.0 Standards Setting Phase 2 Meeting Agenda 62

Appendix J: WIDA ACCESS for ELLS Test Nondisclosure Agreement (NDA) 63

Appendix K: Phase 1 Panelist Training Presentation 64

Appendix L: Phase 1 Group Training Presentation 72

Appendix M: Phase 2 Training Presentation and Handout 79

Appendix N: Phase 1 Impact Information 90

Appendix O: Phase 2 Data Collection Sheets 92

Appendix P: ACCESS 2.0 Standard Setting Study Training Evaluation Form 104

Appendix Q: ACCESS 2.0 Standard Setting Study Panelist Background Form 105

Appendix R: ACCESS 2.0 Standard Setting Study Evaluation Form: All Domains 108

Appendix S: Phase 1 Panelist Rating Form 116

Appendix T: WIDA Performance Level Definitions for Levels 4 Through 6 117

Appendix U: Study Evaluation Summary 119

Executive Summary

In July and August of 2016, WIDA conducted a standard setting study to reexamine the ACCESS for ELLs 2.0 assessment (ACCESS 2.0) proficiency level scores. The following factors motivated WIDA to conduct this study: Migrating from a paper-and-pencil to an online assessment, employing a new centrally scored, revised speaking assessment, and adapting to the influence of college and career ready standards.

The standard setting study occurred in two phases. The first phase, which was led by WIDA, identified scores that panelists felt represented the marginally English proficient English learner (EL) student on the ACCESS 2.0 Listening, Reading, Speaking, and Writing domain tests. Phase 1 occurred in July of 2016. Phase 1 information was used to set up Phase 2, which was led by the Center for Applied Linguistics (CAL). The goal in Phase 2 was to determine where to set cut scores between the six WIDA proficiency levels, as described by the WIDA English Language Development (ELD) Standards for Grades 1–12 in each of the four domains. After each phase, recommended cut scores were smoothed to assure appropriate vertical articulation of cut scores across grades. Information from individual domain scores was used to create composite cut scores.

Phase 1 involved two activities. The first was to describe and document the Listening, Speaking, Reading, and Writing characteristics that represented a minimally competent English proficient, or *borderline*, EL at each grade. Each group displayed the agreed-upon description of this borderline student so that all panelists could refer to it.

The second activity identified domain scale scores that represented the minimally competent student performance at each grade and for each domain test. For the Listening and Reading tests, a modified Bookmarking method was adopted using an Ordered Item Booklet (OIB) online program. For the Speaking and Writing tests, a modified Body of Work method was used, again with online materials. In this Body of Work method, panelists reviewed portfolios of students' work (consisting of all of the student's responses to the Speaking or Writing prompts) which were ordered from the highest to the lowest scores. Judgments were then made about where a minimally competent student's performance would lie. For each domain test, there were three rounds. In Round 1, panelists assigned ratings based on the group's description of the borderline student. In Round 2, panelists could revisit their initial ratings after they discussed Round 1 ratings and saw the group's median scores from Round 1. Prior to the last round, impact information (i.e., the proportion of students above and below the median cut score) based on the group's Round 2 median scores were shared. Panelists then completed Round 3. The Round 3 ratings were used as the group's final recommendation. For all rounds, it was made clear that panelists did not need to change their ratings, only that their final ratings be informed by conversations, group median ratings, and impact information. Groups 2 and 3 did not complete Round 3 of the Speaking domain. Impact information was shared in Round 2 for those groups.

Phase 2 focused on setting cut scores for the six proficiency levels described by the WIDA ELD Standards in each of four domains (Listening, Reading, Writing and Speaking) for Grades 1–12. The methods and procedures used for this phase followed closely those used in Phase 1, with some differences that arose from the different goals of the two phases. Phase 2 only had two rounds. To start Round 1 of Phase 2, initial recommended cut scores, which were calculated prior to convening Phase 2, were provided for all domains except Speaking. To calculate the proposed cut scores, CAL examined smoothed recommended cut scores from Phase 1, reviewed the previous ACCESS 1.0 cut score information, and created initial recommended cut score values. During Phase 2, it was made clear to panelists that these initial recommendations were only a starting point. Panelists were free to make changes to the initial recommended values, and in many cases, they did. After Round 2, impact data was shared. For Phase 2, Round 2 ratings were used as panelists' final recommended cut scores. Following this, CAL used a smoothing procedure on the Phase 2 final recommendations to create the recommended domain, grade, and proficiency level cut scores.

In both Phase 1 and Phase 2, standard setting activities occurred in Grades 1–12. Kindergarten was not part of the standard setting study. To obtain Kindergarten cut scores, researchers at WIDA fit a regression model based on Grades 1, 2, and 3 cut scores. They then interpolated results to obtain preliminary cut scores. Once this was completed, CAL examined the preliminary cuts to see if actual Kindergarten student performances were consistent with identified cut scores. After deliberations between CAL and WIDA, final recommended Kindergarten cut scores were created.

On September 22, 2016, a memo was sent to WIDA states presenting the recommended cut scores for Grades 1–12. Recommended cut scores for Kindergarten were not provided since analysis was underway at that time. On September 27, 2016, a webinar session was held to give WIDA member states the opportunity to ask questions and/or seek clarification about the standard setting recommendations. Most comments at this webinar related to the impact of the new cut scores on accountability and EL reclassification criteria. A memo was sent to the WIDA Executive Committee outlining the procedures used to establish Kindergarten cut scores, and on October 11, 2016, a webinar session was held for the Executive Committee approval of the recommended cut scores, including Kindergarten. The cut scores were approved at that meeting. Subsequently, approved cut score values were provided to the Data Recognition Corporation (DRC) to be used for the 2016–2017 ACCESS 2.0 score reports. States were also given the opportunity to receive an updated state student response (SSR) file with post-standard setting scale scores and proficiency level values.

Purpose of the Standard Setting Study

The confluence of many factors, such as migrating from a paper-and-pencil to an online assessment, employing a new centrally scored revised speaking assessment, and adapting to the influence of college and career ready standards, motivated WIDA to re-examine its language proficiency levels as expressed by ACCESS 2.0. To support this re-examination, a standard setting study was conducted, which occurred in two phases. The first phase (led by

WIDA) sought to identify exit criteria scores on ACCESS 2.0. The second phase (led by CAL) determined where to set cut scores between the six proficiency levels described by the WIDA ELD Standards for Grades 1–12.

Process

In this section we focus on processes common to both phases of the standard setting study. More details are provided for each phase in their respective sections below.

Approval of the study design. Beginning in 2011 member states were made aware of the need for a standard setting study as part of the development and implementation of the new online assessment and the study was included in the ASSETS[1] grant. In June of 2015 the WIDA Board delayed the standard setting study by one year so that the study wouldn't coincide with the ACCESS 2.0 launch in 2015–2016. Members did not want to delay score reports that first year, or manage changes in the cut scores the same year that they were navigating changes in the state content test scores. In December of 2015 the study plan was presented to the WIDA Technical Advisory Committee and the WIDA Executive Committee, and recruitment of panelists began the following month. The Phase 1 and Phase 2 standard setting meetings were held in July and August of 2016.

Design of the standard setting study. Both phases of the standard setting study used the same methods: A modified Bookmarking procedure for the receptive domains (Listening and Reading) and a modified Body of Work procedure for the productive domains (Speaking and Writing). The items and portfolios used for Phase 1 were a subset of those used for Phase 2. The Phase 1 materials focused on the higher end of the English language proficiency scale, while those used for Phase 2 spanned the scope of the scale. Results from Phase 1 were used to inform the CAL proposed cut scores that served as a starting point for Phase 2.

Expert review summary. An outside expert, Dr. Marianne Perie from the University of Kansas, reviewed the plan for the standard setting study in March of 2016. Several minor recommendations were made, many of which were adopted for the standards setting study. A majority of the comments made by Dr. Perie dealt with the timing of rounds and sharing information between and across groups. This review is provided in Appendix A.

Panelist recruitment, selection, and participation. Panelist recruitment and selection took place between January and August of 2016. Panelists were nominated by the state education agencies of WIDA member states. Approximately 60 panelists were chosen for each phase, divided into four groups of up to 15 panelists each. For Phase 1, the panelists represented a broad range of educators, including EL and content teachers, administrators, and district and state policymakers. For Phase 2, nearly all panelists were EL and content teachers. Detailed information regarding panelist selection is available in Appendices B–D.

---

[1] A federally funded Enhanced Assessment Grant awarded to WIDA to develop a "next-generation," online English language proficiency assessment based on the WIDA Standards.

Approval of panelist recommendations. In September of 2016 a memorandum was sent to WIDA states presenting the recommended cut scores for Grades 1–12 and a webinar was held to give WIDA member states the opportunity to ask questions and/or seek clarification about the standard setting recommendations. A memo was sent to the WIDA Executive Committee outlining the procedures used to establish Kindergarten cuts, and on October 11, 2016, a webinar was held for the Executive Committee approval of the recommended cut scores, including Kindergarten. The cut scores were approved at that meeting.

Communication campaign. A communication campaign was established because study leadership recognized that the reasoning behind, and the process for, updating the scoring scale would need to be clearly communicated to SEAs and LEAs. On September 23, 2016, a memo was sent to WIDA states presenting the recommended cut scores for Grades 1–12. Recommended cut scores for Kindergarten were not provided since analysis was underway at that time. On September 27, 2016, a webinar session was held to give WIDA member states the opportunity to ask questions and/or seek clarification about the standard setting recommendations. Most comments during this webinar related to the impact of the new cut scores on accountability and EL reclassification criteria. Following approval by the WIDA Executive Committee on October 11, 2016, approved cut score values were provided to DRC to be used for the 2016–2017 ACCESS 2.0 score reports. States were also given the opportunity to receive an updated SSR file with post-standard setting scale scores and proficiency level values.

In addition, the campaign also facilitated communication tools and resources for the WIDA Professional Learning and Client Services departments, for policy makers, for SEAs and LEAs, educators, parents, and families. A complete list of communication tools and resources is available in Appendix E.

The remainder of this report is divided into two sections. Section I describes the design, conduct, and results of Phase 1, while Section II focuses on Phase 2.

Section I: Phase 1 Report

Introduction

In July and August of 2016, WIDA conducted a standard setting study to reexamine the ACCESS for ELLs 2.0 assessment (ACCESS 2.0) proficiency level scores. Many factors motivated the WIDA to conduct a new standard setting study, including migrating from a paper-and-pencil to an online assessment, employing a new centrally scored, revised speaking assessment, and adapting to the influence of college and career ready standards.

The standard setting study occurred in two phases. The first phase (led by WIDA) identified scores that represented the marginally English proficient EL student on the ACCESS 2.0 Listening, Reading, Speaking and Writing domain tests. Using information from the first phase, a second phase study (led by CAL) determined where to set cut scores between the six

WIDA proficiency levels, as described by the WIDA English Language Development Standards for Grades 1–12. This section of the report describes activities associated with Phase 1 of the standard setting process.

The ASSETS grant required WIDA member states to agree upon a "common definition" of an English learner, which, at a minimum, establishes a common English language proficiency level understood as “proficient” in English. Linquanti, Cook, Bailey, and MacDonald (2016)[2] offer guidance to states on how to establish a common English learner definition. In their work, Linquanti et al. outline four stages involved in establishing a common definition: Identifying potential English learners, classifying English learners, establishing an English proficient performance standard, and reclassifying English learners. For the purposes of the ASSETS grant, Phase 1 of the standard setting study supports the third stage in Linquanti et al.’s guidance establishing an English proficient performance standard.

The Phase 1 standard setting study identifies recommended English proficiency levels below which ELs should not be declared proficient. This minimum level is established for the Listening, Speaking, Reading and Writing test scores in Grades 1–12.

Methods

Standard setting methods. For the Writing and Speaking domains, a modified Body of Work method was used. For the Listening and Reading domains, a modified Bookmarking method was used. These methods are described in detail below.

Panelists. The goal was to have 60 panelists participate in each phase meeting, representing each of the WIDA member states. All panelists had to have experience teaching English learners. Following panelist selection, educators were divided into four groups, each including panelists representing five grade levels. Group 1 represented Grade Levels 1–5, Group 2 Grade Levels 4–8, Group 3 Grade Levels 6–10, and Group 4 Grade Levels 8–12. Each group was further divided into three tables of four or five panelists each. The Phase 1 meeting ended up having 59 panelists due to a late cancellation.

For each phase of the project, panelists were assigned to the appropriate group based on their experience teaching the corresponding grade levels and content areas. Panelists were asked in the Standard Setting Study Nominee Qualifications survey (Appendix D) if they had experience working with ELs with disabilities and 83% responded *yes* in Phase 1. In addition, educators with specialization or credentials in special education (five in both Phase 1 and Phase 2) were assigned to each group.

---

[2] Linquanti, R., Cook, H. G., Bailey, A. L., & MacDonald, R. (2016). *Moving toward a more common definition of English learner: Collected guidance for states and multi-state assessment consortia*. Washington, D.C.: Council of Chief State School Officers.

The work of each group was led by a facilitator, a CAL or WIDA staff member with previous experience in standard setting, and each group had a CAL or WIDA staff member serving as a helper. In addition, panelists who were also State Education Agency (SEA) or Local Education Agency (LEA) staff experienced with the study received additional training and served as Table Leaders during Phase 1. For detailed information on how panelists were assigned to groups and tables based on experience and credentials, see Appendix F.

Materials. Phase 1 panelists were provided the following materials. Examples of meeting materials are in Appendices G-T

1. Administrative documentation
   - Phase 1 meeting agenda and Table Leader meeting agenda
   - Non-disclosure agreement form
   - Panelist background questionnaire
   - WIDA ACCESS 2.0 Phase 1 Standard Setting Study Evaluation Form: All Domains
   - Panelist Rating Form
2. Training materials
   - Panelist training presentation handout (consisting of selected slides from the training PowerPoint)
   - Group training presentation
   - Training evaluation forms
3. A Chromebook computer and audio headset
4. For discussions about English proficiency
   - A handout on federal law and materials from the common EL definition work (ESSA Sec. 80101, 20)
   - WIDA Performance Level definitions for Levels 4 through 6
   - Sample Model Performance Indicators (MPIs) for the targeted grade levels and for domains (wida.us)
   - Can Do Descriptors at targeted grades (wida.us)
5. For the Listening and Reading tests
   - An online Ordered Item Booklet program containing test items and associated materials (including scripts, directions, and audio material used by students) for the Listening and Reading tests. Test item difficulties ranged from Proficiency Level 6.0 down to 4.0 or the scale score closest to 4.0.
6. For the Writing and Speaking tests
   - Student Writing booklets for Grades 1–3 (for Grades 4-12, materials were online)
   - Student ordered work-sets or portfolios in an online program
     - For Writing, booklets include all Writing samples from one student at each score point down to Level 4.0
     - For Speaking, audio responses from one student at each score point down to Level 4.0.
7. Phase 1 impact information

Process. The Phase 1 study involved three activities. First, panelists participated in joint training that described the purpose, process, and objectives for the standard setting study. This training was done in large groups. The day before this training pre-selected panelists were shown the large-group training presentation and given the opportunity to ask questions. They served as Table Leaders the following day. Their task was to facilitate the process and address basic questions about the standard setting. During the standard setting activity, panelists were organized into four grade-based groups. Each group had three tables with four or five panelists at each table, and each group had two facilitators (one from WIDA and one from CAL). Group assignments are shown in Table 1.

Table 1. *Phase 1 Standards Setting Groups by Grade*

Group	**Grade**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	Group 1 with three tables of four to five panelists											
2				Group 2 with three tables of four to five panelists.								
3						Group 3 with three tables of four to five panelists						
4								Group 4 with three tables of four to five panelists				

In the second activity, panelists described the Listening, Speaking, Reading, and Writing performances that represented borderline English proficient students at their tables and in their groups. Subsequently, groups created operational definitions of "English proficient" at each grade. To support conversations, three sources of information were drawn upon. The first source was a description of what English proficient means in federal law, specifically ESSA Section 8101(20). This section of law defines ELs and is important in understanding what characterizes borderline English proficient students. The second source was the WIDA Performance Level definitions and other WIDA standards resources (e.g., MPIs and Can Do Descriptors), and the final source was the panelists' own personal expertise with ELs.

The English proficient discussion was facilitated in a large group setting with panelist sitting at their assigned tables in their assigned groups. Facilitators gave a presentation on the federal definition of ELs, directed panelists to materials that outlined the elements of that definition, highlighted that English proficiency might look different by cluster, if not by grade, and presented the range of English proficient scores currently adopted by WIDA states. Each table within each group then drafted a brief written description of the borderline "English proficient" student for their assigned grades. Tables then shared their descriptions with the group. After sharing, groups created a common description of the borderline English proficient student, and that was used for standard setting rounds.

Listening and Reading methodology. The standard setting panelists' task was to determine the scale score that represented a borderline student's performance in Listening and Reading. A modified Bookmarking method was used to accomplish this. A web-based, OIB program was created and populated with items from the 2015–2016 ACCESS 2.0 online test administration. Items in the program were selected by CAL and approved by WIDA. Selected items spanned scale score ranges from Proficiency Levels 4.0–6.0. The lower scale score range was chosen based on current minimum domain exit criteria requirements in WIDA states. It was considered unlikely that panelists would assign borderline students' scores below a 4.0 given the commonly recognized need to increase the academic language rigor of ACCESS 2.0 proficiency levels based on college and career ready standards. The higher scale score values in the selected items represented the upper end of the scale score range for that domain and cluster, which were all at Proficiency Level 6.0.

Apart from the proficiency level and scale score range criteria, several other steps were followed to select items for the OIB program. They were as follows:

1) Determine the difference between the observed empirical proficiency level score (e.g., Level 1, Level 2, etc.) and the "a priori" proficiency level designation made by item developers.
2) Set aside items where differences between proficiency level designations were either greater than one or less than two.
3) Eliminate any aberrant outlying items.
4) Create groups of remaining items where observed scale scores differed by 1 point (include "trains" of items if, e.g., item A is 231, item B is 232, and item C is 233).
5) Choose one item from each group created in Step 4, using the following criteria:
    a) if the result from step 1 is smaller for one item than for all other items within the group, choose that item
    b) otherwise, from among the items with the smallest value from Step 1, if one item is closer to the middle of the difference between the items above and below the group, choose that item
    c) otherwise, choose the item that tests the standard that has the least representation in that grade and domain
    d) otherwise, choose one item to keep.
6) Review the content of items and make final selection decisions based on the judgment of perceived item difficulty and the coverage of WIDA's standards (social instructional language, and the language of English language arts, mathematics, science, and social studies).

Once items were selected and approved, they were uploaded into the OIB program. Figure 1 shows a display from this program.

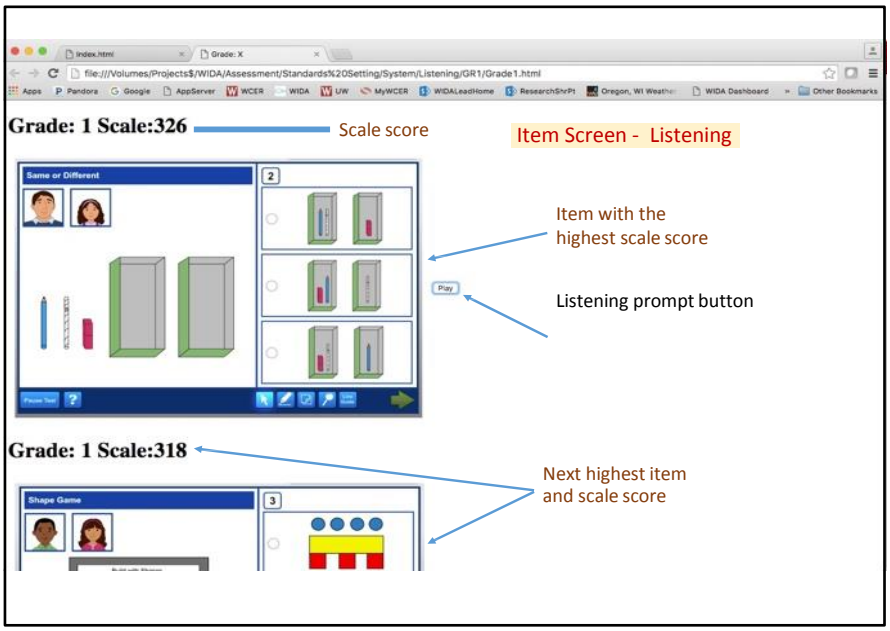


Figure 1. Grade 1 Listening test item from an Ordered Item Booklet in the web-based display.

In this figure, a graphic of a test item as it appeared on the online test screen is shown. The cluster and scale score for each item are at the top of each item display. A button to the right of items is for any associated audio passage (if applicable). Items for the Listening and Reading OIB were sorted in descending order from the most to least difficult.

Panelists were led through three rounds of ratings. Across all rounds, panelists were directed to make their ratings independently. In Round 1, panelists used all provided or created materials (e.g., WIDA Performance Definitions, borderline student descriptions) to identify the item at which a marginally English proficient EL would correctly respond 67% of the time. Once panelists' ratings were completed, each table discussed the group's Round 1 median, highest, and lowest scale score ratings. Each table then discussed any thoughts they had about Round 1.

Taking their Round 1 information and discussions into account, panelists started Round 2 by independently identifying the item they believed 67% of the marginally English proficient students would correctly answer. It was made clear that panelists did not have to change their ratings. However, Round 2 and 3 were opportunities for panelists to look at their ratings after they received more information. Once Round 2 ratings were complete, panelists discussed their ratings at each table. Group facilitators then presented impact data on the number and percent of ELs identified as meeting or not meeting the group's median score after Round 2.

After discussion, panelists were directed to begin Round 3. Facilitators made clear that Round 3 ratings were to be done independently and no further discussion about that domain or grade would occur once that round was complete. Panelists were also told that Round 3 ratings would be used to inform final cut scale scores for their assigned domains and grades. After Round 3, panelists were asked to fill out an evaluation form for that domain.

Writing and Speaking methodology. For the Speaking and Writing tests, a modified form of the Body of Work method was used. This method had panelists review a portfolio of all of a student's responses on the Writing or Speaking test. Panelists were then told to select the score of the student they thought represented the borderline case, starting at the highest grade in their group. Another web-based program was created to support panelists' Speaking and Writing test ratings. The selection process for student portfolio inclusion in the web-based program was akin to that for the Listening and Reading tests. The Writing and Speaking web-based program had several screens to support panelists' decisions. Figure 2 shows the initial screen for a Cluster 4-5 Writing assessment.

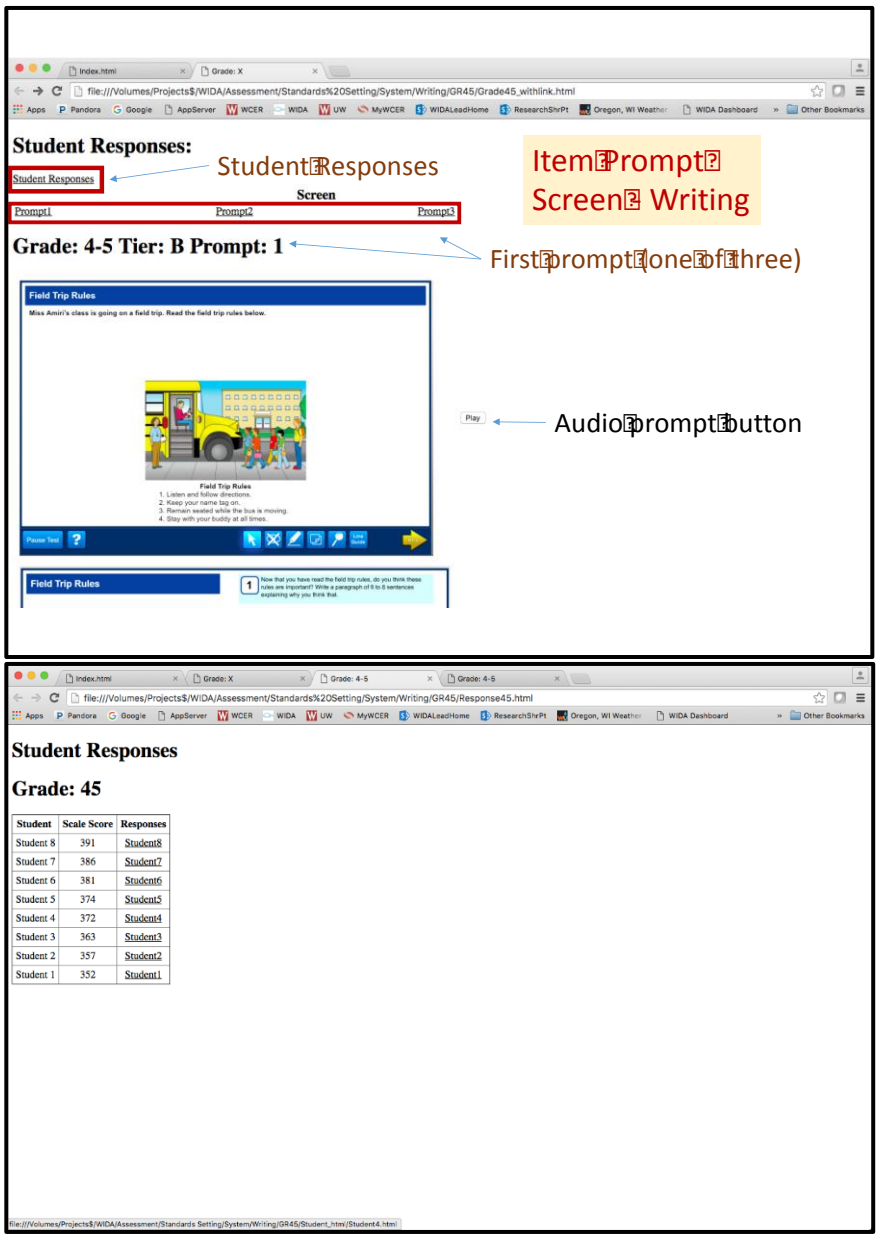


Figure 2. Grades 4-5 Writing item from a Student Portfolio in the web-based display.

Figure 2 has two panels. The top panel shows one of several items associated with a student's portfolio. To rate a portfolio, panelists selected a student scale score (panel at the bottom) and reviewed the associated item(s) and the student's response (Prompt 1, Prompt 2, and Prompt 3). Using this information, panelists made their selections. Three rounds of ratings were conducted for the Writing and Speaking test, identical in form to that done with the Listening and Reading tests. After Round 3 ratings, panelists were told that their final scale score values in their assigned domains and grades would be used to inform the final cuts. As with Reading and Listening, panelists were asked to fill out an evaluation for each domain.

Results

The results presented below reflect panelists' recommendations (before and after smoothing). This information was used by CAL to set up and frame discussions for Phase 2 of the standard setting process. It is important to note that the scale score values presented in this section do not necessarily reflect final recommended Phase 1 cuts. Once Phase 1 and Phase 2 deliberations were complete, CAL re-centered domain scale score values such that all domain scores were centered at a value of 350 and at Proficiency Level 4.0 in Grade 5. This had been the characteristic of the ACCESS score scale when originally developed. This practice has been continued. By no means did this scale re-centering process change Phase 1 or Phase 2 final cuts as they relate to proficiency level interpretations. This does, however, change the final scale score values representing those proficiency levels. For the Reading and Listening domain scale scores, that difference is relatively small. For the Writing and Speaking domain scale scores, that difference is larger. Regardless, panelist recommendations (after smoothing) relative to rankings and relationships to proficiency levels remain the same.

Domain results. For all domains, Round 3 ratings were used as the final recommendations of panelists.

Listening.

Table 2. *Listening Round 3 Median Cut Scores by Group*

Grade	Group 1	Group 2	Group 3	Group 4	Median Score
1	311				311
2	340				340
3	363				363
4	372	369			371
5	383	381			382
6		399	411		405
7		405	421		413
8		421	431	405	419
9			433	399	416
10			441	401	421
11				417	417
12				423	423

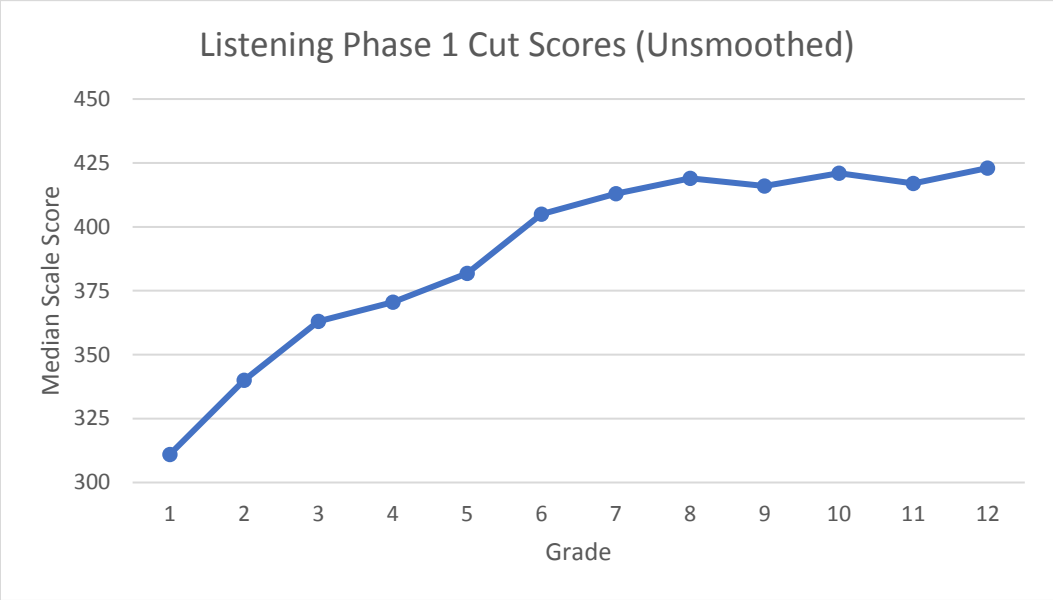


Figure 3. Panelists' Round 3 Listening cut scores across raters by grade.

Reading.

Table 3. *Reading Round 3 Median Cut Scores by Group*

Grade	Group 1	Group 2	Group 3	Group 4	Across Groups
1	329				329
2	346				346
3	356				356
4	371	371			371
5	376	378			377
6		390	398		394
7		394	406		400
8		406	410	394	403
9			413	398	406
10			417	406	412
11				410	410
12				417	417

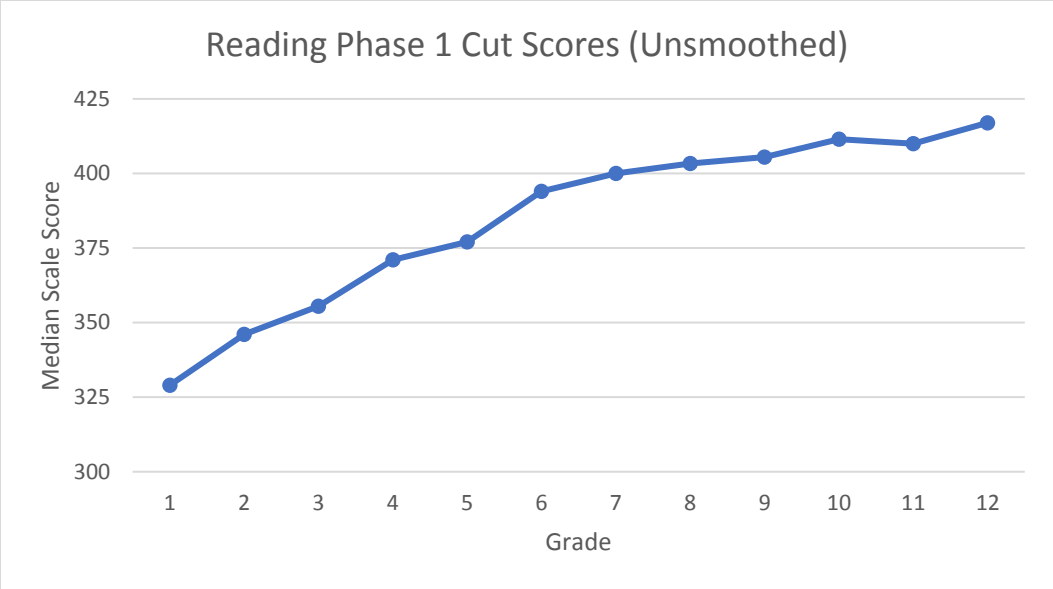


Figure 4. Panelists' Round 3 Reading cut scores across raters by grade.

Speaking.

Table 4. *Speaking Round 2 or 3 Median Cut Scores by Group*

Grade	Group 1	Group 2	Group 3	Group 4	Across Groups
1	390				390
2	411				411
3	425				425
4	437	399			418
5	450	425			438
6		414	435		425
7		435	468		452
8		468	497	414	460
9			471	407	439
10			499	430	465
11				450	450
12				470	470

Note. Groups 2 and 3 did not complete Round 3 of the Speaking domain. Impact information was shared in Round 2 for those groups.

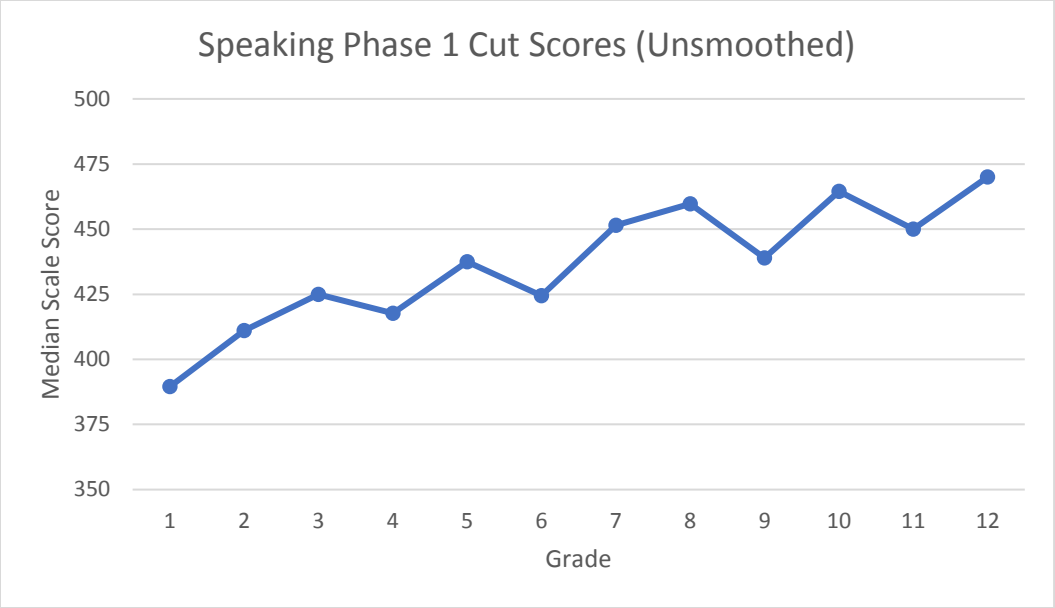


Figure 5. Panelists' Round 3 Speaking cut scores across raters by grade.

Writing.

Table 5. *Writing Round 3 Median Cut Scores by Group*

Grade	Group 1	Group 2	Group 3	Group 4	Across Groups
1	314				314
2	326				326
3	331				331
4	374	363			369
5	386	372			379
6		376	382		379
7		382	390		386
8		390	391	382	388
9			423	413	418
10			431	417	424
11				420	420
12				423	423

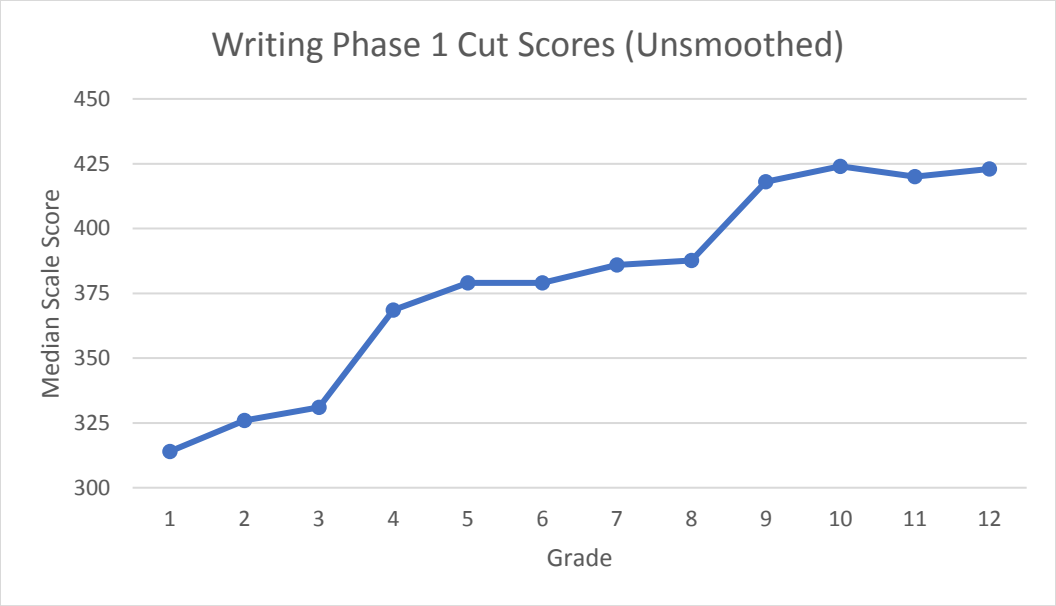


Figure 6. Panelists' Round 3 Writing cut scores across raters by grade.

Vertical articulation. As is evident in the figures in the last section, the recommended median cut scores across grades are not typically smooth or systematic. In some cases, there are large jumps between grades, e.g., Grades 3 and 4 in Writing. In others, higher grades have lower median cut scores, e.g., Grades 8 and 9 in Speaking. Organizing the groups such that they overlap grades is intended to ameliorate this; however, it is often the case that even with overlapping grades, smoothing is required. The goal of smoothing is to provide a meaningful, vertically articulated series of cut scores across grades.[3]

The smoothing routine used for Phase 1 is a non-linear growth function of the following form,

$$Smoothed\ Score = C + A(1 - e^{-kg}),$$

where C is the lowest observed score; A is the amplitude (highest minus the lowest score plus one); k is a non-linear growth function, and g is grade. Using Microsoft Excel's advanced solver function and the above non-linear growth function, panelists' original cut scores can be smoothed to a fitted curve representing their judgments.[4] The following figures overlay the fitted cut scores over the observed cut scores. The following table lists the smoothed cut scores for each domain. The values displayed in this table are those sent to CAL to prepare for Phase 2 of the standard setting study.

[3] Given that ACCESS 2.0 is on a vertical scale, increasing scale score expectations across grades is consonant with the increased academic English language ability students require to be proficient. That is, the amount and complexity of language expected of an early elementary student is much less than that of a high school student. The ACCESS 2.0 vertical scale should reflect that, as should the potentially proficient cut scores.
[4] For more on this see http://people.chem.ucsb.edu/laverman/leroy/Chem116/PDF116CL/Solver.pdf.

Table 6. *Smoothed Phase 1 Recommended Cut Scores*

Grade	Phase 1 Smoothed Cut Scores			
	Listening	Reading	Speaking	Writing
1	310	328	392	311
2	339	345	404	328
3	361	359	415	343
4	378	371	424	357
5	391	381	432	369
6	401	390	439	380
7	409	397	445	390
8	415	403	450	400
9	419	408	455	408
10	423	412	459	415
11	425	415	462	422
12	427	418	465	428

The following figures overlay the Phase 1, Round 3 cut scores (unsmoothed) and the smoothed cut score values.

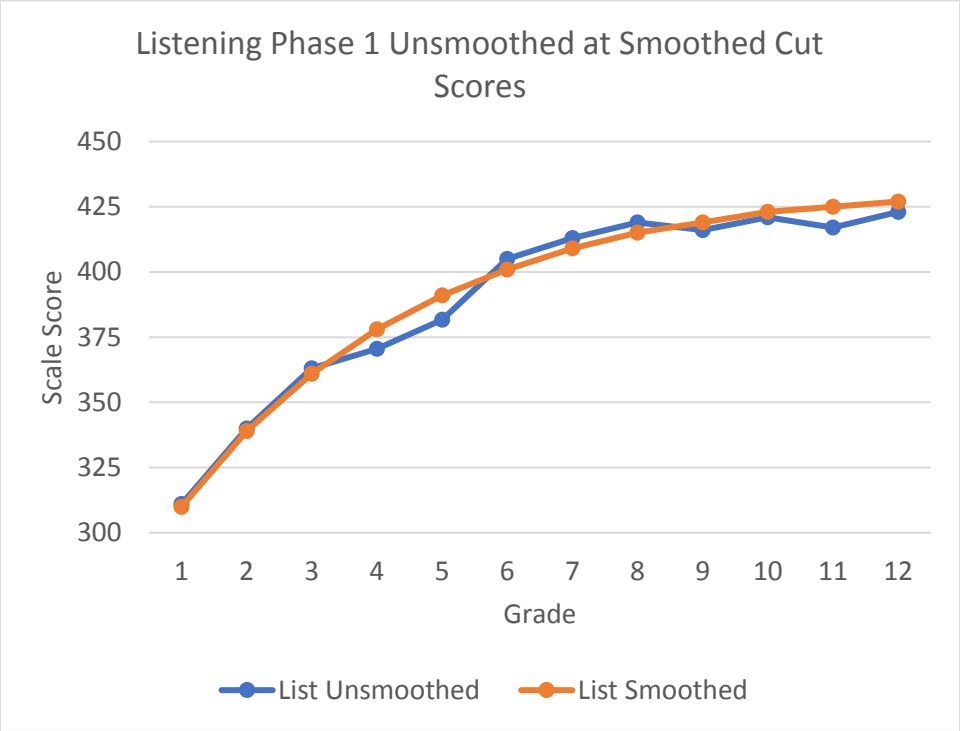


Figure 7. Listening Phase 1 smoothed and unsmoothed cut scores by grade.

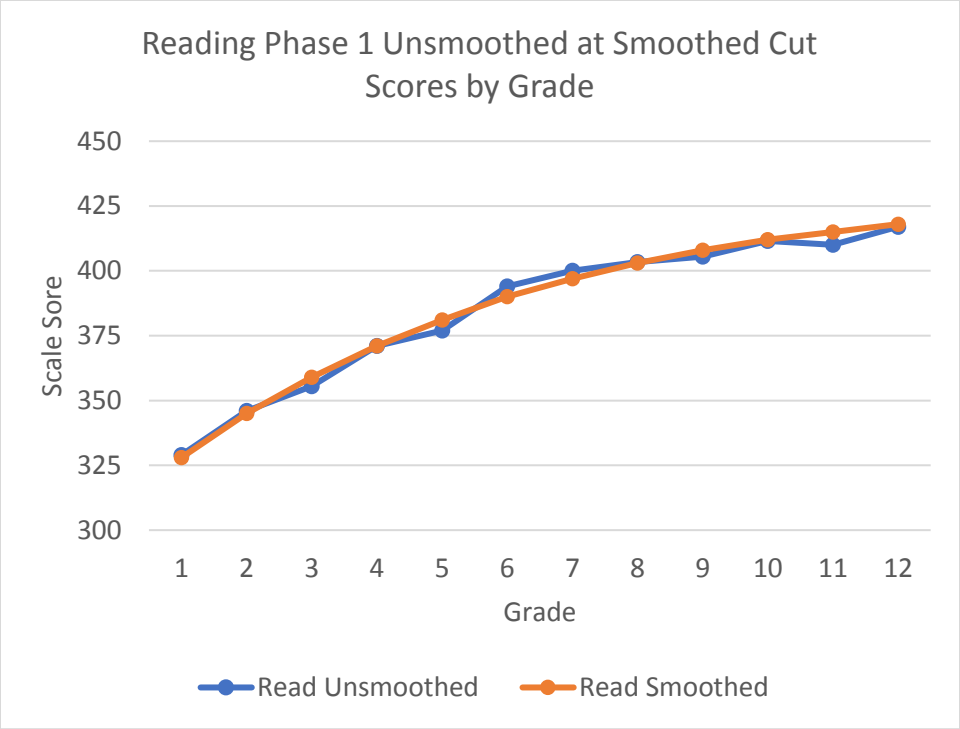


Figure 8. Reading Phase 1 smoothed and unsmoothed cut scores by grade.

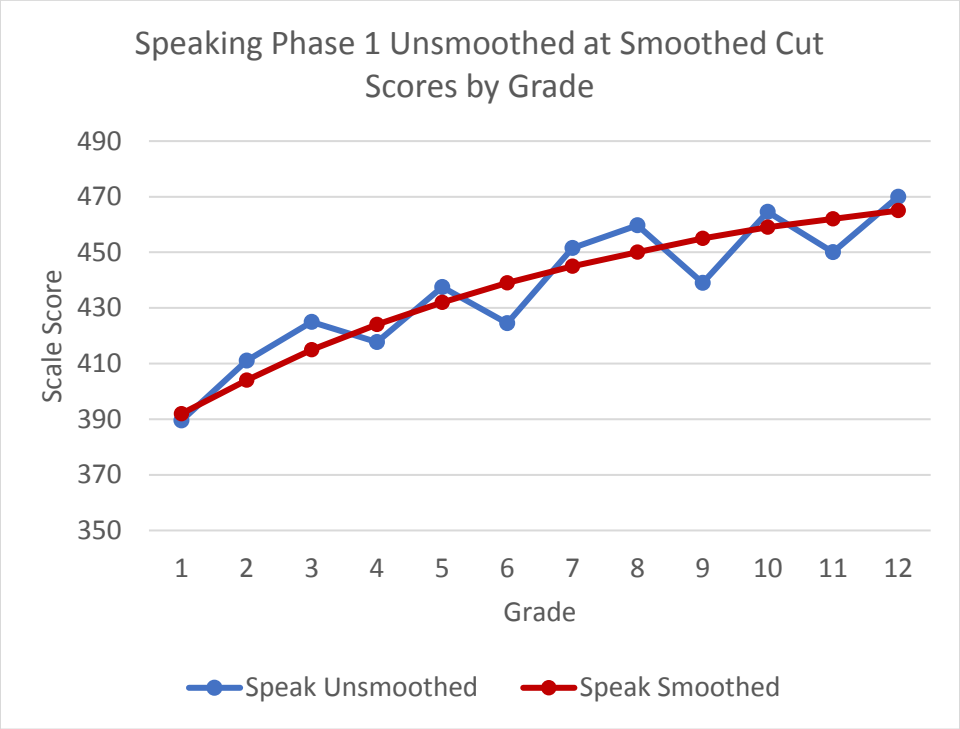


Figure 9. Speaking Phase 1 smoothed and unsmoothed cut scores by grade.

In panelists' Round 3 ratings of Speaking (unsmoothed), four adjacent grade-level sequences are not monotonic: Grades 3 to 4, Grades 5 to 6, Grades 8 to 9, and Grades 10 to 11. Three out of the four non-monotonic sequences are at grade-level cluster borders. For references, ACCESS for ELLs Online has five grade-level clusters: Grade 1, Grades 2–3, Grades 4–5, Grades 6–8, and Grades 9–12. Additionally, all non-monotonic sequences are at grades where two groups either began to overlap (Group 2, Grade 4; Group 3, Grade 6) or end overlapping (Group 3, Grade 10). At least two factors contributed to the up and down cut score assignments for Speaking. First, each new cluster had new student portfolios with new items. The new items and student performances at the higher grade level may have colored panelist responses. Second, groups had different perceptions of how the minimally competent student should perform. Generally (and not surprisingly), groups in lower grades assigned higher cut scores compared to their higher grade group colleagues for the same grade. These cut scores assignment differences are most pronounced in Speaking.

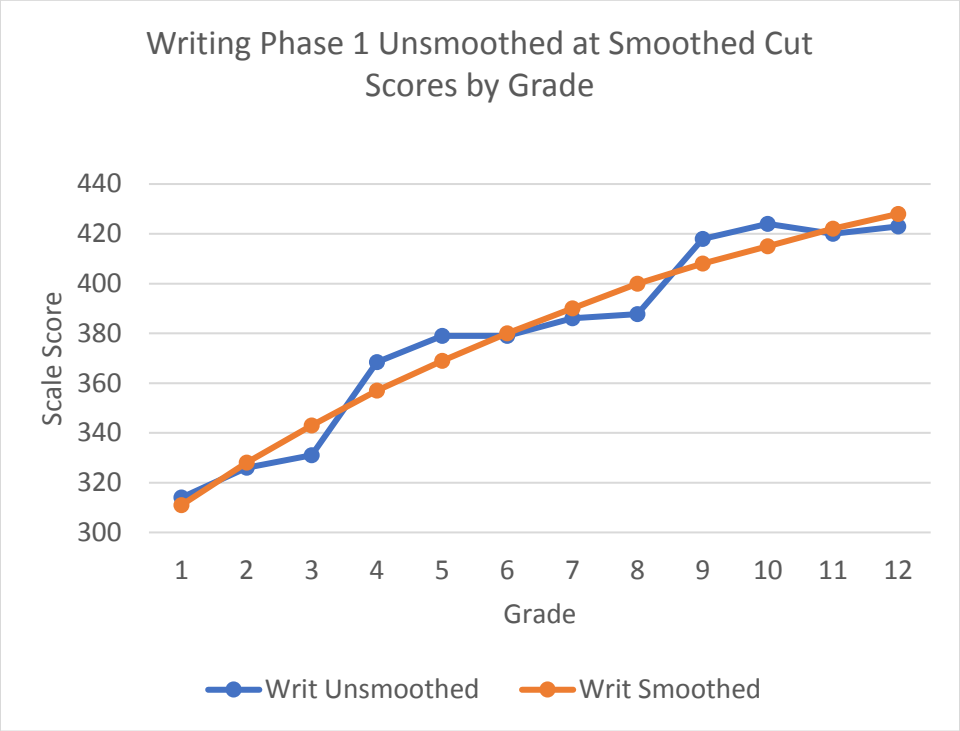


Figure 10. Speaking Phase 1 smoothed and unsmoothed cut scores by grade.

Training evaluation summary

During the standard setting meetings, panelists were asked to complete an anonymous evaluation form on the training process. As indicated in Table 7, almost all panelists rated the training portion of the Phase 1 meeting as *excellent* or *good*.

Table 7. *Phase 1 Training Evaluation Responses (55 Responses)*

Survey Questions	Responses			
	Excellent	**Good**	**Fair**	**Poor**
The **purpose** of the standard setting meeting was communicated clearly	*43 (78.2%)*	*12 (21.8%)*		
The standard setting training was **organized** and **easy to follow**	*40 (72.7%)*	*14 (25.5%)*	*1 (1.8%)*	
I understand my **role and responsibilities** at this standard setting meeting	*33 (60%)*	*21 (38.2%)*	*1 (1.8%)*	
The presenter **responded effectively** to audience questions and comments	*46 (83.6%)*	*9 (16.4%)*		
I feel **prepared** to begin the standards setting process	*30 (54.5%)*	*24 (43.6%)*	*1 (1.8%)*	

Twenty panelists commented on the evaluation form and were consistently positive when speaking about the training process.

> *Excellent organization and attention to detail.*
>
> *One addition could be to walk through (step by step) a sample item together as far as what the expectations are for the participants.*
>
> *The framing of the training was well organized and the presenter (Gary) was knowledgeable about exactly what was being communicated to participants.*
>
> *Excellent training and process. However, the schedule may need [to be] adjusted for future due to the length of discussion to determine the border line.*
>
> *The morning training was very helpful. The speaker was clear and to the point. Everyone seems extremely organized. I really appreciate that. No question seems wrong, and I appreciate the safe environment being established. The 2/3 broadline [sp] student could have been revisited more in depth right before we began Round 1.*

These quotes also illustrate how some panelists used the training evaluation form to comment on the standard setting process in addition to the training process. The subject of these comments (borderline student definition, use of time, etc.) reflect comments from the standard setting process evaluation form.

Section II: Phase 2 Report

Introduction

The purpose of Phase 2 of the standard setting was to determine where to set the cut scores between proficiency levels for all grade levels in all domains. In preparation for Phase 2, CAL developed proposed cut scores for the domains of Listening, Reading, and Writing. In addition, CAL selected items to be used in the OIBs for Listening and Reading, and the portfolios to be used in Speaking and Writing. The procedures for creating the proposed cut scores and for selecting items and portfolios are described below.

Developing proposals for cut scores. To develop the proposals for the cut scores, the scale score for Proficiency Level 1.0 for each grade was fixed at its previous value, and the Phase 1 cuts were fixed at Level 5.5 for Listening and Reading and Level 4.5 for Writing. The other cut scores were adjusted so that the proportional width of the proficiency level bands would be maintained. Because of the extensive changes to the administration and scoring of the Speaking test, no cut score proposal was prepared for that domain. The proposed cut scores are shown in Tables 8–10.

Table 8. *Proposed Cut Scores: Listening*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	221	258	277	297	323
2	246	284	303	326	352
3	265	302	322	348	374
4	279	317	336	365	391
5	291	328	348	378	404
6	300	338	357	388	414
7	309	347	366	396	422
8	317	354	373	402	428
9	323	361	380	406	432
10	329	367	386	410	436
11	335	373	392	412	438
12	340	378	397	414	440

Table 9. *Proposed Cut Scores: Reading*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	260	285	302	320	336
2	280	305	322	337	353
3	294	320	336	351	367
4	306	331	348	363	379
5	315	340	357	373	389
6	322	348	364	382	398
7	329	355	371	389	405
8	335	360	377	395	411
9	340	366	382	400	416
10	345	371	387	404	420
11	349	375	391	407	423
12	353	379	395	410	426

Table 10. *Proposed Cut Scores: Writing*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	236	266	296	326	359
2	262	293	313	343	385
3	281	311	328	358	404
4	295	326	342	372	418
5	307	338	354	384	430
6	317	348	365	395	440
7	326	356	375	405	449
8	333	364	385	415	456
9	340	371	393	423	463
10	346	377	400	430	469
11	352	383	407	437	475
12	357	388	413	443	480

Choosing items for Listening and Reading Ordered Item Booklets. The process for choosing Listening and Reading items for the OIBs was similar to that used in Phase 1; however, to determine all cut scores, items had to range across the scale score range. In addition, previous experience has demonstrated that the task can be difficult for panelists when items are close together in terms of difficulty, or when there is a large mismatch between the intended difficulty of an item and its empirical difficulty. To keep the number and spread of items at a manageable level, the following steps were followed:

1) Determine the difference between a priori proficiency level and empirical proficiency level (subtract empirical from a priori).
2) Eliminate items whose results from Step 1 are either >1 or <-2.
3) Group items whose scale score value differs by 1 point or less (include "train" of items if, e.g., item A is 231, item B is 232, and item C is 233).
4) Choose one item from each group created in Step 3, using the following criteria:
   a) if result from Step 1 is smaller for one item than for all other items within the group, choose that item
   b) otherwise, from among the items with the smallest value from Step 1, if one item is closer to the middle of the difference between the items above and below the group, choose that item
   c) otherwise, choose the item that tests the standard that has the least representation in that grade and domain
   d) otherwise, choose one item at random.

Table 11 shows the number of items that were selected for each domain per cluster.

Table 11. *Number of Items Selected per Cluster*

Domain	Cluster				
	1	**2-3**	**4-5**	**6-8**	**9-12**
Listening	29	28	30	36	21
Reading	25	27	26	33	33

Choosing Speaking and Writing Portfolios. Speaking and Writing portfolios consisted of all Speaking or Writing samples produced by a student in response to all of the Speaking or Writing prompts that student saw. As with the OIBs, since the purpose of this phase of the standard setting was to set cut scores across the spectrum of proficiency levels, the portfolios selected represented the range of observed raw scores on the test. On the other hand, in order to ensure that the difference in proficiency between adjacent portfolios was detectable by panelists, it was necessary to sample across the raw score points, rather than choosing portfolios at all points. In addition, portfolios were chosen that had relatively flat profiles; that is, the ratings on the different tasks did not differ greatly.

It was decided to choose ten portfolios for each domain, starting at Raw Score 3, and then choosing portfolios at intervals of about three raw score points. To choose the portfolios, the following steps were taken:

- WIDA identified up to ten portfolios at each raw score point that have relatively flat profiles.
- CAL reviewed the portfolios to ensure that:
  - Ratings were accurate (with adjustments made if necessary in the judgment of expert raters at CAL)
  - The samples were audible or legible

    - The responses were on-topic
- Based on that review, a final selection of portfolios was made.

For each cluster, one of the Speaking portfolios was chosen as a practice portfolio.

Methods

Standard setting methods. For the Writing and Speaking domains, a modified Body of Work method was used. For the Listening and Reading domains, a modified Bookmarking method was used. These methods are described in detail below.

Panelists. Overall, the procedure for selecting panelists for Phase 2 was similar to that for Phase 1, with some minor differences. Since 19 panelists participated in both meetings we did not assign Table Leaders in Phase 2. In the qualifications survey, 93% of Phase 2 panelists indicated that they had experience working with ELs with disabilities, and five educators held additional credentials, degrees, or specialization in special education. The Phase 2 meeting had 54 panelists, due to some late cancellations and a travel shutdown.

Materials. All panelists had access to computers with an internet connection to individualized Google Sheets data collection spreadsheets on which they recorded their judgments. In addition, they had online access to the OIBs for Reading and Listening (including audio files for Listening items), to Writing tasks for Tiers B/C and all Speaking tasks, and to student portfolios for Speaking and Writing. Examples of meeting materials are in Appendices G-T.

Handouts for panelists included the following:

- Phase 2 meeting agenda
- Non-disclosure agreement form
- Panelist background questionnaire
- Training handout consisting of selected slides from the training PowerPoint. Slides included a tool, referred to as the *cheat sheet*, that demonstrated how to translate a proficiency level score (e.g., 4.2) into percentages.
- Training evaluation forms
- Domain evaluation forms
- Proficiency level descriptors
- Standards booklet (1 per table)
- Data response sheets
- Tier A Writing books

Process. Panelists met as a whole group on the first day to receive general training. The training included information on the background of the test, the need for the current standard setting, the essentials of standard setting, and instruction on the Body of Work procedure used for Speaking and Writing. After the general introduction, panelists met in their groups. After introductions, each group started with the Speaking domain, followed by Writing, Reading, and Listening.

Speaking. Panelists first reviewed the Speaking tasks in the highest cluster assigned to their group, and then discussed the linguistic features they might expect of students at the borderline of Proficiency Levels 2–6 when responding to those tasks.

The panelists then listened to a sample portfolio, and were asked to rate it as if it were the product of a student in the highest grade in the cluster. They were instructed to consider the following questions in deciding on a rating:

1. Given your understanding of the WIDA ELD Standards, which WIDA ELD proficiency level (1–6) does the performance you hear most reflect for that grade level?
2. Then decide — How confident am I that this performance fits that WIDA ELD proficiency level (100% – 50%)?
3. Then decide — If I'm not 100% confident of my selection in Step 2, which adjacent WIDA ELD proficiency level (higher or lower) do I think might also be characterized by this performance?

They then recorded their ratings on the data response sheet and transferred their ratings to their individual data collection spreadsheets. Facilitators and helpers circulated within their groups to answer questions and ensure that the panelists stayed on task. Facilitators also monitored their spreadsheets, and when they saw that all panelists had entered their ratings, they revealed the results to the panelists on a screen. The results included all responses from the panelists (shown anonymously), along with the group averages and standard deviations. The panel discussed the process and results as a group, and the facilitator answered any questions they had.

Panelists then began rating the portfolios for their groups, starting with the portfolio that received the lowest raw score in the highest cluster. Panelists were asked to listen to the portfolio and rate it for the highest grade in the cluster, following the same procedures described for the sample portfolio. They then rated the same portfolio for the other grade levels in the cluster. When the facilitators saw that all panelists had entered ratings for all grade levels for the first portfolio, they revealed the ratings to the panelists and had them discuss their ratings at the table. After the discussion, panelists conducted a second round, in which they were given the opportunity to confirm or adjust their ratings. They then moved on to the next portfolio in the cluster and repeated the process, continuing through all of the portfolios in the cluster. After they had finished one cluster, they went on to the next and repeated the process. When they had finished rating all of the portfolios for their group in that cluster, the facilitators revealed the cuts that would result from the panelists' ratings, along with impact

data in the form of the distribution across proficiency levels of the population of students from the most current operational administration that would result if the panels' cuts had been in force for that operational testing year. Panelists were then asked to fill out the evaluation forms for the Speaking portion of the standard setting.

Writing. Procedures for Writing were similar to those for Speaking. However, because a proposal had been created for Writing cut scores ahead of the meeting, panelists also saw the proficiency level scores that a portfolio would receive based on the proposed cut scores, and were asked to consider that score as a starting point, with the option to confirm or adjust as they deemed necessary.

Reading and Listening. Training for the Reading and Listening Bookmarking procedure was conducted in the groups, with the facilitator leading the training. After the training, facilitators modeled the task for the panelists by having the panelists look at an item from the highest cluster in the middle of the range of all items for that cluster, and leading the panelists in a discussion regarding the probability that, first, a student at Proficiency Level 4.0 would answer that item correctly, and next that a student at Proficiency Level 3.0 would answer that question correctly.

After that discussion and a review of the procedures, panelists were asked to consider first where to set the Proficiency Level 2.0 cut for the highest grade in the highest cluster assigned to the panel. They were asked to start with the easiest item in the OIB and work upward. They were instructed to keep in mind a student at the border between Levels 1 and 2, and ask themselves, in light of the performance level descriptors, at what point would the borderline student no longer have a 50% or higher probability of getting an item correct?

To confirm their answer, they were also instructed to continue examining items beyond the first one which, in their estimation, the borderline student would no longer have a 50% or higher probability of getting correct. Once they were satisfied with their answers, they recorded the scale score of that item on their rater scoring sheet or, if they felt that the point at which the borderline student would fall below a 50% probability lay between two items, they recorded a scale score between those two items. They then entered their answers in the online sheet, and followed the same procedures to establish the 2.0 cut scores for all grade levels in their groups.

When facilitators saw that all panelists had recorded scores for the 2.0 cut for all grades in their group, they revealed the results to the panelists, who then discussed the results at their tables. Panelists were then asked to conduct a second round in which they had the opportunity to confirm or adjust their answers.

After the panelists had established their 2.0 cuts, they worked individually to establish the 3.0, 4.0, 5.0 and 6.0 cuts for all grade levels in their group. When all panelists had finished, the results were revealed and they discussed them in groups, and then had a chance to confirm

or adjust their answers. After all panelists had finished both rounds, the facilitators showed them the final results, along with impact data in the form of the distribution across proficiency levels of the population of students from the most current operational administration that would result if the panels' cuts had been in force for that operational testing year. Panelists were then asked to fill out the evaluation forms for the Reading portion of the standard setting.

Results

Domain results.

Speaking. As typical with results from a Body of Work standard setting methodology, CAL conducted logistic regression analyses (by grade) to relate Speaking and Writing scale scores to the odds of a portfolio being considered by the panelists to be representative of one or another of two adjacent proficiency levels (e.g., the ratio of the probability that a portfolio is characteristic of Level 3 to the probability of the portfolio being considered as Level 2). The logistic regression model may be represented as follows:

$log(Odds) = Intercept_j + \beta_j \times Scale\ Score$, where

$Intercept_j$ is a cut-based intercept;

β_j is a cut-based slope parameter.

From the data collected from the standard setting panelists, CAL used logistic regression to determine the scale score at which a corresponding proficiency measure has a 50% chance of being considered at a certain proficiency level and also a 50% chance of being considered at the next higher proficiency level (i.e., Levels 2.0, 3.0, 4.0, 5.0, and 6.0). Scale scores representing this 50% probability are panelists' judgments (as a group) of where the cut scores lie between adjacent WIDA proficiency levels. The panelists' cut scores are presented in Figure 11.

Table 12 shows the values of the panelists' cut scores across the grades on the Speaking scoring scale. Figure 11 and Table 12 also include the scale score values of the Phase 1 cuts (later approved by the Executive Committee of the WIDA Board). Note that the scale score values presented in this proposal, including the Phase 1 cuts, are not the ones used in the Phase 1 or the Phase 2 study. Rather, they have been adjusted so that the Level 4.0 cut at Grade 5 is set to be 350, which is the center of the underlying Speaking scale. (This had been done with the ACCESS score scales in each domain when the scales where originally developed. The ACCESS 2.0 score scales in Listening, Reading, and Writing all have this characteristic, even if 350 is no longer the Level 4.0 cut for Grade 5.) Thus, following the analysis of the Phase 2 results, CAL made this adjustment to the Speaking score scale in order to make the range of the Speaking scale scores similar to that of the other domains. It must be emphasized that such an adjustment does not alter the underlying proficiency measures (i.e., proficiency level interpretations from both phases of the standard setting study).

For each proficiency level cut, cut score values are expected to increase monotonically across the grades, reflecting the grade-level developmental trajectory. Nevertheless, Figure 11

reveals that some of the proficiency level cuts exhibit reversed patterns in the cut score values. For example, moving from Grade 1 to Grade 2, the cut scores drop for the Level 2.0, Level 3.0, and Level 6.0 cuts, respectively. If the intent is to maintain a monotonic increase of the cut scores across the grade levels for each proficiency level cut, some adjustments will need to be made to the cut scores resulting from panelist judgments in Phase 2 of the standard setting. Below we describe a smoothing method for adjusting the cut scores across grades, which alleviates the reversed patterns in the cut score values.

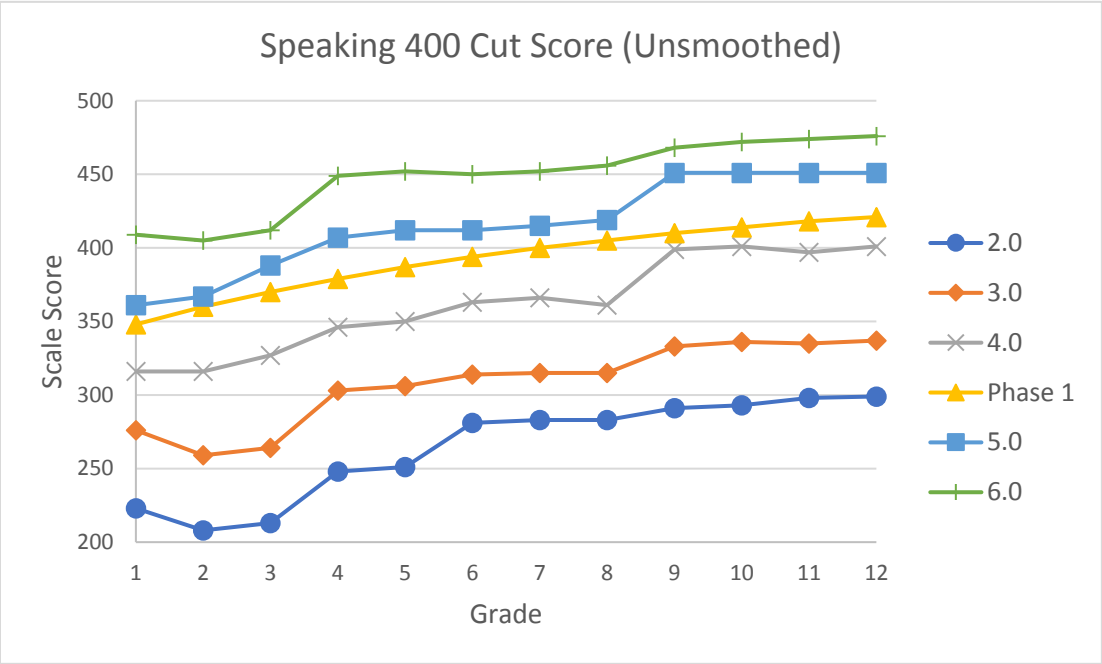


Figure 11. Panelists' Speaking cut scores across grades.

Table 12. *Panelists' Initial Speaking Cut Scores across Grades (Unsmoothed)*

Grade	WIDA Proficiency Levels					
	2.0	**3.0**	**4.0**	**Phase 1 cut**	**5.0**	**6.0**
1	223	276	316	348	361	409
2	208	259	316	360	367	405
3	213	264	327	370	388	412
4	248	303	346	379	407	449
5	251	306	350	387	412	452
6	281	314	363	394	412	450
7	283	315	366	400	415	452
8	283	315	361	405	419	456
9	291	333	399	410	451	468
10	293	336	401	414	451	472
11	298	335	397	418	451	474
12	299	337	401	421	451	476

Writing. The process for analyzing the Writing results was similar to that for Speaking. Results from Writing are shown in Figure 12 and Table 13. Note that, as with Speaking, the monotonicity assumption was not always met, and so the scores needed to go through a smoothing process.

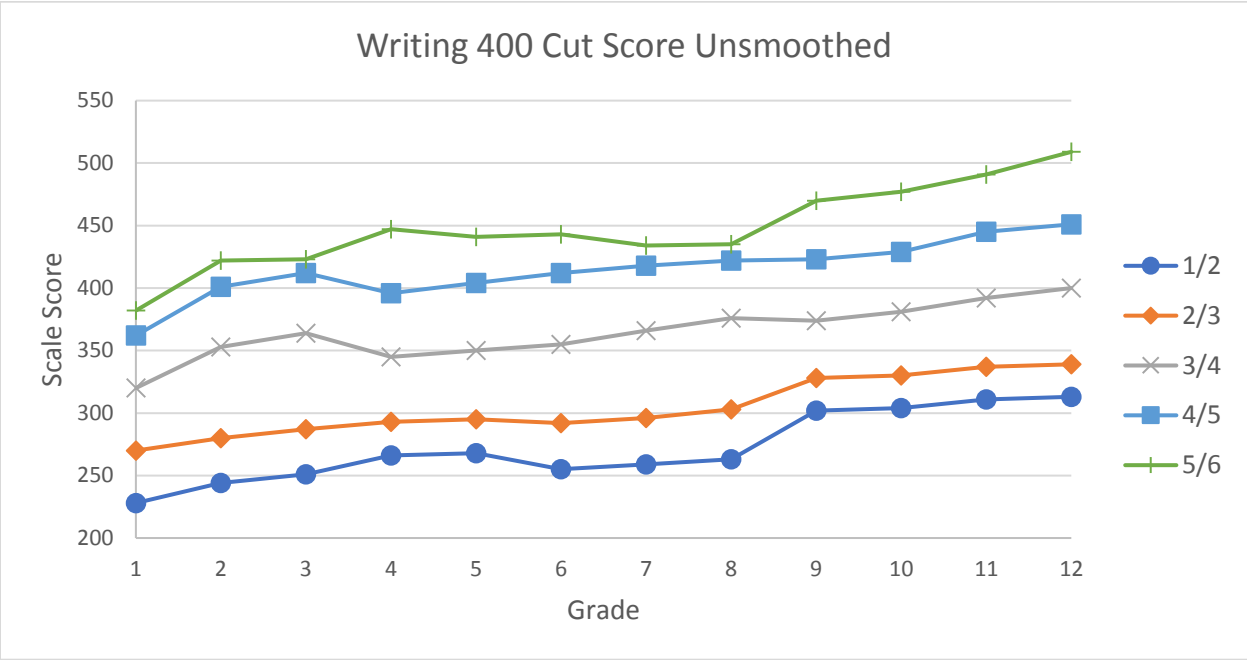


Figure 12. Panelists' Writing cut scores across grades.

Table 13. *Panelists' Initial Writing Cut Scores across Grades (Unsmoothed)*

	WIDA Proficiency Levels					
Grade	**2.0**	**3.0**	**4.0**	**Phase 1 cut**	**5.0**	**6.0**
1	228	270	320	311	362	382
2	244	280	353	328	401	422
3	251	287	364	343	412	423
4	266	293	345	357	396	447
5	268	295	350	369	404	441
6	255	292	355	380	412	443
7	259	296	366	390	418	434
8	263	303	376	400	422	435
9	302	328	374	408	423	470
10	304	330	381	415	429	477
11	311	337	392	422	445	491
12	313	339	400	428	451	509

Listening and Reading. Figure 13 (Listening) and Figure 14 (Reading) plot panelists' mean judgments in terms of scale scores for each grade and cut, with cuts following a coloring convention (the absent red is for the minimum score for each grade—which panelists did not evaluate). In addition to the panelists' judgments, the figures also include the scale score values of the results from Phase 1 (Indigo). Note that in Phase 1, the panelists did not assign a proficiency level interpretation to those exit cuts; rather, WIDA proposed assigning a Proficiency Level interpretation of 5.5 to those cuts, and obtained tentative approval from the Executive Committee of the WIDA Board. Table 14 and Table 15 show the panelists' mean judgments in terms of scale scores for each grade level cut, along with the Phase 1 cut and the scale score that would be interpreted as Level 5.5 based on the panelists' 5.0 and 6.0 cuts from Phase 2.

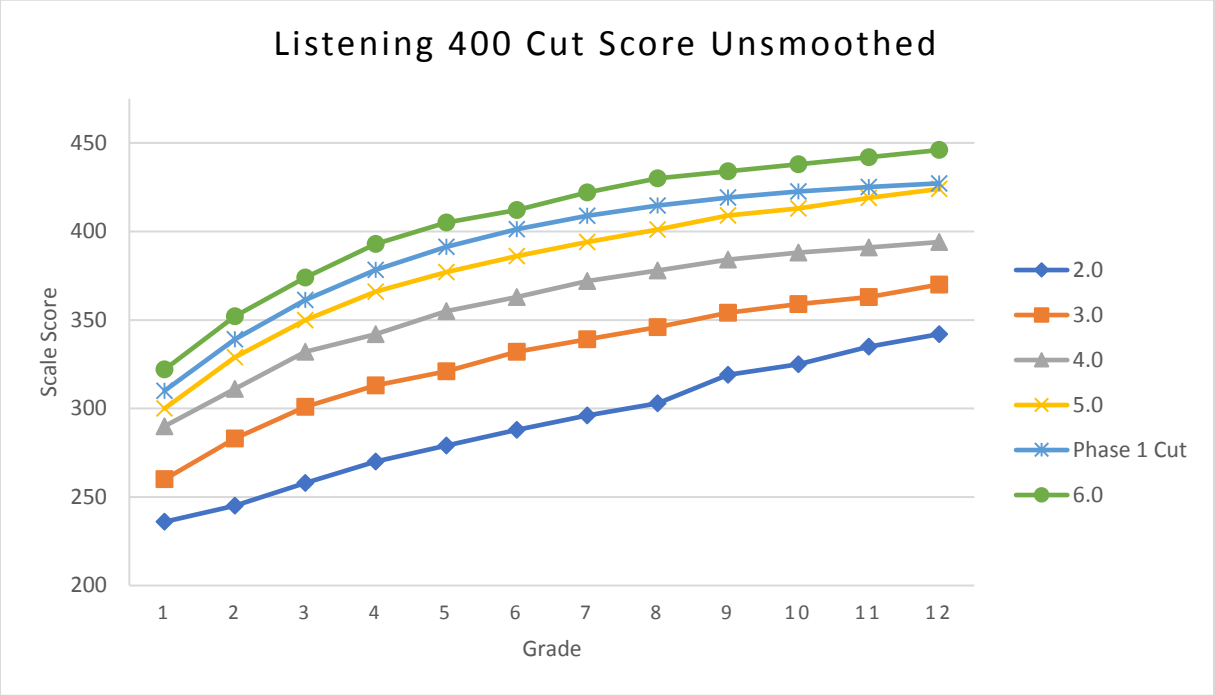


Figure 13. Panelists' Listening cut scores across grades.

Table 14. *Panelists' Initial Listening Cut Scores across Grades (Unsmoothed)*

Grade	WIDA Proficiency Levels						
	2.0	3.0	4.0	5.0	Phase 1 cut	5.5	6.0
1	236	260	290	300	310	311	322
2	245	283	311	329	339	341	352
3	258	301	332	350	361	362	374
4	270	313	342	366	378	380	393
5	279	321	355	377	391	391	405
6	288	332	363	386	401	399	412
7	296	339	372	394	409	408	422
8	303	346	378	401	415	416	430
9	319	354	384	409	419	422	434
10	325	359	388	413	423	426	438
11	335	363	391	419	425	431	442
12	342	370	394	424	427	435	446

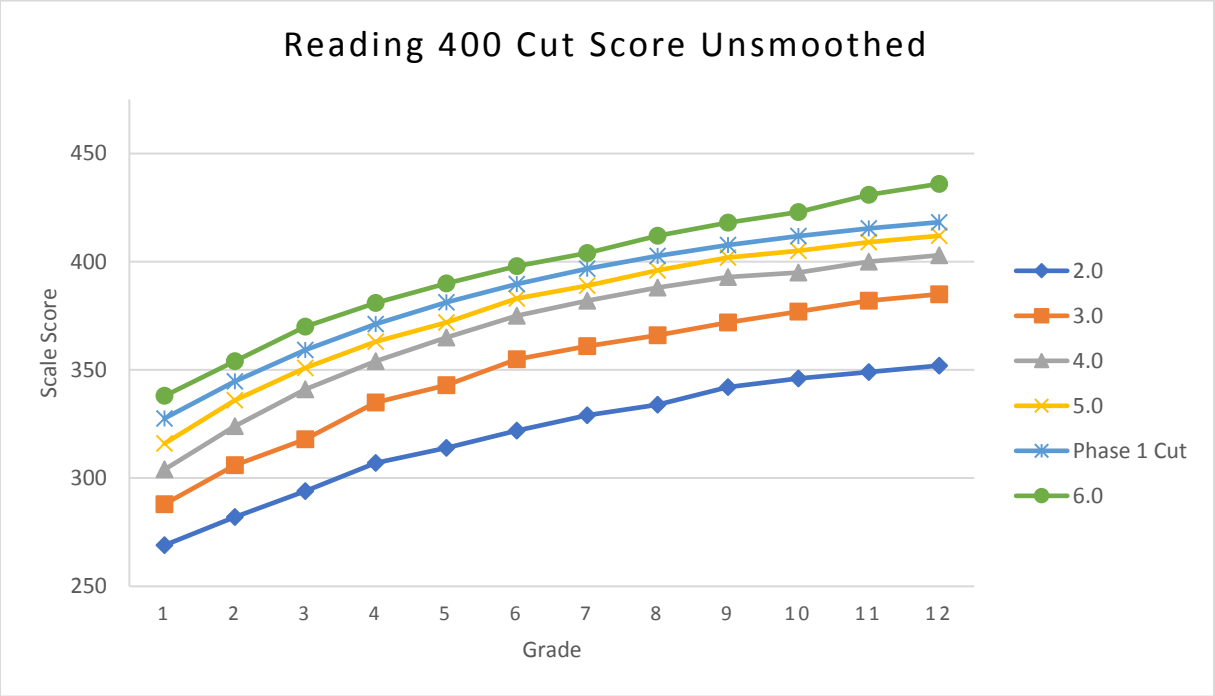


Figure 14. Panelists' Reading cut scores across grades.

Table 15. *Panelists' Initial Reading Cut Scores across Grades (Unsmoothed)*

Grade	WIDA Proficiency Levels						
	2.0	3.0	4.0	5.0	Phase 1 cut	5.5	6.0
1	269	288	304	316	328	327	338
2	282	306	324	336	345	345	354
3	294	318	341	351	359	361	370
4	307	335	354	363	371	372	381
5	314	343	365	372	381	381	390
6	322	355	375	383	390	391	398
7	329	361	382	389	397	397	404
8	334	366	388	396	403	404	412
9	342	372	393	402	408	410	418
10	346	377	395	405	412	414	423
11	349	382	400	409	415	420	431
12	352	385	403	412	418	424	436

While the Phase 1 cuts (and historical proportional distance between cuts) drove the initial scale score values proposed to panelists for each grade and cut, panelists were not provided with Phase 1 cuts directly. Nevertheless, for most grade levels the Level PL 5.5 score that results from the panelist' judgment is within two scale score points of the Phase 1 cuts; thus, to a large extent, the panelist kept the 5.0, 5.5, and 6.0 curves almost equidistant. The only cases where this difference is larger are in Grade 9–12 for Listening and Grades 11–12 for Reading.

Smoothing Procedures.

Speaking. CAL smoothed the cut scores resulting from Panelist judgments by fitting a quadratic polynomial function across the grade levels for each PL cut. The quadratic function may be represented as follows:

$Scale\ Score = Intercept_j + \beta_{1j} \times Grade + \beta_{2j} \times Grade^2$, where

$Intercept_j$ is a cut-based intercept;

β_{1j} and β_{2j} are cut-based quadratic coefficients.

Figure 15 shows the fitted curves of the cut scores, along with the exit cuts, and Table 16 presents the fitted cut score values across the grades. As expected, the polynomial smoothing helps to mitigate the reversed patterns in the cut score values observed from the unsmoothed results. Furthermore, the exit cuts resulting from Phase 1 of the standard setting would still be interpreted as between Levels 4 and 5.

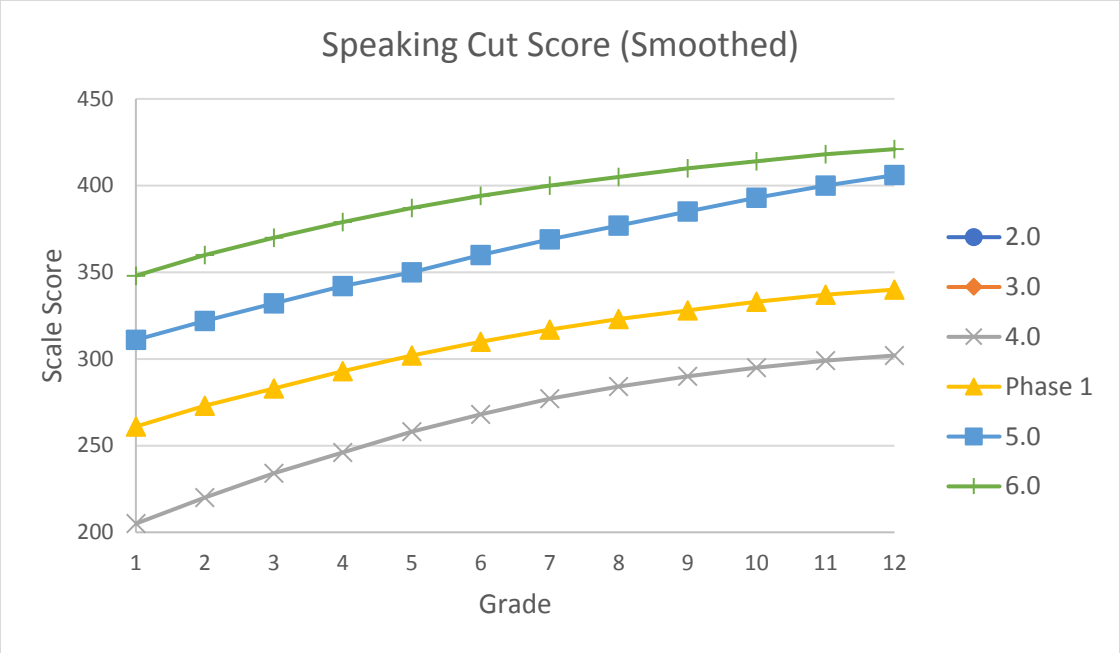


Figure 15. Quadratic function fitted curves for Speaking across grades.

Table 16. *Quadratic Function Fitted Cut Scores for Speaking across Grades (Final Recommended Cuts)*

Grade	WIDA Proficiency Levels					
	2.0	**3.0**	**4.0**	**Phase 1 cut**	**5.0**	**6.0**
1	205	261	311	348	361	403
2	220	273	322	360	374	415
3	234	283	332	370	386	425
4	246	293	342	379	397	435
5	258	302	350	387	407	443
6	268	310	360	394	417	451
7	277	317	369	400	425	457
8	284	323	377	405	433	463
9	290	328	385	410	440	468
10	295	333	393	414	446	471
11	299	337	400	418	451	474
12	302	340	406	421	455	476

Writing. A similar approach was used for Writing. Since the Writing domain consists of all performance tasks that are scored using a generic scoring scale, it is important for CAL's content experts to review the unsmoothed and smoothed cut scores in relation to the scoring scale. It was found that Grade 4 students can be awarded Level 2 if their average rating per task was below 1. This is deemed unacceptable from the content perspective. CAL recognized that the smoothing created a mathematical artifact where the Grades 4–5 cut scores were lowered significantly in order to fit the curve. In addition, there might have been some keyboarding effect in Grades 4–5 which might have led to Grades 4–5 tasks being more difficult than expected. In order to align the proficiency level cut to the scoring scale, CAL fixed the Grade 4 2.0 cut to the pre-smoothed value. Consequently, Grades 5 and Grades 6 2.0 cuts were also adjusted such that the final cut scores increase monotonically. Figure 16 and Figure 17 show the fitted curves of the cut scores, and Table 17 presents the fitted cut score values across the grades.

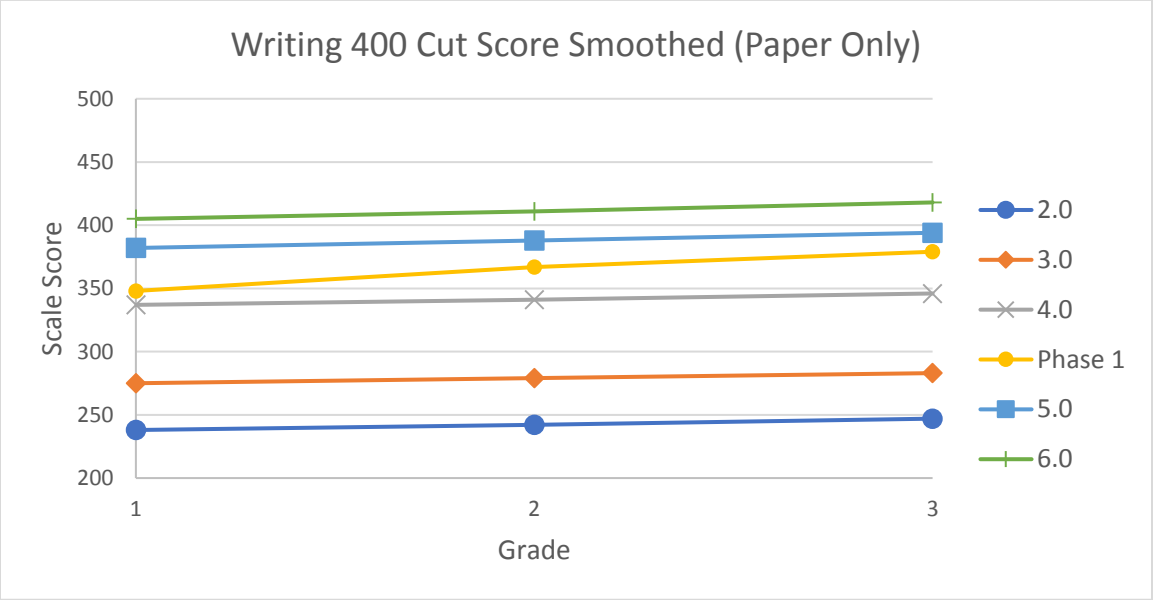


Figure 16. Quadratic function fitted curves for Writing for Grades 1–3 (paper).

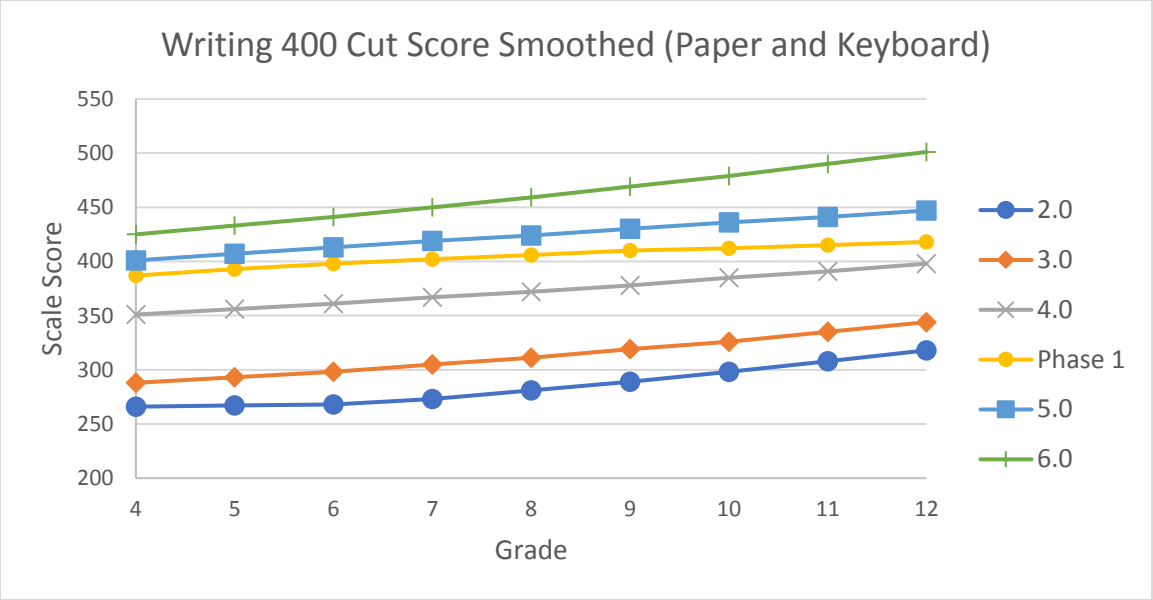


Figure 17. Quadratic function fitted curves for Writing for Grades 4–12 (paper and online).

Table 17. *Quadratic Function Fitted Cut Scores for Writing across Grades (Final Recommended Cuts)*

Grade	WIDA Proficiency Levels					
	2.0	**3.0**	**4.0**	**Phase 1 cut**	**5.0**	**6.0**
1	238	275	337	348	382	405
2	242	279	341	367	388	411
3	247	283	346	379	394	418
4	266	288	351	387	401	425
5	267	293	356	393	407	433
6	268	298	361	398	413	441
7	273	305	367	402	419	450
8	281	311	372	406	424	459
9	289	319	378	410	430	469
10	298	326	385	412	436	479
11	308	335	391	415	441	490
12	318	344	398	418	447	501

Listening and Reading. To capture the monotonically increasing curve of cut scores across grade, grade was transformed to LN(Grade+1). Taking the log ensures that the curve will increase monotonically (it can't decrease the way a polynomial could) and the +1 allows the final prediction function to be extended down to Kindergarten (panelists only considered Grade 1 and beyond). An interaction model was then applied, with each curve allowed to differ in slope:

Scale Score = Cut*LN(Grade+1) + LN(Grade+1) + Cut

Figure 18 and Figure 19 plot the model's fitted curves along with panelists' judgments. Table 18 and Table 19 show the values for the cuts resulting from this method.

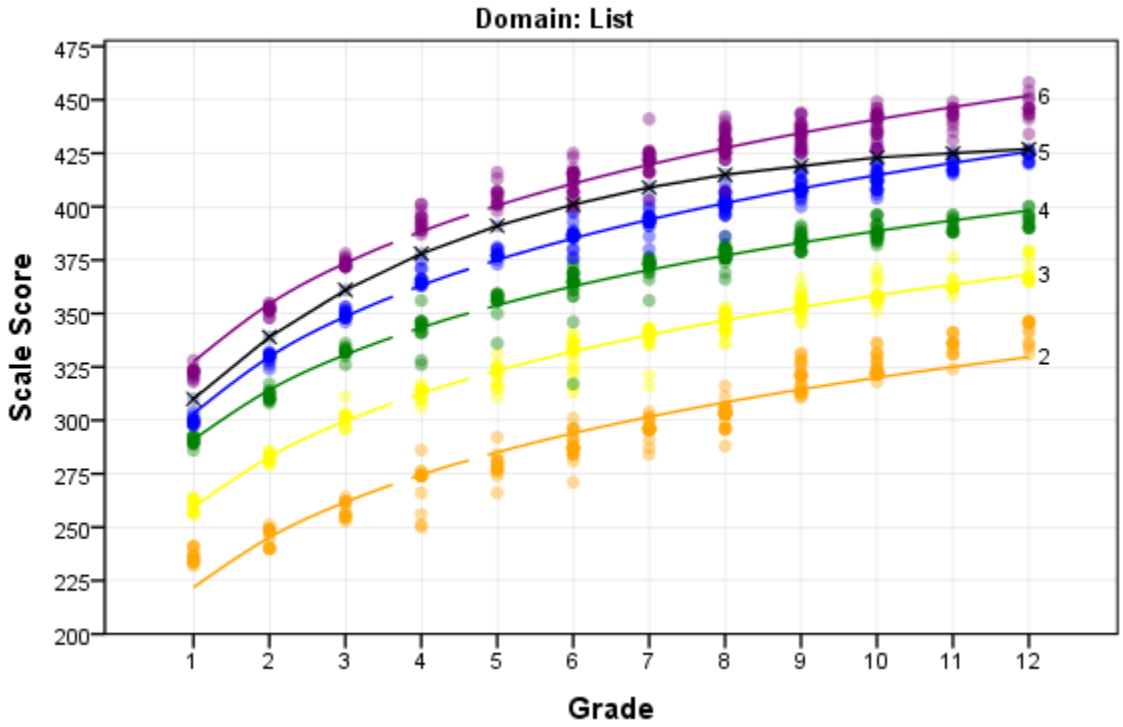


Figure 18. Interaction model fitted curves, with panelist values (•) and exit cuts (X) for Listening.

Table 18. *Interaction Model for Listening*

Grade	WIDA Proficiency Levels						
	2.0	**3.0**	**4.0**	**5.0**	**Phase 1 cut**	**5.5**	**6.0**
1	222	259	291	303	310	315	327
2	245	283	314	330	339	342	354
3	262	300	331	349	361	362	374
4	275	313	343	363	378	376	388
5	285	323	354	375	391	388	401
6	294	332	363	385	401	398	411
7	302	340	370	394	409	407	420
8	308	347	377	402	415	415	427
9	314	353	383	409	419	422	434
10	320	358	389	415	423	428	441
11	325	364	394	420	425	434	447
12	330	368	398	426	427	439	452

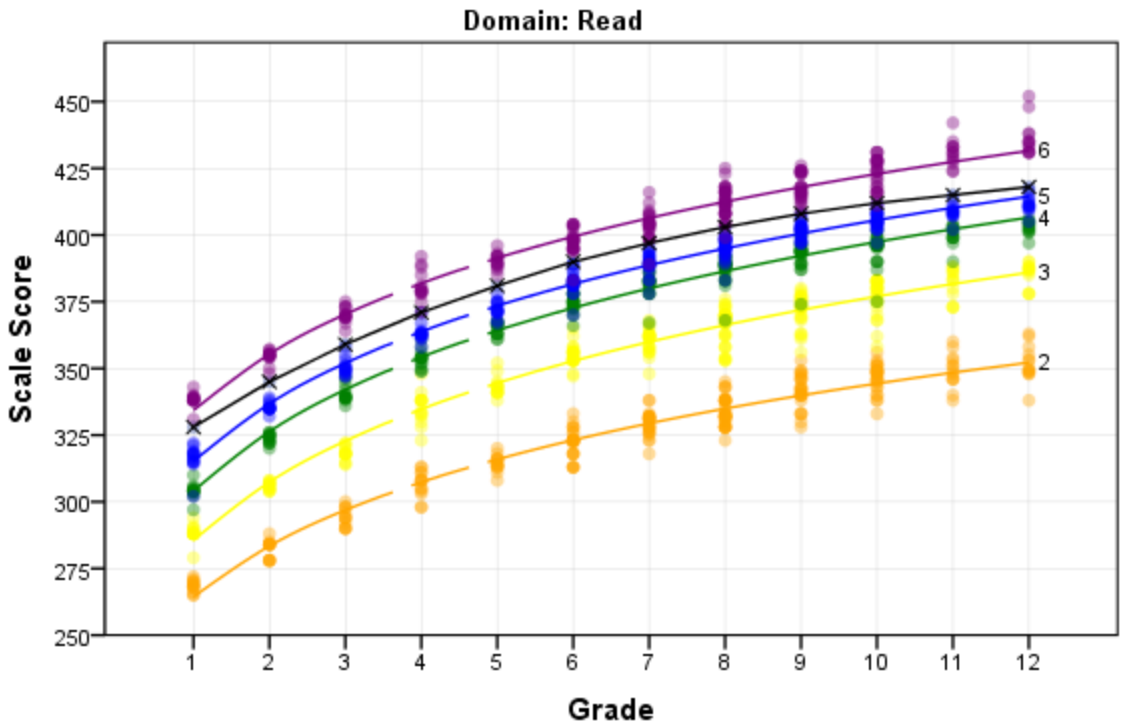


Figure 19. Interaction model fitted curves, with panelist values (•) and exit cuts (X) for Reading.

Table 19. *Interaction Model for Reading (Final Recommended Cuts)*

Grade	WIDA Proficiency Levels						
	2.0	3.0	4.0	5.0	Phase 1 cut	5.5	6.0
1	264	286	304	315	328	325	334
2	283	307	326	337	345	346	355
3	297	323	342	352	359	361	370
4	307	335	354	364	371	373	382
5	316	345	364	373	381	382	391
6	323	353	373	382	390	391	399
7	329	360	380	389	397	398	406
8	335	366	386	395	403	404	412
9	340	372	392	401	408	410	418
10	344	377	397	406	412	415	423
11	348	382	402	410	415	419	427
12	352	386	407	414	418	423	432

Further adjustments were made to the Listening cut scores to account for discrepancies between the Grade 1 2.0 cut derived from the model and the panelists' means. Specifically, the Grade 1 and Grades 10–12 values were set to the panelist means (see Figure 20 and Table 20).

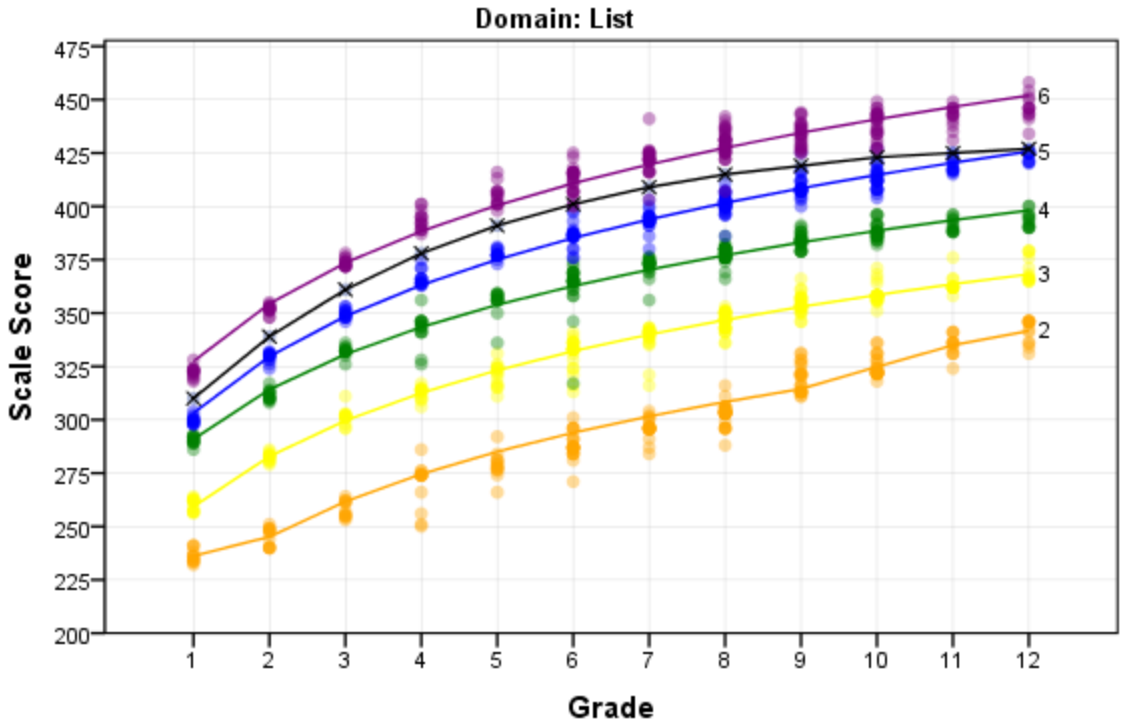


Figure 20. Amended interaction model fitted curves (for Cut 2.0, Grade 1 and Grades 10–12 cuts are set to panelist means), with panelist values (•) and exit cuts (X) for Listening.

Table 20. *Amended Interaction Model for Listening (Final Recommended Cuts)*

Grade	WIDA Proficiency Levels						
	2.0	**3.0**	**4.0**	**5.0**	**Phase 1 cut**	**5.5**	**6.0**
1	236	259	291	303	310	315	327
2	245	283	314	330	339	342	354
3	262	300	331	349	361	362	374
4	275	313	343	363	378	376	388
5	285	323	354	375	391	388	401
6	294	332	363	385	401	398	411
7	302	340	370	394	409	407	420
8	308	347	377	402	415	415	427
9	314	353	383	409	419	422	434
10	325	358	389	415	423	428	441
11	335	364	394	420	425	434	447
12	342	368	398	426	427	439	452

Creation of composites. In addition to domain scores, four weighted composite scores are calculated, as follows:

- Oral: Rounded up (0.5 * Listening + 0.5 * Speaking)
- Literacy: Rounded up (0.5 * Reading + 0.5 * Writing)
- Comprehension: Rounded up (0.3 * Listening + 0.7 * Reading)
- Overall: Rounded up (0.15 * Listening + 0.35 * Reading + 0.35 * Writing + 0.15 * Speaking)

To calculate the cut scores for these composites, the cut scores from the domains are plugged into the formulas. For example, to determine the Level 2.0 cut for Grade 1 Literacy, first find the 2.0 cuts for Grade 1 Reading and Grade 1 Writing (264 and 238). Then plug that into the formula above (Rounded up (0.5 * Reading + 0.5 * Writing)), and the result is 251. Table 21 through Table 24 show the resulting composite cuts.

Table 21. *Final Cut Scores: Oral*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	221	260	301	332	365
2	233	278	318	352	385
3	248	292	332	368	400
4	261	303	343	380	412
5	272	313	352	391	422
6	281	321	362	401	431
7	290	329	370	410	439
8	296	335	377	418	445
9	302	341	384	425	451
10	310	346	391	431	456
11	317	351	397	436	461
12	322	354	402	441	464

Table 22. *Final Cut Scores: Literacy*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	251	281	321	349	370
2	263	293	334	363	383
3	272	303	344	373	394
4	287	312	353	383	404
5	292	319	360	390	412
6	296	326	367	398	420
7	301	333	374	404	428
8	308	339	379	410	436
9	315	346	385	416	444
10	321	352	391	421	451
11	328	359	397	426	459
12	335	365	403	431	467

Table 23. *Final Cut Scores: Comprehension*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	256	278	300	311	332
2	272	300	322	335	355
3	287	316	339	351	371
4	297	328	351	364	384
5	307	338	361	374	394
6	314	347	370	383	403
7	321	354	377	391	410
8	327	360	383	397	417
9	332	366	389	403	423
10	338	371	395	409	428
11	344	377	400	413	433
12	349	381	404	418	438

Table 24. *Final Cut Scores: Overall*

Grade	WIDA Proficiency Levels				
	2.0	**3.0**	**4.0**	**5.0**	**6.0**
1	242	274	315	344	368
2	254	289	329	359	383
3	265	300	340	371	396
4	279	309	350	382	406
5	286	317	358	390	415
6	291	324	365	399	423
7	298	331	372	406	431
8	304	337	378	412	438
9	311	344	385	418	446
10	318	350	391	424	453
11	325	356	397	429	459
12	331	362	402	434	466

Training evaluation summary

During the standard setting meetings panelists were asked to complete an anonymous evaluation form on the training process, As in Phase 1, almost all Panelists rated the training portion of the Phase 2 meeting as *excellent* or *good*.

Table 25. *Phase 2 Training Evaluation Responses (53 Responses)*

Survey questions	Responses			
	Excellent	**Good**	**Fair**	**Poor**
The **purpose** of the standard setting meeting was communicated clearly	*36 (67.9%)*	*16 (41.5%)*		*1 (1.9%)*
The standard setting training was **organized** and **easy to follow**	*34 (64.2%)*	*18 (25%)*	*1 (1.9%)*	
I understand my **role and responsibilities** at this standard setting meeting	*37 (69.8%)*	*16 (30.2%)*		
The presenter **responded effectively** to audience questions and comments	*38 (73.1%)*	*13 (34%)*	*1 (1.9%)*	
I feel **prepared** to begin the standards setting process	*28 (54.8%)*	*22 (30.2%)*	*2 (3.8%)*	*1 (1.9%)*

In Phase 2, nine Panelists commented on the training process. These comments reflect that a third of the Panelists had attended the Phase 1 meeting training session.

> *Excellent organization and attention to detail. Excellent intro, background, and setting expectations.*

Very slow pacing

I think it helped that I had already attended the July session. I do believe you have taken issues from July into consideration and made adjustments accordingly.

Before coming to the meeting I did not receive any explanation of the purpose of the meeting so I didn't feel prepared to begin the process. But after I got here everyone looked to ensure I understood what I was doing.

Study evaluation summary

At the end of the Phase 1 and Phase 2 meetings panelists were asked to complete evaluations of the standard setting study, specifically regarding a) their opinion of the adequacy of the seven components of standard setting and, b) their confidence in the ability of the group's recommended cut scores to adequately and appropriately classify students as borderline English proficient. In both Phase 1 and Phase 2 panelists, on average, rated the adequacy of the study components as a little better than *good* in all four domains. In Phase 1 panelists, on average, rated their confidence in their recommended cut scores as 3.8, where (3 = Medium, 4 = High, and 5= Very High). In Phase 2 that average rating was 4.2. See Appendix U for a summary of panelist evaluation of, and comments on, the study.

Appendix A
Review of the Plan for the Standard Setting Study by Dr. Marianne Perie

To: Gary Cook
From: Marianne Perie
RE: Review of WIDA Standard Setting Plan
Date: March 27, 2016

After reviewing the plan for setting cut scores on the plan for setting cut scores on ACCESS 2.0, I think it looks like a solid approach. I have identified a few areas for consideration within each phase and those are detailed in this memo.

Phase I

The three biggest variables in any cut score methodology are the facilitator, the panelists, and the methodology. Not enough time is spent describing the panel, but great attention should be paid to recruiting them. The number of panelists should be identified. For the approach described, I recommend a minimum of 6 per grade, which would result in 72 total (2 per state) and 18 per panel. Then, you can have 3 groups of 6 within each panel, allowing for better estimates of both intra- and inter-panel variation. Additional information regarding the characteristics of the panelists is also needed. What groups are you trying to represent in what proportion? Multiple groups are mentioned, but not their experience with the target students. I would recommend that the largest group be educators (bullets 2 & 3). Policymakers are important, but should not dominate the group. In addition, demographic representativeness must also be considered. This could be difficult when dealing with state nominations. Perhaps consider asking states to nominate many more panelists than you need to help ensure a representative panel. You do not want individual state recruiting to result in a panel of white male policymakers.

The English Proficient activity is well designed. Encouraging discussion across all four modalities focusing specifically on conjunctive versus compensatory decisions will clarify the decisions they will be making.

For the bookmark procedure, the ordered item booklet (OIB) is described as being sorted from most to least difficult. Typically, OIBs with items sorted hardest to easiest tend to result in higher cut scores than when items are scored easiest to hardest. I would state that specifically and justify your decision to sort the items from most to least difficult. Perhaps, describe the policy purpose as being weighted against false positives (saying someone is English proficient when they are not), which is more damaging than false negatives for the English proficient cut for ELPAs.

I would reconsider sharing bookmark placements across tables after Round 1. To be able to give a real interpanel variance rating, withhold sharing bookmarks across tables until Round 3. This will allow you to distinguish within table variance from across table variance and isolate any effects of a single, strong-opinioned, panelist. I recommend this approach at least for the first grade. After the first grade is complete, your facilitators can determine if there is a panelist that could skew the cut scores. If there is, continue to keep tables separate. If it's not an issue, then go ahead and share ratings after Round 1 for the second grade.

For the listening piece, it will be important for panelists to take notes to aid in the discussion. Give them good instructions and a method for noting important features of the items to avoid multiple replays of the items.

For speaking and writing, consider showing a couple of examples of work around the same cut score with less consistent responses across tasks. The procedure to use a consistent set of responses makes sense for Round 1, but once panelists are circling a cut score, they should see other sets of responses with the same total score.

I wasn't clear on how the task of describing a minimally competent student in the composite task differs from the marginally English proficient discussion at the beginning? Is the focus more on grade-level content standards? The purpose of this task will need to be made clear to panelists.

To help with the composite cut scores, can you provide profiles of student results? For instance, provide a sample of some of the most common ways students achieve the initial cut score. Demonstrate several ways that a student can reach the same result. I recommend facilitating a discussion of whether just summing the four domains is sufficient or if a student should be able to perform above the minimum in at least one domain. The facilitation will be important in guiding this discussion.

Finally, I was not clear on what an "aggregate recommendation" entailed. Typically, we either work towards consensus or take the median as the final recommendation.

The plan for the Phase I process looks like a strong approach. In reviewing the agenda, I think you've given too little time for Steps 2 and 3 and too much time for Step 5, so be flexible about moving Step 4 to Day 2 as needed.

Phase II

In reviewing the Phase II description, I again found this to be a strong approach and have only a few recommendations.

Again, you need a description of how panels will be organized. I recommend 2-3 tables per group. Typically, tables are arranged with no more than eight and no fewer than five panelists per table to facilitate discussion in which everyone participates. I saw no mention of having stand-by panelists, but you may want to consider that as a backup plan to substitute in the event of a no show.

I did not see mention of performance level descriptions. There needs to be more explanation of how the panelists will be directed in defining the interpretation for the remaining performance levels. Specifically, what are they comparing the proposed cuts to? What is their specific judgmental task?

One recommendation is to ask panelists to provide written justifications for any changes to initial proposed cut scores. Those justifications should be based on the content requirement of the items. They can help test developers see key components of items that are important to student classification and can provide additional evidence for policymakers reviewing impact data.

Again, for speaking, it will be important to provide a note-talking mechanism. I would also like them to listen to a body of work with the same total score but demonstrating different ways to reach that score.

It is not clear from the description of the speaking portion if the panelists complete both grades 5 and 4 before seeing impact data. Another approach could be completing grades 5 and 3, reviewing impact data and then working on grades 4 and 2. I also was not clear on the type of articulation expected within and across grade clusters. It would seem that an indication of number of years in the EL system would be an important conditioning variable for understanding trends in impact data.

My only concern with the Phase 2 plan is with the use of a regression approach that includes initial and panelist recommended cut scores. Using a regression to smooth results with initial cuts and panelist cuts as input will de-emphasize panelist ratings. If you feel you have changed ACCESS, why not give panelist cuts extra weight since they have already seen (and ostensibly rejected) the equipercentile cuts. Also, why expect smooth results across grade bands? Wouldn't years in the US be a mitigating factor, rather than age?

It also appears that only Phase 1 considers the composite cut, which seems odd. Why not complete the regression analysis during the workshop and show Phase 2 panelists composite cut for other levels? Then, you could collect their reaction to the composite cut score in a survey to determine if it might need to be raised or lowered.

Further consideration of clarifying the target for panelists in Phase 2, and weighting their ratings over the initial cuts in the analyses is recommended. Otherwise, this standard setting plan is solid. The approach combines multiple aspects of research-based methods and follows best practices for standard setting.

Thank you for the opportunity to review this work.

Appendix B
Recruitment and Selection of Panelists

Recruitment of panelists for the Phase 1 and 2 meetings began January 14, 2016. WIDA member SEA representatives received a communication from WIDA Client Relations asking them to nominate educators (content teachers, teachers of English learners and students with disabilities, school administrators, district coordinators, and SEAs); see Appendix C. Nominees were sent a link to the Standard Setting Qualifications online survey asking about their education, credentials, experience, and current position; see Appendix D. A small number of nominees did not complete the qualifications survey, but answered the questions in a phone interview or were known to the selection group from previous work with WIDA. Recruitment continued through August 1, 2016 in order to give some states that did not initially nominate a panelist an opportunity to participate in the project, and to identify and recruit educators of specific grades or content areas for the Phase 2 meeting.

Table B1. *Panelist Participation by WIDA State*

	WIDA Member State	Phase 1 Panelists	Phase 2 Panelists	Panelists Attending Both
1	Alabama	1	1	1
2	Alaska	1	1	1
3	Colorado	1	2	0
4	CNMI	1	1	1
5	Delaware	2	0	0
6	District of Colombia	0	1	0
7	Florida	2	2	0
8	Georgia	3	0	0
9	Hawaii	2	2	2
10	Idaho	0	2	0
11	Illinois	6	3	2
12	Indiana	1	1	1
13	Kentucky	2	0	0
14	Maine	2	2	1
15	Maryland	2	4	1
16	Massachusetts	3	2	1
17	Michigan	1	1	1
18	Minnesota	1	1	0
19	Missouri	0	3	0
20	Montana	2	1	1
21	Nevada	3	3	0
22	New Hampshire	3	4	3
23	New Jersey	2	0	0
24	New Mexico	0	3	0
25	North Carolina	1	5	0
26	North Dakota	2	0	0
27	Oklahoma	2	1	0

	WIDA Member State	Phase 1 Panelists	Phase 2 Panelists	Panelists Attending Both
28	Pennsylvania	0	0	0
29	Rhode Island	1	1	0
30	South Carolina	0	0	0
31	South Dakota	3	0	0
32	Tennessee	2	1	0
33	US Virgin Islands	0	1	0
34	Utah	1	1	1
35	Vermont	2	1	0
36	Virginia	0	0	0
37	Wisconsin	2	2	1
38	Wyoming	2	1	1
	Total	**59**	**54**	**19**

Based on the Nominee Qualifications survey results, panelists were invited to participate in one, or both of the two standard setting meetings. All selected panelists had to have experience teaching or supporting English learners, and be very familiar or somewhat familiar with the WIDA ELD standards, the six WIDA proficiency levels, and the WIDA ACCESS for ELLs assessment.

Table B2. *Panelist Familiarity with the WIDA ELD Standards, the Six WIDA Proficiency Levels, and the WIDA ACCESS for ELLs Assessment*

Panelist Familiarity with:	Phase 1 (59 panelists)	Phase 2 (54 panelists)
WIDA ELD Standards		
Very familiar	78.6%	75.9%
Somewhat familiar	21.4%	24.1%
Not familiar	0%	0%
Six WIDA proficiency levels		
Very familiar	87.5%	87%
Somewhat familiar	12.5%	13%
Not familiar	0%	0%
WIDA ACCESS for ELLs		
Very familiar	87.5%	94.4%
Somewhat familiar	12.5%	5.6%
Not familiar	0%	0%

We selected educators who were (a) certified to teach all content areas and all grades, (b) had experience teaching Special Education, (c) were familiar with language service-based instruction, (d) who had at least two years of experience teaching, and (e) who were demographically diverse.

Table B3. *Panelist Teaching Certification or Specialization, Additional Certifications, and Highest Degree Obtained*

Certification, Specialization, or Degree	Phase 1 (59 panelists)	Phase 2 (54 panelists)
Teaching certification		
Elementary	37 (62.7%)	38 (70.4%)
Middle	32 (54.3%)	33 (61.1%)
Secondary	28 (47.5%)	26 (48.1%)
Content certification or specialization		
English Language Arts	14 (23.7%)	21 (38.9%)
Mathematics	3 (5.1%)	2 (3.7%)
Science	4 (6.8%)	3 (5.6%)
Social Studies	7 (11.9%)	7 (13%)
Special Education	5 (8.5%)	5 (9.3%)
ESL/Bilingual	47 (79.7%)	43 (79.6%)
Additional certification		
Administrator	18 (30.5%)	10 (18.5%)
School Psychologist	1 (1.7%)	0
Gifted/Talented	1 (1.7%)	0
Foreign Language (not English)	6 (10.2%)	13 (24.1%)
Highest level of education completed		
Bachelor's degree	2 (3.4%)	3 (5.6%)
Some graduate study	5 (8.5%)	7 (13%)
Master's degree	34 (57.6%)	30 (55.6%)
Some doctoral study	9 (15.3%)	8 (14.8%)
Doctoral degree	9 (15.3%)	6 (11.1%)

Note. Panelists often selected several certifications. Elementary and secondary certification includes middle school depending on the state or region. If a panelist indicated that their certification included 6-8 grades it is indicated in this table. Many ESL/Bilingual certifications are K-12, even if the content area certification is elementary or secondary. Elementary and middle certification is considered to cover Grades 1–8.

Table B4. *Panelists who Provided Instruction in These Language Services in Their Current Position or Job Title*

Language Services	Phase 1 (37 panelist teachers)	Phase 2 (36 panelist teachers)
Content area tutoring	10 (27%)	9 (25%)
Content-based ESL	22 (59.5%)	20 (55.6%)
Developmental bilingual programs (two-way bilingual programs)	1 (2.7%)	0
Heritage language	3 (8.1%)	0
Pull-out ESL	21 (56.8)	17 (47.2%)
Sheltered English instruction	19 (51.4%)	13 (36.1%)
Structured English Immersion / SDAIE	3 (8.1%)	2 (5.6%)
Transitional bilingual	3 (8.1%)	0
Dual language two-way immersion	2 (5.4%)	1 (2.8%)
No services	3 (8.1%)	5 (13.9%)
Other (i.e., supervise teachers, reading services)	9 (24.3%)	7 (19.4%)

Note. Panelists could select more than one service could be selected.

Table B5. *Panelist General Work Experience*

Experience	Phase 1 (59 panelists)	Phase 2 (54 panelists)
Number of years as educator		
Less than 2 years	0	0
2-5 years	1 (1.7%)	3 (5.6%)
6-10 years	10 (16.9%)	6 (11.1%)
11-20 years	24 (40.7%)	28 (51.9%)
21 or more years	24 (40.7%)	17 (31.5%)
Number of years worked for current school district/employer		
Less than 2 years	4 (6.9)	2 (3.7%)
2-5 years	13 (22.4%)	11 (20.4%)
6-10 years	17 (29.3%)	19 (35.2%)
11-20 years	19 (32.8%)	17 (31.5%)
21 or more years	5 (8.6%)	5 (9.3%)

Table B6. *Panelist Demographic Information*

Demographic	Phase 1 (59 panelists)	Phase 2 (54 panelists)
Gender		
Female	52 (88.1%)	50 (92.6%)
Male	7 (11.9%)	4 (7.4%)
Ethnicity (multiple responses allowed)		
American Indian or Alaska Native	1 (1.7%)	1 (1.9%)
Asian	4 (6.8%)	2 (3.7%)
Black/African American	3 (5.1%)	1 (1.9%)
Pacific Islander/Native Hawaiian	0 (0%)	0 (1.7%)
Hispanic or Latino	13 (22%)	14 (25.9%)
White	42 (71.2)	42 (77.8%)
Native language		
English	49 (83.1%)	45 (83.3%)
Spanish	6 (10.2%)	9 (16.7%)
Other	5 (8.5%)	3 (5.6%)

Note. The larger greater percentage of male panelists in Phase 1 is due to SEA representatives participating as panelists. According to the most recent population survey released by the U.S. Bureau of Labor, male educators constitute just 3.2% of pre-K and Kindergarten teachers, 19.3% of the elementary and middle school teacher population, and 40.8% of the high school level teaching staff. These numbers suggest a clear female majority in the teaching profession, especially in the earlier grades.

Appendix C
Email Sent to WIDA Member State Education Agencies requesting Panelist Nominations

WIDA is soliciting from all 38 WIDA states participants in a Standard Setting Study for ACCESS for ELLs 2.0, and to date, we don't have a representative slated from (insert state).

Participants should have experience teaching or directly supporting the instruction of ELs in K–12 public schools. Sixty participants are being sought for each of two phases of WIDA ACCESS for ELLs 2.0 2016 Standard Setting activities. These meetings will be held to identify exit criterion scores on ACCESS 2.0 (Phase 1) and to determine where to set cut scores between the six proficiency levels described by the WIDA English Language Development (ELD) standards for grades 1–12 (Phase 2). The dates for the Phase 1 meetings are July 12th-July 13th, and for Phase 2 Aug. 10th, 11th, and 12th (1/2 day on Aug. 12th). Both meetings will be in Madison. WIDA will cover all travel, lodging, and incidentals.

Would you please think of an educator you would like to nominate? I remember as an educator how much I loved opportunities to travel, collaborate with others, especially when all expenses were paid! [Name], copied in this email, will follow-up with whomever as soon as he receives your nominee's contact information.

We definitely don't want to miss the voice of (insert state)!!

Appendix D
ACCESS 2.0 Standard Setting Study Nominee Qualifications Survey

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

ACCESS for ELLs 2.0 Standard Setting Study

Your state department has nominated you to be a panelist at the upcoming ACCESS for ELLs 2.0 Standard Setting Study. WIDA is required to fully document the personal and professional background of each panelist in order to show that selected panelists represent the appropriate variety of expertise, content area specialties, and racial/ethnic group.

The following survey should take you 5-10 minutes to complete, and will provide WIDA with the appropriate information to make selections of panelists from the nominations that states have made. Even if you have already responded to an email from *[name]* at WIDA with your qualifications, please complete this survey as it is more encompassing that what was previously asked. If you have any questions, contact *[name and contact information].*

Please enter your name below.

Do you, or have you worked with ELLs?

- ❍ Yes
- ❍ No

If No Is Selected, Then Skip To: Thank you for completing this survey....

In which state do you work?

- ❍ Alabama
- ❍ Alaska
- ❍ Colorado
- ❍ Delaware
- ❍ District of Columbia
- ❍ Florida
- ❍ Georgia
- ❍ Hawaii
- ❍ Idaho
- ❍ Illinois
- ❍ Indiana
- ❍ Kentucky
- ❍ Maine
- ❍ Maryland
- ❍ Massachusetts
- ❍ Michigan
- ❍ Minnesota
- ❍ Missouri
- ❍ Montana
- ❍ Nevada
- ❍ New Hampshire
- ❍ New Jersey
- ❍ New Mexico
- ❍ North Carolina
- ❍ North Dakota
- ❍ Northern Mariana Islands
- ❍ Oklahoma
- ❍ Pennsylvania
- ❍ Rhode Island
- ❍ South Carolina
- ❍ South Dakota
- ❍ Tennessee
- ❍ U.S. Virgin Islands
- ❍ Utah
- ❍ Vermont
- ❍ Virginia
- ❍ Wisconsin
- ❍ Wyoming

Please list the school(s) and/or district(s) that you work with. (N/A if not applicable)

What is your current position? Eg., District Programs Administrator, Principal, Resource Educator, EL Teacher, Mathematics or Science Teacher, etc.

What other relevant positions have you held in the past? E.g., District Programs Administrator, Principal, Resource Educator, EL Teacher, Mathematics or Science Teacher, etc.

How many years of experience do you have in your...

Current position?
Past position? (List position and number of years; N/A if not applicable)

What are your roles in your current position? E.g., I support Title I Teachers as an EL Resource Teacher, I teach mathematics.

What were your roles in your past position(s)? E.g., I support Title I Teachers as an EL Resource Teacher, I teach mathematics.

Please list your current qualifications. E.g., high school science certification. (N/A if not applicable)

Please select the answer that most closely describes your current title(s). (You may select more than one)

- ❑ ELL Teacher
- ❑ Content Teacher
- ❑ Special Education Teacher
- ❑ School Administrator (E.g., Principal, Assistant Principal, School Curriculum Coordinator)
- ❑ District Administrator (E.g., Superintendent, District Curriculum Coordinator, Federal Program Support Administrator)
- ❑ State Administrator

Display This Question:

If Please select the answer that most closely describes your past title(s):
ELL Teacher Is Selected
Or Content Teacher Is Selected
Or Special Education Teacher Is Selected

Please indicate the content areas that you currently teach by selecting the corresponding grade levels (1-12) to the right of the content areas. (You may select more than one grade and/or content area)

	1	2	3	4	5	6	7	8	9	10	11	12
English Language Arts	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Mathematics	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Science	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Social Studies	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Other	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

Please select the answer that most closely describes your past title(s). (You may select more than one)

- ❑ ELL Teacher
- ❑ Content Teacher
- ❑ Special Education Teacher
- ❑ School Administrator (E.g., Principal, Assistant Principal, School Curriculum Coordinator)
- ❑ District Administrator (E.g., Superintendent, District Curriculum Coordinator, Federal Program Support Administrator)
- ❑ State Administrator

Display This Question:

If Please select the answer that most closely describes your past title(s):
ELL Teacher Is Selected
Or Content Teacher Is Selected
Or Special Education Teacher Is Selected

Please indicate the content areas that you taught in the past by selecting the corresponding grade levels (1-12) to the right of the content areas. (You may select more than one grade and/or content area)

	1	2	3	4	5	6	7	8	9	10	11	12
English Language Arts	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Mathematics	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Science	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Social Studies	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Other	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

Do you... (you may select more than one)

	Yes	No
Have experience working with ELs with disabilities?	❍	❍
Teach general ELL classes?	❍	❍

Please select your race/ethnicity (optional response). You may select more than one.

- ❑ Hispanic/Latino
- ❑ American Indian or Alaska Native
- ❑ Asian
- ❑ Black or African American
- ❑ Native Hawaiian or Other Pacific Islander
- ❑ White
- ❑ Other _____

Thank you for completing this survey. A WIDA project team member will be in touch with you at the end of April or beginning of May to let you know if you have been selected to participate in the ACCESS for ELLs 2.0 Standard Setting Study.

Appendix E
Standard Setting Communication Campaign Tools and Resources

Tool/Resource	Audience	Communication Location
Standard Setting Recommendations memo to SEAs	SEAs	SEA Secure Portal of the WIDA website
Standard Setting flyers: • Overview • Teacher • Administrator • School psychologist and Spec. Educ. teacher	SEAs, LEAs, teachers, administrators, school psychologists, special education teachers	Public WIDA website
SEA and LEA webinars	SEAs, LEAs, educators	SEA Secure Portal and Public WIDA website
State-Specific follow-up webinar	Specific to educators in a member state	Specific state page on wida.us
Getting Students Ready for Testing flyers (online and paper)	LEAs, educators	Public WIDA website
Impact Data	SEAs	secure data transfer protocol
ACCESS for ELLs 2.0: New Rigorous Score Scale video and Discussion Guide	SEAs, LEAs, educators	Public WIDA website
Interpretive Guide for Score Reports- section on Standard Setting	SEAs, LEAs, educators	Public WIDA website
U.S. Education Department Letter	U.S. Education Department	Email communication
Legislators Letter	SEAs	SEA Secure Portal of WIDA website
Superintendents and School Boards Letter	SEAs	SEA Secure Portal of WIDA website
Slides for talking to parents	LEAs, parents, families	Public WIDA website
Slides for talking to other educators	LEAs, educators	Public WIDA website
Score lookup calculator and FAQ	SEAs, LEAs, educators	Secure portal of WIDA website
Understanding your child's 2017 ACCESS for ELLs 2.0 scores	Parents and families	Public WIDA website
2017 Individual Student Score Notes	Educators, Parents and families	Public WIDA website
SEA Exit Criteria Memo	SEAs	SEA Secure Portal
Summary of ACCESS 2.0 Standard Setting Study	SEAs	Secure portal of WIDA website
2017 Speaking Scores flyer	SEAs, LEAs, educators	Public WIDA website
SEA outreach plan and checklist	SEAs	Public WIDA website
LEA outreach plan and checklist	LEAs	Public WIDA website
Printed and thumb drive resource kits for SEAs	SEAs	Mailed to SEAs

Appendix F
Assignment of Panelists to Groups and Tables

The target number of panelists was 60 for each meeting, divided into four groups, each including panelists representing five grade levels. Group 1 represented Grade Levels 1–5, Group 2 Grade Levels 4–8, Group 3 Grade Levels 6–10, and Group 4 Grade Levels 8–12. Each group was further divided into three tables of four or five panelists each. The Phase 1 meeting ended up having 59 panelists due to a late cancellation, and the Phase 2 meeting had 54 panelists due to the shutdown of a major airline and some late cancellations.

Groups

For each phase of the project, panelists were assigned to the appropriate group based on their experience teaching the corresponding grade levels. In addition, teachers of English learners and of all content areas in each of the five grades levels in each group, and panelists representing each of the WIDA member states were assigned across groups. Panelists were asked in the Standard Setting Qualifications online survey if they have experience working with English learners with disabilities and 83% responded *yes* in Phase 1, and 93% in Phase 2. In addition, educators with specialization or credentials in special education (five in both Phase 1 and Phase 2) were assigned to each group.

Tables

For Phase 1 the goal was to select one SEA, LEA, educator with special education experience, two content teachers, and multiple ethnicities per each of the three tables in a group. Table assignments of Phase 2 panelists differed only in that SEAs and LEAs were not specifically recruited for that meeting.

If it was not possible to have all these targets represented in a specific table, they were represented in the larger group. For example in Table F2, Phase 1, Group 1, Table 2 there wasn't an SEA participating, but there were two other SEAs in the group. Members of all tables in a group were able to participate in and contribute to the whole-group discussions.

Table F1. *Number of Years Currently Teaching Grades 1–12*

Grade	Number of Years Teaching Grade													
	Phase 1							Phase 2						
	<2	2-5	6-10	11-20	>=21	Total	NA	<2	2-5	6-10	11-20	>=21	Total	NA
1	2	6	9	2	0	19	11	1	5	5	4	0	15	28
2	2	5	8	1	1	17	10	1	5	4	4	0	14	28
3	3	7	6	3	0	19	9	1	8	3	3	0	15	28
4	4	7	5	1	0	17	10	1	8	3	3	0	15	28
5	1	4	7	1	0	13	10	1	7	3	3	0	14	28
6	2	5	6	2	0	15	9	3	3	4	5	0	15	27
7	1	5	4	2	0	12	11	3	3	2	6	0	14	27
8	2	7	4	2	0	15	10	3	3	2	6	0	14	27
9	5	1	4	6	0	16	10	4	4	0	3	0	11	27
10	6	0	5	6	0	17	10	4	4	0	3	0	11	27
11	2	0	5	6	0	13	11	2	5	0	3	0	10	27
12	1	0	4	6	1	12	11	2	5	0	3	0	10	27

Note. NA is the number of panelist that identified as non-teaching current positions.

Table F2. *Distribution of Panelists to Groups and Tables in the Phase 1 Standard Setting Meeting*

Group	Table (#)	SEA	LEA	Content Areas										ELL SWD	Grade Clusters								Non-white
				ESL		ELA		SCI		MA		SS			1-5		4-8		6-10		8-12		
				Now	Past	Now	Past	Now	Past	Now	Past	Now	Past		Now	Past	Now	Past	Now	Past	Now	Past	
1	1 (4)	1	2	1	3	1	3	1	2	1	3	1	2	4	0	4	0	3	0	0	0	0	2
(1-5)	2 (5)	0	3	5	0	2	3	0	4	1	3	1	4	3	4	1	0	1	0	0	0	0	3
	3 (5)	1	1	4	1	2	3	1	3	2	2	2	3	5	2	3	1	2	0	1	0	1	2
TOTAL	**(14)**	**2**	**6**	**10**	**4**	**5**	**9**	**2**	**9**	**4**	**8**	**4**	**9**	**12**	**6**	**8**	**1**	**6**	**0**	**1**	**0**	**1**	**7**
2	1 (5)	1	2	4	1	4	1	1	2	2	1	2	1	4	3	2	2	2	2	1	1	0	1
(4-8)	2 (5)	1	2	2	3	1	4	0	3	0	3	0	4	3	2	2	0	5	0	2	0	0	1
	3 (5)	2	2	1	4	1	1	0	3	0	3	0	2	2	1	1	0	2	0	1	0	2	2
TOTAL	**(15)**	**4**	**6**	**7**	**8**	**6**	**6**	**1**	**8**	**2**	**7**	**2**	**7**	**9**	**6**	**5**	**2**	**9**	**2**	**4**	**1**	**2**	**4**
3	1 (5)	3	0	3	2	1	1	0	1	0	2	0	1	4	0	2	2	2	2	1	1	2	1
(6-10)	2 (5)	2	1	4	1	2	1	0	1	0	1	1	1	5	1	1	0	2	1	3	1	3	2
	3 (5)	0	2	4	0	2	2	1	2	1	2	2	2	5	0	2	2	2	3	1	0	0	2
TOTAL	**(15)**	**5**	**3**	**11**	**3**	**5**	**4**	**1**	**4**	**1**	**5**	**3**	**4**	**14**	**1**	**5**	**4**	**6**	**6**	**5**	**2**	**5**	**5**
4	1 (5)	1	2	3	1	2	2	2	1	0	2	1	1	5	2	0	2	2	1	2	2	3	2
(9-12)	2 (5)	1	3	4	1	1	2	0	1	0	2	1	1	5	0	0	1	3	1	3	0	4	2
	3 (5)	1	2	5	0	2	0	1	0	1	0	1	1	4	1	0	1	0	2	1	2	1	2
TOTAL	**(15)**	**3**	**7**	**12**	**2**	**5**	**4**	**3**	**2**	**1**	**4**	**3**	**3**	**14**	**3**	**0**	**4**	**5**	**4**	**6**	**4**	**8**	**6**
Mtg Total	**59**	**14**	**22**	**40**	**17**	**21**	**23**	**7**	**23**	**8**	**24**	**12**	**23**	**49**	**16**	**18**	**11**	**26**	**12**	**16**	**7**	**16**	**22**

Note. In the *Table* column the number in () is the number of panelists at this table. Red numbers are content areas or grades taught in the past. Panelists identified as SEAs or LEAs are currently in those positions. Panelists were asked if they had experience working with English learners with disabilities (ELL SWD), so this column does not distinguish between current or past experience. Cells shaded in blue represent the grade cluster for the corresponding group.

Table F3. *Distribution of Panelists to Groups and Tables in the Phase 2 Standard Setting Meeting*

Group	Table (#)	SEA	LEA	Content Areas										ELL SWD	Grade Clusters								Non-white
				ESL		ELA		SCI		MA		SS			1-5		4-8		6-10		8-12		
				Now	Past	Now	Past	Now	Past	Now	Past	Now	Past		Now	Past	Now	Past	Now	Past	Now	Past	
1	1 (4)	1	2	1	2	1	3	1	2	1	2	0	3	4	1	3	1	2	0	0	0	0	1
(1-5)	2 (5)	0	2	2	1	2	2	1	2	1	3	1	2	5	3	2	1	2	0	0	0	0	2
	3 (5)	0	1	2	1	3	2	2	2	2	2	3	1	5	3	2	1	2	0	1	0	1	2
TOTAL	**(14)**	**1**	**5**	**5**	**4**	**6**	**7**	**4**	**6**	**4**	**7**	**4**	**6**	**14**	**7**	**7**	**3**	**6**	**0**	**1**	**0**	**1**	**5**
2	1 (5)	0	2	5	0	4	1	1	2	1	2	1	2	4	4	1	3	1	2	0	0	0	2
(4-8)	2 (4)	1	0	3	1	1	1	0	2	0	2	0	1	4	2	1	1	2	0	1	0	0	1
	3 (5)	1	2	1	2	0	3	0	3	0	2	0	3	4	1	2	0	4	0	1	0	1	2
TOTAL	**(14)**	**2**	**4**	**9**	**3**	**5**	**5**	**1**	**7**	**1**	**6**	**1**	**6**	**12**	**7**	**4**	**4**	**7**	**2**	**2**	**0**	**1**	**5**
3	1 (4)	0	0	3	0	2	0	0	1	1	0	2	0	3	0	3	3	1	3	1	2	1	2
(6-10)	2 (4)	0	1	2	0	2	0	0	1	0	1	0	1	3	1	1	2	0	3	1	1	1	0
	3 (5)	0	2	2	1	3	1	0	1	0	1	1	0	5	1	0	1	4	3	2	1	1	1
TOTAL	**(13)**	**0**	**3**	**7**	**1**	**7**	**1**	**0**	**3**	**1**	**2**	**3**	**1**	**11**	**2**	**4**	**6**	**5**	**9**	**4**	**4**	**3**	**3**
4	1 (4)	0	3	3	0	2	1	1	0	0	1	2	0	4	1	0	2	0	1	1	0	3	3
(9-12)	2 (4)	0	3	1	1	0	4	0	1	0	1	0	2	4	0	0	0	2	0	3	0	4	1
	3 (5)	0	1	4	0	0	2	0	1	0	1	0	3	5	0	1	0	1	1	1	3	2	1
TOTAL	**(13)**	**0**	**7**	**8**	**1**	**2**	**7**	**1**	**2**	**0**	**3**	**2**	**5**	**13**	**1**	**1**	**2**	**3**	**2**	**5**	**3**	**9**	**5**
Mtg Total	**54**	**3**	**19**	**29**	**9**	**20**	**20**	**6**	**18**	**6**	**18**	**10**	**18**	**50**	**17**	**16**	**15**	**21**	**13**	**12**	**7**	**14**	**18**

Note. In the *Table* column the number in () is the number of panelists at this table. Red numbers are content areas or grade taught in the past. Panelists identified as SEAs or LEAs are currently in those positions. Panelists were asked if they had experience working with English learners with disabilities (ELL SWD), so this column does not distinguish between current or past experience. Cells shaded in blue represent the grade cluster for the corresponding group.

Appendix G

ACCESS 2.0 Standards Setting Phase 1 Meeting Agenda

Wisconsin Center for Education Research (WCER)
University of Wisconsin-Madison
1025 West Johnson Street, MD #23
Madison, WI 53706

866.276.7735 • www.wida.us • info@wida.us

ACCESS 2.0 Standards Setting Phase I Meeting
Data Recognition Corporation (DRC) Scoring Center, 208 E. Olin Ave., Madison, WI
July 12-13, 2016
Agenda

Objectives of the Meeting: To determine an English proficiency score below which EL students should not be reclassified and what proficiency level value that score should represent.

Tuesday, July 12th

Breakfast at the Concourse Hotel

9:00 a.m. – 10:30 a.m.	Training: Describe what English proficiency means
10:30 a.m. – 10:45 a.m.	**Break**
10:45 a.m. – 12:00 p.m.	Round 1: Identify the minimum English proficiency level score
12:00 p.m. – 1:00 p.m.	**Lunch**
1:00 p.m. – 2:30 p.m.	Round 2: Consider Round 1 discussions
2:30 p.m. – 2:45 p.m.	**Break**
2:45 p.m. – 4:00 p.m.	Round 3: Review impact data

Wednesday, July 13th

Breakfast at the Concourse Hotel

9:00 a.m. – 10:30 a.m.	Round 3 continued
10:30 a.m. – 10:45 a.m.	**Break**
10:45 a.m. – 12:00 p.m.	Round 4: Final deliberations
12:00 p.m. – 1:00 p.m.	**Lunch**
1:00 p.m. – 3:00 p.m.	Overall composite and domain cut-score discussions

Appendix H

ACCESS 2.0 Standards Setting Phase 1 Table Leader Training Agenda

Wisconsin Center for Education Research (WCER)
University of Wisconsin-Madison
1025 West Johnson Street, MD #23
Madison, WI 53706

866.276.7735 • www.wida.us • info@wida.us

ACCESS 2.0 Standards Setting Table Leader Training
Data Recognition Corporation (DRC) Scoring Center, 208 E. Olin Ave., Madison, WI
July 11, 2016
Agenda

Monday, July 11th

1:30 p.m. – 1:30 p.m.	Introductions
1:30 p.m. – 2:00 p.m.	Role of Table Leaders
2:00 p.m. – 2:30 p.m.	Training
2:30 p.m. – 2:45 p.m.	**Break**
2:45 p.m. – 4:00 p.m.	Training

Appendix I

ACCESS 2.0 Standards Setting Phase 2 Meeting Agenda

WIDA

Wisconsin Center for Education Research (WCER)
University of Wisconsin-Madison
1025 West Johnson Street, MD #23
Madison, WI 53706

866.276.7335 • www.wida.us • info@wida.us

ACCESS 2.0 Standards Setting Phase II Meeting
Data Recognition Corporation (DRC) Scoring Center, 208 E. Olin Ave., Madison, WI
August 10-12, 2016
Agenda

Objectives of the Meeting: To advise WIDA on the proficiency level cut scores for all domains.

DAY 1: Wednesday, August 10th

7:30 a.m. – 8:30 a.m.	**Breakfast at the Sheraton**
9:00 a.m. – 9:15 a.m.	Welcome and Introductions
9:15 a.m. – 10:30 a.m.	Background and Overview • Review of Phase 1 meeting • Training on Body of Work procedures
10:30 a.m. – 10:45 a.m.	**Break (divide into groups)**
10:45 a.m. – 12:00 p.m.	Speaking
12:00 p.m. – 1:00 p.m.	**Lunch**
1:00 p.m. – 2:30 p.m.	Speaking
2:30 p.m. – 2:45 p.m.	**Break**
2:45 p.m. – 4:45 p.m.	Speaking

DAY 2: Thursday, August 11th

7:30 a.m. – 8:30 a.m.	**Breakfast at the Sheraton**
9:00 a.m. – 10:30 a.m.	Writing
10:30 a.m. – 10:45 a.m.	**Break**
10:45 a.m. – 12:00 p.m.	Writing
12:00 p.m. – 1:00 p.m.	**Lunch**
1:00 p.m. – 1:30 p.m.	Training on Bookmarking procedures
1:30 p.m. – 2:30 p.m.	Reading
2:30 p.m. – 2:45 p.m.	**Break**
2:45 p.m. – 4:45 p.m.	Reading

DAY 3: Friday, August 12th

7:30 a.m. – 8:30 a.m.	**Breakfast**
9:00 a.m. – 10:30 a.m.	Listening
10:30 a.m. – 10:45 a.m.	**Break**
10:45 a.m. – 12:00 p.m.	Listening
12:00 p.m. – 1:00 p.m.	**Lunch**

Appendix J

WIDA ACCESS for ELLs Test Nondisclosure Agreement (NDA)

The same NDA was used for Table Leaders in Phase 1 and for Panelists in Phase 2, but the dates and meeting title (highlighted section) were modified.

WIDA ACCESS for ELLs® Test Nondisclosure Agreement

The Board of Regents of the University of Wisconsin System (the "Regents"), on behalf of the WIDA Consortium ("WIDA") is the copyright owner of the secure, unpublished ACCESS for ELLs® test (the "Test"), as defined in 37 C.F.R. § 202.20(b)(4). WIDA treats the Test as constituting valuable and proprietary trade secret. WIDA maintains the Test as confidential and secure, and only provides access to the Test to individuals who are legally bound to maintain the confidentiality and security of the Test.

In consideration for my participation as a Panelist in the ACCESS 2.0 Standards Setting Phase 1 meeting, July 12-13, 2016, I,

_____ (print name) of

_____ (print organization)

agree as follows:

1. The Test is confidential, proprietary information and material of the Regents.
2. I shall use the Test only in my role as a Panelist according to the directions supplied to me by the meeting facilitator.
3. I shall not copy, modify or distribute any components of the Test for any purpose.
4. The Test must be kept secure and confidential, as disclosure of the Test to third parties could adversely affect the validity of the test items, results or the commercial value of the Test. I shall keep all Test materials secure and confidential at all times in accordance with any instructions that I receive from the Test facilitator.

_____ _____

Signature Date

_____ _____

WIDA meeting facilitator acknowledgment Date

Appendix K

Phase 1 Panelist Training Presentation

WIDA Consortium Phase 1 Standard Setting Training

Tuesday, July 12, 13, 2016
DRC Scoring Facility
Madison, WI

Overview

- Why are we here?
- What are we supposed to do?
- Steps in the Standard Setting process
- The Standard Setting User Interface (SSUI)

Why do a standard setting?

- A confluence of factors motivate the need to re-examine the meaning of ACCESS for ELLs proficiency levels, e.g.,
  - Migrating from a paper-and-pencil mode of assessment
  - Refining the speaking assessment design and using centralized scoring
  - The influence of career and college ready academic language use expectations
- The process used to re-examine proficiency levels is called "Standard Setting" or "Setting Performance Standards."

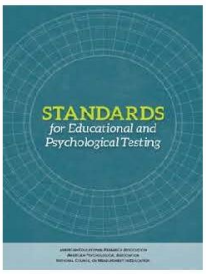
3

Quotable Quotes

- "The setting of performance standards is a blend of judgment, psychometrics, and practicality" (Hambleton & Pitoniak, 2006).
- "Cut scores embody value judgments as well as technical and empirical considerations" (Testing Standards, 2014).
- Performance standards are socially-constructed opinions of student performance, often informed by empirical information.
- "It is a process by which qualified panelists, following carefully developed and documented procedures that mitigate against arbitrariness, assign interpretative meaning to performances on tests" (Kenyon & Römhild, 2013).

12/19/16 4

The assessment professions' rulebook or "Testing Standards"



Definitions from Testing Standards

- **Performance level**: a label or statement classifying a test taker's competency in a particular domain.
- **Performance-level descriptor**: a description of what test takers know and can do at specific performance levels.
- **Cut-score**: a specified point on a score scale, such that

12/18/16 6

Testing Standards

- **Standard 5.21**: When proposed interpretations involve one or more cuts scores, the **rationale and procedures** used for establishing cut scores should be **documented clearly**.

12/19/16 7

Testing Standards

- **Standard 5.22:** When cut scores defining pass-fail or proficiency levels are based on direct judgments about the adequacy of item or test performances, 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.

This is why you're here!!

12/19/16 8

Testing Standards

- **Standard 5.23**: When feasible and appropriate, cut scores defining categories with distinct substantive interpretations should be **informed by sound empirical data** concerning the relation of **test performance to the relevant criteria**.

12/19/16 9

To summarize …

- We need a documented process and a procedure for identifying cutscores.
- We need expert participants to use their judgment to identify these cutscores.
- We need to use available, relevant information on test performance to use inform cutscore selection.

Steps in the Standard Setting Process

- Understand the assessment, its purposes and the performance level descriptors
- Build a description of the borderline student, i.e., the student who is just ready be reclassified as no longer an EL
- Engage in standard setting rounds
  - Bookmark for reading & listening
  - Body of Work for writing & speaking

ACCESS for ELLs Purposes

- Identify students' current level of English language proficiency along the developmental continuum.
- Identify students' progress in attaining English proficiency
- One of multiple measures used to determine whether students are prepared to exit English language support programs.
- Provides districts with information that will help them evaluate the effectiveness of their ESL/bilingual programs.

ACCESS for ELLs 2.0

Listening	Process, understand, interpret and evaluate spoken language in a variety of situations
Reading	Process, understand, interpret and evaluate written language, symbols and text with understanding and fluency
Writing	Engage in written communication in a variety of situations for a variety of purposes and audiences
Speaking	Engage in oral communication in a variety of situations for a variety of purposes and audiences

ACCESS for ELLs 2.0 Online Grade Clusters	1	2-3	4-5	6-8	9-12
ACCESS for ELLs 2.0 Paper Grade Clusters	1	2	3	4-5	6-8	9-12

***Note: Kindergarten will remain an interactive, paper-based kit.**

WIDA Performance Definitions – Listening and Reading Grades K-12

	Discourse Dimension Linguistic Complexity	Sentence Dimension Language Forms & Conventions	Word/Phrase Dimension Vocabulary Usage
At each grade, toward the end of a given level of English language proficiency, and with instructional support English learners will process…			
Level 6 Reaching	▯ Text and speech from a wide range of perspectives and registers (e.g., sarcasm, humor) ▯ Abstract and often nuanced language of each content area	▯ Dense complex sentences indicative of particular text types & genres ▯ A full range of grade-level sentence patterns associated with each content area	▯ Extensive technical, specific & general language of content area topics and themes ▯ Subtle content area words, terms, and expressions
Level 5 Bridging	▯ Rich descriptive discourse with complex sentences Cohesive and organized related ideas	▯ Compound, complex grammatical constructions (e.g., multiple phrases and clauses) A broad range of sentence patterns characteristic of particular content areas	▯ Technical and abstract content-area language Words and expressions with shades of meaning across content areas
Level 4 Expanding	▯ Connected discourse with a variety of sentences ▯ Expanded related ideas	▯ A variety of complex grammatical constructions ▯ Sentence patterns characteristic of particular content areas	▯ Specific and some technical content-area language ▯ Words or expressions with multiple meanings across content areas

WIDA Performance Definitions – Speaking and Writing Grades K-12

	Discourse Dimension Linguistic Complexity	Sentence Dimension Language Forms & Conventions	Word/Phrase Dimension Vocabulary Usage
At each grade, toward the end of a given level of English language proficiency, and with instructional support English learners will produce…			
Level 6 Reaching	▯ Text and speech with a tightly woven flow of sentences matched to purpose, situation, and audience ▯ Linguistically dense text and speech to communicate complex ideas associated with each content area	▯ A variety of grammatical structures indicative of particular genres ▯ A full range of sentence patterns (e.g., embedded clauses & complex noun phrases) and conventions associated with each content area	▯ Technical, specific, & general language of content area topics and themes ▯ Words, terms, and expressions with precise connotative meaning
Level 5 Bridging	▯ Multiple, complex sentences ▯ Organized, cohesive, and coherent expression of ideas	▯ A variety of grammatical structures matched to purpose ▯ A broad range of sentence patterns characteristic of particular content areas	▯ Technical and abstract content-area language, including content-specific collocations ▯ Words and expressions with precise meaning across content areas
Level 4 Expanding	▯ Short, expanded, and some complex sentences ▯ Organized expression of ideas with emerging cohesion	▯ A variety of grammatical structures ▯ Sentence patterns characteristic of particular content areas	▯ Specific and some technical content-area language ▯ Words and expressions with expressive meaning through use of collocations and idioms across content areas

Background for the discussion of the borderline English proficient student

What is an English learner according to ESSA?

An EL is a student whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny [that student]

- the ability to meet the State's proficient level of achievement;
- the ability to successfully achieve in classrooms where the language of instruction is English; or
- the opportunity to fully participate in society

ESSA § 8101

By extension, an English proficient student …

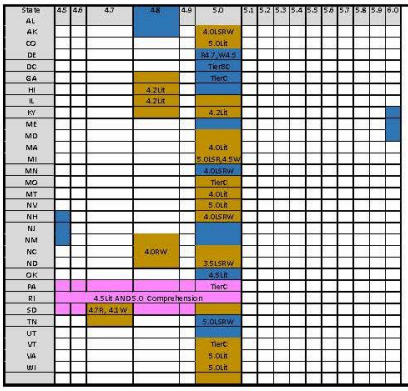
- Has the ability to perform proficiently on state assessments in English;
- **Can meaningfully participate in an English-only classroom, and**
- **Has the opportunity to fully participate in society where the common language is English.**

What would a borderline student (at the grade cluster under consideration) look like in terms of English language proficiency in:

Reading
Listening
Writing,
and
Speaking?

Here are some things to think about…

- What language is needed in the new college and career standards?
- English proficient does NOT mean fluent.
- OCR and DOJ think of English proficient as performing similarly to non-EL classroom peers.
- English proficient does NOT mean proficient in ELA.
- We are deliberating & recommending scores below which we believe ELs should not be reclassified.
- It might be helpful to look at what other states do.

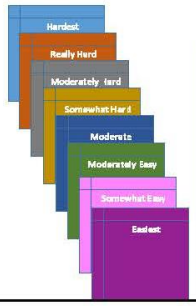


WIDA States' Exit Criteria

Most states have an Overall CPL of 5.0

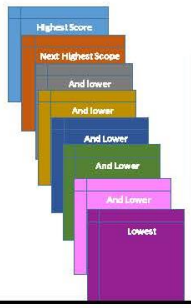
The minimum domain or literacy cuts range from 3.5 to 5.0

Bookmark Method



The goal is to identify the item where 2/3 or more of the borderline students will successfully answer, starting from the hardest item going downward.

Body of Work Method



The goal is to identify which student portfolio best represents the writing or speaking of a borderline student.

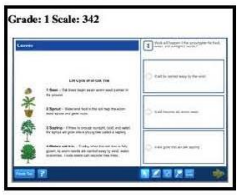
Put together, we have student portfolios.

Description of Rounds

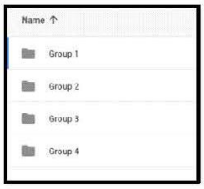
- Preliminaries
  - Do all the grades in a domain before going to another one (reading, listening, writing and speaking)
  - Do the highest grade first, then the next highest, then the next until done
- Domain Rounds
  - Round 1: Initial assignment of cuts
  - Round 2: After looking at group statistics and discussion adjust or confirm cuts
  - Round 3: After looking at impact data and discussion adjust or confirm cuts
- Overall Composite Rounds
  - Round 1: After looking at impact data and discussion assign initial cuts
  - Round 2: After looking at group statistics and discussion adjust or confirm cuts

Online Standard Setting

Standard Setting User Interface

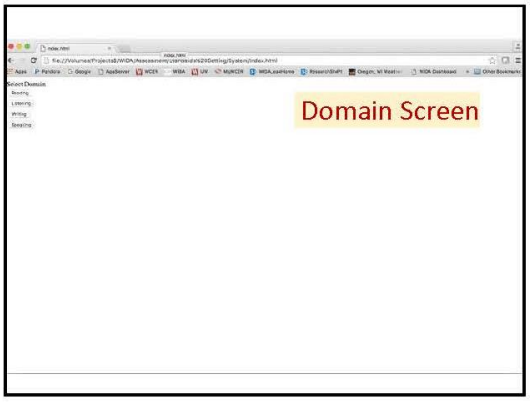


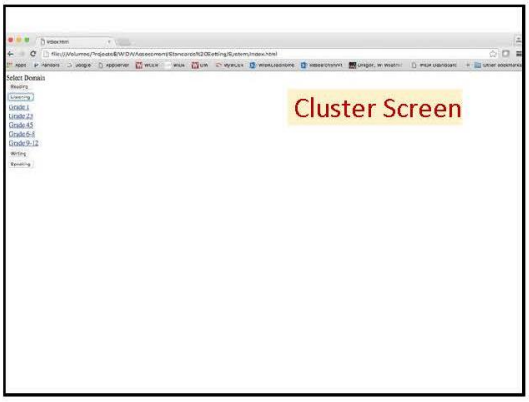
Panelist Score Entry Form

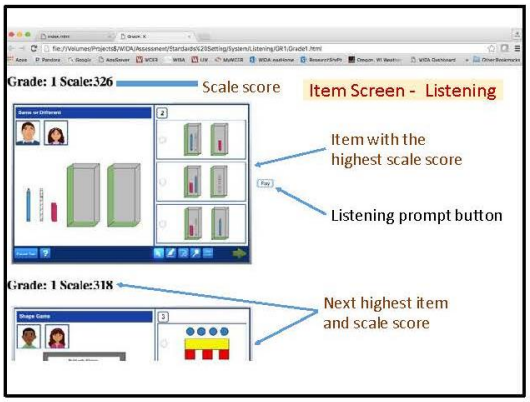


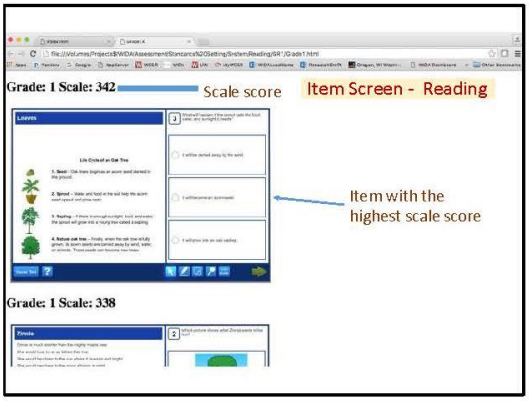
Standard Setting User Interface

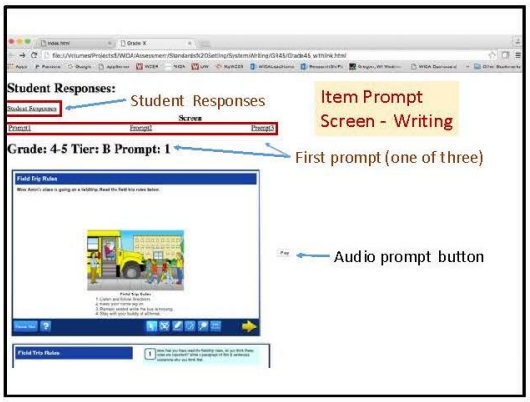
- Domain Screen
- Cluster Screen
- Rating Screens
  - Ordered items (reading and listening)
  - Student portfolios (writing and speaking)

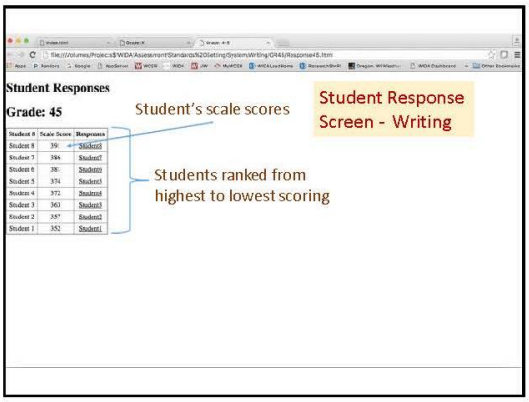


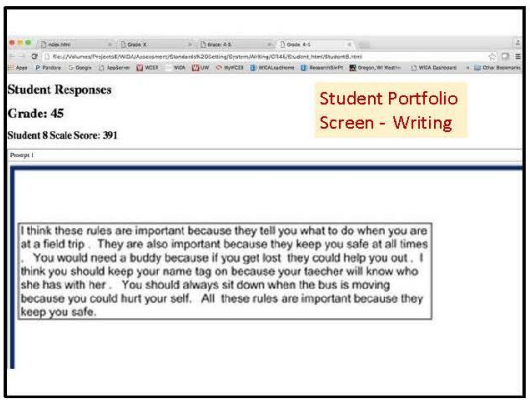


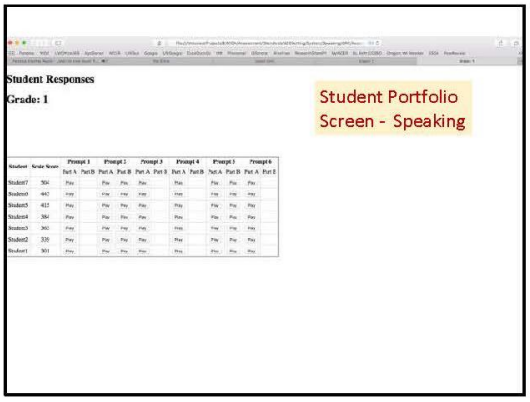








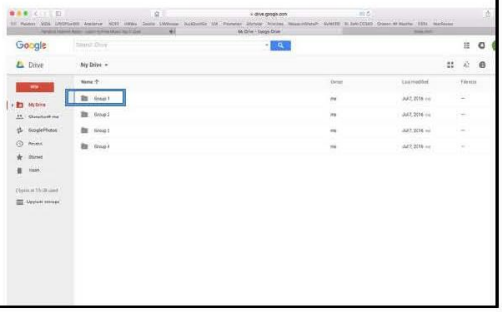




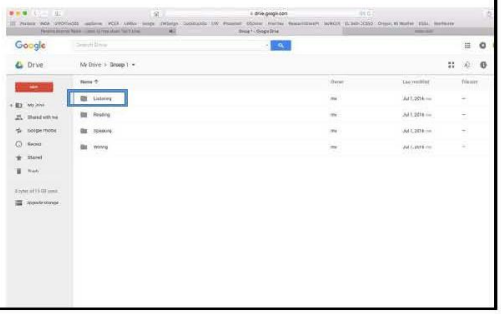
Panelist Score Entry Form

- This online tool is used in addition to recording your scores on paper
- The online tool aggregates information across tables

Recording Scores During Rounds



Recording Scores During Rounds



Your Task

- Think of a student or students (you know) in the target grade level currently functioning at the **borderline** between being classified as an EL and becoming reclassified
- **What are the language behaviors that this student shows in speaking, in writing, in reading and listening?**
  - e.g., writing, while not perfect, is understandable and similar to non-EL peers in grammar, organization and vocabulary.
  - e.g., speaking, while not necessarily fluent, engages comfortably in classroom and social discussions on a variety of topics

Your Task for reading & listening

- Work on your own through the items in the User Interface, starting with the first (hardest) item
- For each item, ask yourself, **What language skills are needed to correctly respond to this item?**
- Next, ask yourself: **Would the borderline student have at least a 67% probability (a two out of three chance) of getting this item correct?** (That is, more likely to get it correct rather than incorrect) Or, think of 100 borderline students: would 67 get it right and 33 get it wrong?)
- If no, move on to the next item, until your answer is **YES.**

Your Task for reading & listening

- Then look at the next easiest item(s) and confirm that you continue to say Yes to the question: **Would the borderline student have more than a 67% probability (a two out of three chance) of getting this item correct?**
- As you continue working through the items, feel free to finalize the placement of you bookmark **after you review all the items**
- Record your scale score both on the provided paper and in the online panelist score entry form

Your Task for writing & speaking

- Work on your own through the student portfolios, starting with the highest scoring student going downward
- For each set of prompts, ask yourself: **What language skills are needed to respond to these prompts?**
- Next, ask yourself: **Would 67% of the borderline students write or speak like this?** Or, think of 100 borderline students: would 67 of them write or speak like this?
- If no, move on to the student portfolio, until your answer is **YES.**

Your Task for writing & speaking

- Then look at the next lowest portfolio and confirm that you continue to say Yes to the question: **Would 67% (2 out of 3) of borderline students write or speak like this?**
- As you continue working through the portfolios, feel free to finalize the placement **after you review all student portfolios**
- Record your scale score both on the provided paper and in the online panelist score entry form

Assignments

- Four groups
  - Group 1: grades 1 to 5
  - Group 2: grades 4 to 8
  - Group 3: grades 6 to 10
  - Group 4: grades 8 to 12
- Three tables in each group
- Four or five panelists at each table, one of whom is the table leader
- Two facilitators in each group

Responsibilities be for TABLE LEADERS

- Facilitate the standard setting process for their table
  - Make sure group panelists understand their roles and responsibilities
  - Make sure panelists understand and can use the standard setting user interface (SSUI)
  - Make sure panelists have meaningful discussions…
    - Of WIDA PLDs for their grades and domains
    - Of the minimally competent student
    - During each round of the standard setting process
  - Ask the group facilitator to help with questions the table leader cannot answer
- Serve as a panelist

Responsibilities for PANELISTS

- Participate in training
- Ask questions to clarify understanding
- Complete standard setting surveys
- Review all items within an ordered item book (reading and listening) for assigned grades
- Review all student portfolios (writing and speaking) for assigned grades

Appendix L
Phase 1 Group Training Presentation

WIDA ACCESS 2.0 Standard Setting Group Presentation

July 12, 13, 2016

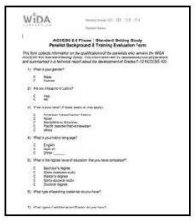
Madison, WI

Introductions – Getting to know you

- Name?
- What state are you from?
- How long have you been in your current field?
- What other educational roles have you filled?
- Experience with large-scale assessments, especially ACCESS for ELLs
  - Administering assessments
  - Item writing
  - Local or consortium committees
  - Alignment studies
  - Other standard setting

Preliminaries

- Fill out the demographic and training evaluation survey.
- It allows us to record the background of who was at the standard setting
- It provides you the opportunity to comment on the standard setting training.
- We will be filling out other surveys after we finish each domain.



Responsibilities be for TABLE LEADERS

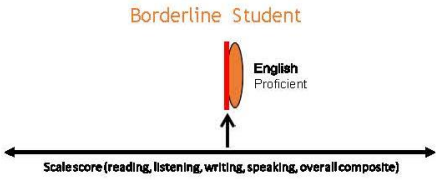
- Facilitate the standard setting process for their table
- Serve as a panelist

Responsibilities for PANELISTS

- Participate in training.
- Ask questions to clarify understanding.
- Complete standard setting surveys.
- Review all items within an ordered item book (reading and listening) for assigned grades.
- Review all student portfolios (writing and speaking) for assigned grades.
- Assign scores.

Our Task

- Identify the scale score in reading, listening, writing, speaking, and the overall composite that represents a borderline English proficient student.



Our Agenda

- Discuss the borderline English proficient student
  - Discuss what it means to be a borderline student within each cluster (1, 2-3, 4-5, 6-8, or 9-12).
  - Record key words or phrases that identify these students' language behavior.
- Conduct standard setting rounds
  - Starting with the highest grade first
  - Do all grades in reading first
  - Then do the same for listening, writing, and speaking in that order
  - Last of all do the overall composite

The Borderline Student

- At your table, write out statements that describe the characteristics of the borderline student in each domain, within each cluster.
- Start with speaking, then writing, listening and reading
- You may want to use the WIDA performance definitions to help create your statements.
- The following slide provides examples of things you might want to write...

Borderline Student Example Statements

- For Speaking, this student
  - While not necessarily fluent, engages comfortably in classroom and social discussions on a variety of topics
  - Uses a variety of sentence structures with varied levels of complexity
  - Uses grade appropriate academic vocabulary
- For Listening, this student, is able to comprehend
  - Practically all social and academic discussions
  - A range of sentence patterns characteristic of particular content areas
  - Words and expressions with shades of meaning across content areas

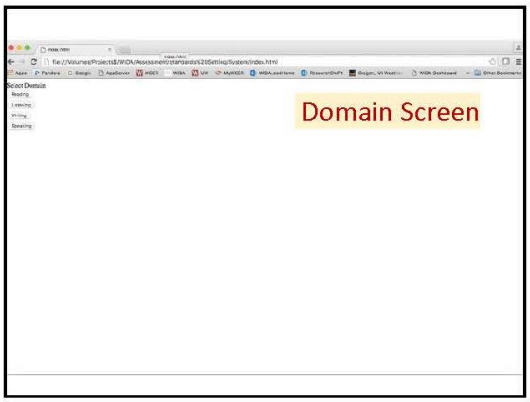
Borderline Student Activity

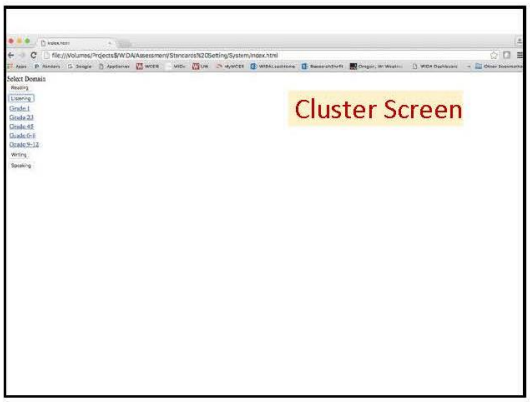
- At your table, quickly draft statements for each domain, for each cluster. Limit your self to only 1 or 2 statements for each domain/cluster. (15 minutes)
  - Your table leader will facilitate this discussion.
  - Be prepared to share your statements.
- Share your statements, one domain at a time (25 minutes)
- Discuss and agree upon the statements you want to write on the poster paper to represent the borderline student's language. (20 minutes)

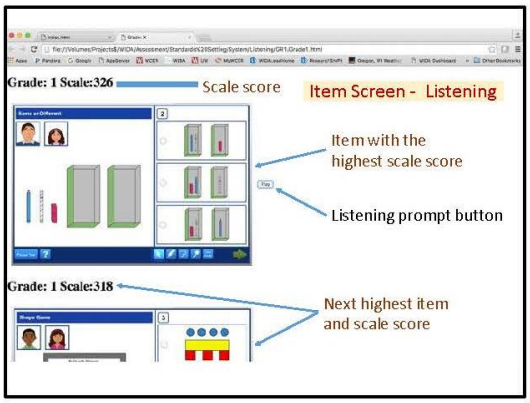
Tools you will need for the Standard Setting

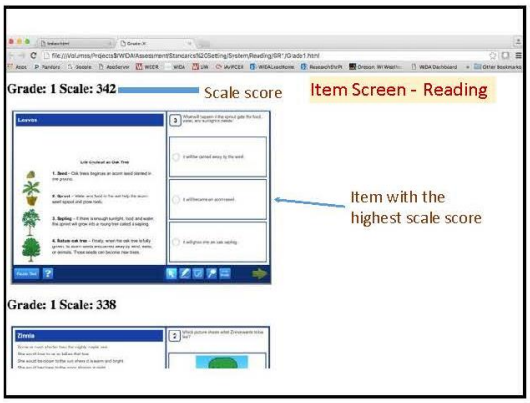
- Description of the borderline student
- A computer with..
  - The Standard Setting User Interface (SSUI)
  - Online Panelist Score Entry Form
- A paper rating form

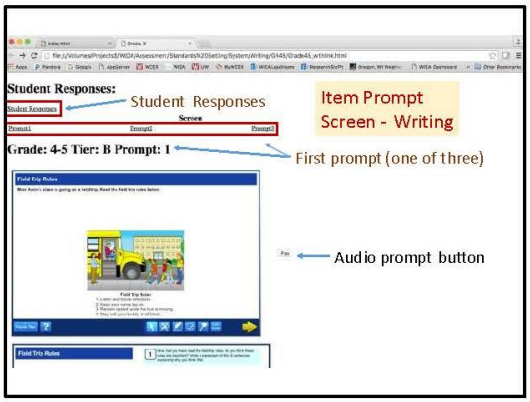
Let's take some time to familiarize ourselves with the Standard Setting User Interface.

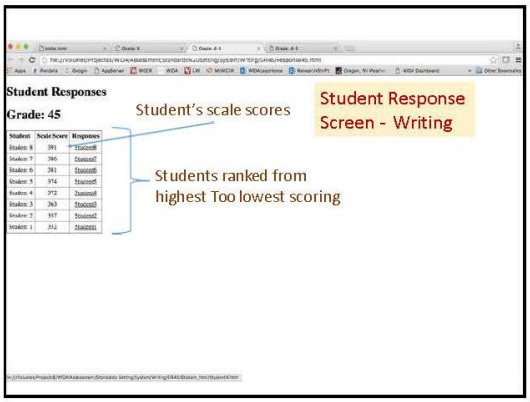

Domain Screen

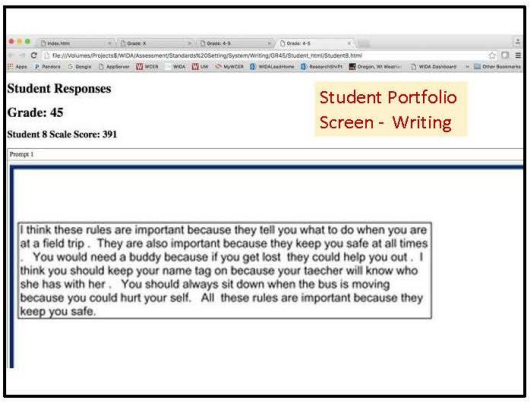

Cluster Screen


Grade: 1 Scale:326
Scale score
Item Screen - Listening
Item with the highest scale score
Listening prompt button
Grade: 1 Scale:318
Next highest item and scale score


Grade: 1 Scale: 342
Scale score
Item Screen - Reading
Item with the highest scale score
Grade: 1 Scale: 338

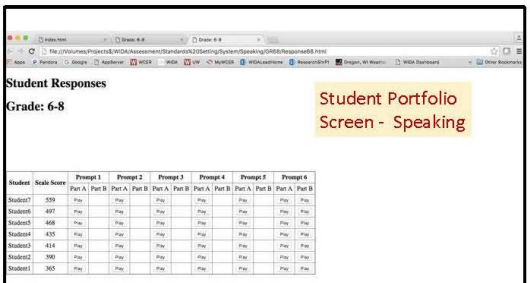

Student Responses:
Student Responses
Item Prompt Screen - Writing
Grade: 4-5 Tier: B Prompt: 1
First prompt (one of three)
Audio prompt button


Student Responses
Grade: 45
Student's scale scores
Student Response Screen - Writing
Students ranked from highest Too lowest scoring



Student Portfolio Screen - Writing

Item Prompt Screen - Speaking

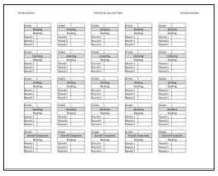


Student Portfolio Screen - Speaking

Let's take some time to familiarize ourselves with the Panelist Score Entry Forms

Panelist Entry Forms

Paper Form



Online Form

We would like panelists to enter information onto both forms.

Now Let's begin Standard Setting

Your Task for reading & listening

- Work on your own through the items in the User Interface, starting with the first (hardest) item
- For each item, ask yourself, **What language skills are needed to correctly respond to this item?**
- Next, ask yourself: **Would the borderline student have at least a 67% probability (a two out of three chance) of getting this item correct?** (That is, more likely to get it correct rather than incorrect) Or, think of 100 borderline students: would 67 get it right and 33 get it wrong?)
- If no, move on to the next item, until your answer is **YES**.

Your Task for reading & listening

- Then look at the next easiest item(s) and confirm that you continue to say Yes to the question: **Would the borderline student have more than a 67% probability (a two out of three chance) of getting this item correct?**
- As you continue working through the items, feel free to finalize the placement of you bookmark **after you review all the items**
- Record your scale score both on the provided paper and in the online panelist score entry form

Reading Standard Setting

Reading Round 1

- Review the description for the Borderline student.
- Look at all the items in the SSUI.
- Select the score you think represents the Borderline student at the highest grade in your group.
- Record your scale score on the paper and online forms.

Reading Round 1 Results

- Review your group's Round 1 statistics:

Grades	Median	Min	Max
12, 10, 8, 5			
11, 9, 7 ,4			
10, 8, 6, 3			
9, 7, 5, 2			
8, 6, 4, 1			

- At your table discuss Round 1 results
  - Is the median Too high? Too low? Just right?
  - What do you think about the range in scores?

Reading Round 2

- Review the description for the Borderline student.
- Look at all the items in the SSUI.
- Select the score you think represents the Borderline student considering Round 1 results.
- Record your scale score on the paper and online forms.

Round 2 Results & Impact Data

Grades	Median	Min	Max
12, 10, 8, 5			
11, 9, 7 ,4			
10, 8, 6, 3			
9, 7, 5, 2			
8, 6, 4, 1			

- Based on the Group's median____% of students would meet this score.
- At your table discuss Round 2 results & impact data.

Reading Round 3

- Review the description for the Borderline student.
- Look at all the items in the SSUI.
- Select the score you think represents the Borderline student considering Round 2 results & impact data.
- Record your scale score on the paper and online forms.
- You are done with this grade; go on to other grades.

Listening Standard Setting

Listening Round 1

- Review the description for the Borderline student.
- Look & listen to all the items in the SSUI.
- Select the score you think represents the Borderline student at the highest grade in your group.
- Record your scale score on the paper and online forms.

Listening Round 1 Results

Grades	Median	Min	Max
12, 10, 8, 5			
11, 9, 7 ,4			
10, 8, 6, 3			
9, 7, 5, 2			
8, 6, 4, 1			

- At your table discuss Round 1 results
  - Is the median Too high? Too low? Just right?
  - What do you think about the range in scores?

Listening Round 2

- Review the description for the Borderline student.
- Look at all the items in the SSUI.
- Select the score you think represents the Borderline student considering Round 1 results.
- Record your scale score on the paper and online forms.

Round 2 Results & Impact Data

Grades	Median	Min	Max
12, 10, 8, 5			
11, 9, 7 ,4			
10, 8, 6, 3			
9, 7, 5, 2			
8, 6, 4, 1			

- Based on the Group's median____% of students would meet this score.
- At your table discuss Round 2 results & impact data.

Listening Round 3

- Review the description for the Borderline student.
- Look at all the items in the SSUI.
- Select the score you think represents the Borderline student considering Round 2 results & impact data.
- Record your scale score on the paper and online forms.
- You are done with this grade; go on to other grades.

Writing Standard Setting

Writing Round 1

- Review the description for the Borderline student.
- Look & listen to all the prompts in the SSUI before you look at student portfolios.
- Now look at the student portfolios.
- Select the score of the student you think represents the Borderline case starting at the highest grade in your group.
- Record your scale score on the paper and online forms.

Writing Round 1 Results

- Review your group's Round 1 statistics:

Grades	Median	Min	Max
12, 10, 8, 5			
11, 9, 7 ,4			
10, 8, 6, 3			
9, 7, 5, 2			
8, 6, 4, 1			

- At your table discuss Round 1 results
  - Is the median Too high? Too low? Just right?
  - What do you think about the range in scores?

Writing Round 2

- Review the description for the Borderline student.
- Scan all portfolios in the SSUI.
- Select the score you think represents the Borderline student considering Round 1 results.
- Record your scale score on the paper and online forms.

Appendix M

Phase 2 Training Presentation and Handout

CAL CENTER FOR APPLIED LINGUISTICS

ACCESS 2.0 Standard Setting: Phase 2

Madison, WI
August 10-12, 2016
Prepared by: David MacGregor

WIDA CONSORTIUM

Ch... ch... changes

Reading	
ACCESS 1.0	ACCESS 2.0 Series 400
Tiered	Multi-stage Adaptive

Listening	
ACCESS 1.0	ACCESS 2.0 Series 400
Tiered	Multi-stage Adaptive
Teacher-read script	Pre-recorded, Media Delivered

- Same Scale
- Same item format
- Same folder structure

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Why do we need a new standard setting anyway?

Over the past decade:

- The Consortium has expanded from the original 3 states to 38 US State Education Agencies
- Educators within the Consortium have gained experience and expertise in applying the Standards
- College and career readiness standards have come into effect
- A common understanding of ELLs has been established for the Consortium (Phase1)
- The test has changed

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Ch... ch... changes

Writing	
ACCESS 1.0	ACCESS 2.0 Series 400
Tiered	Tier selection based on L and R test
Handwritten responses	Grades 1-3: Handwriting only
	Grades 4-5: Handwriting or keyboarding, at discretion of SEAs and/or local educators
	Grades 6-12: Keyboarding as default
18-point scoring rubric (0-0-0 to 6-6-6)	11 point scoring rubric (0, 1, 1+..., 6)

- Centrally scored
- Same Scale
- Same item format
- Same folder structure

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Ch... ch... changes

Overall	
ACCESS 1.0	ACCESS 2.0 Series 400
Paper-based	Online
Grade-level clusters:	Grade-level clusters:
K	K
1-2	1
	2-3
3-5	4-5
6-8	6-8
9-12	9-12

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Ch... ch... changes

Speaking	
ACCESS 1.0	ACCESS 2.0 Series 400
One-on-one interview	Pre-recorded prompts
Opportunity for follow-up questions	Use of model responses
Adaptive	Tier selection based on L and R test
Responses rated live by administrator	Responses recorded and centrally scored
	New scale

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Focus on borderline student: S and W

At each grade, toward the end of a given level of English language proficiency, and with instructional support, English learners will produce ...

Level 4 Expanding	▪ Short, expanded, and some complex sentences ▪ Organized expression of ideas with emerging cohesion	▪ A variety of grammatical structures ▪ Sentence patterns characteristic of particular content areas	▪ Specific and some technical content-area language ▪ Words and expressions with expressive meaning through use of collocations and idioms across content areas

- By this definition, a student at PL 5.0 should be able to **produce** language with all of these characteristics

Our support for you

- Student performances on the test (S and W)
- Test items (R and L)
- Our proposals for cuts (W, R and L)
  - You will confirm or adjust those proposals
- Performance Definitions
- The ELD Standards
- The online Standard Setting User Interface (SSUI)
- This training

Focus on borderline student: R and L

At each grade, toward the end of a given level of English language proficiency, and with instructional support, English learners will process...

Level 4 Expanding	▪ Connected discourse with a variety of sentences ▪ Expanded related ideas	▪ A variety of complex grammatical constructions ▪ Sentence patterns characteristic of particular content areas	▪ Specific and some technical content-area language ▪ Words or expressions with multiple meanings across content areas

- By this definition, a student at PL 5.0 should be able to **process** language with all of these characteristics

Materials for Speaking

- Recorded examples of students responding to the test items ("portfolios"), presented in order from lowest to highest raw scores
- WIDA ELD Performance Definitions
- Speaking tasks
- Data recording note sheets (paper)
- Panelist Data sheet (online)

Your mission

- Help us determine at what point along the ACCESS scale the cut score between two proficiency levels should be set
  - By domain
  - By grade
- I.e., what scale score will be interpreted as a 2.0? 3.0? Etc.

Process

1. Review the performance levels definitions
2. Review the Tier A tasks online
3. Discuss at your table:
   1. For Tier A: how might a borderline PL 2 student respond to these tasks? Borderline PL 3?
4. Discuss with group
5. Review the Tier B tasks online
6. Discuss at your table:
   1. For Tier BC: how might a borderline PL 3 student respond to these tasks? Borderline PL 4? PL5? PL6?
7. Discuss with group

Process

1. Listen to the portfolios one at a time
   1. Practice portfolio
   2. 9 portfolios per cluster presented from lowest raw score to highest raw score
2. Starting with highest grade in the cluster, ask yourself…

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Process (cont'd)

- Repeat the process for the next lower grade

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Key Questions

1. Given your understanding of the **WIDA ELD Standards**, which WIDA ELD proficiency level (1–6) does the performance you hear most reflect for that grade level?
   1. (In other words, decide – At which WIDA ELD proficiency level [1–6] is the student I am listening to currently functioning in speaking?)
2. Then decide -- How confident am I that this performance fits that WIDA ELD proficiency level (100% – 50%)?
3. Then decide -- If I'm not 100% confident of my selection in #2, which adjacent WIDA ELD proficiency level (higher or lower) do I think might also be characterized by this performance?

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Recording your judgment

Now if this student were in 11th grade, this would be a more solid PL 2

Maybe 70%…

I'll put the other 30% in PL 1

Portfolio 1							
Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11	30	70					
Grade 10							
Grade 9							
Grade 8							
Round 2							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12							
Grade 11							
Grade 10							
Grade 9							
Grade 8							

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Recording your judgment

Hmm, this sounds like a PL 2 Student in grade 12…

But kind of weak, maybe just 60%…

So I'll put the other 40% in PL 1

So I think this student has crossed over to PL 2, but just barely

Portfolio 1							
Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11							
Grade 10							
Grade 9							
Grade 8							
Round 2							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12							
Grade 11							
Grade 10							
Grade 9							
Grade 8							

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Recording your judgment

Repeat for the other grade levels in the cluster

If this student is in Grade 9, I think this is a solid PL 2, so I'll put 100% in PL 2

Portfolio 1							
Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11	30	70					
Grade 10	20	80					
Grade 9		100					
Grade 8							
Round 2							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12							
Grade 11							
Grade 10							
Grade 9							
Grade 8							

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Recording your judgment

After discussion with your table, do round 2.

I'm mostly happy with my choices, but my colleagues convinced me that for a ninth grader this is more impressive.

Portfolio 1

Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11	30	70					
Grade 10	20	80					
Grade 9		100					
Grade 8							

Round 2							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11	30	70					
Grade 10	20	80					
Grade 9		90	10				
Grade 8							

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 19

Recording your judgment in the online data sheet

If the total doesn't add up to 100%, the font stays red

Portfolio 1

Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	30	60	0	0	0	0	90
Grade 11	0	0	0	0	0	0	0
Grade 10	0	0	0	0	0	0	0
Grade 9	0	0	0	0	0	0	0
Grade 8	0	0	0	0	0	0	0

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 22

Recording your judgment

When you cross into another cluster, remember to listen to the portfolio for that cluster

Since this is a different portfolio, I'm going to ignore my judgments from the other grades

Portfolio 1

Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12							
Grade 11							
Grade 10							
Grade 9							
Grade 8	10	90					

Round 2							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12							
Grade 11							
Grade 10							
Grade 9							
Grade 8							

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 20

Process

- When everyone is finished with Portfolio 1, the facilitator will reveal the results
- Discuss at your table:
  - How did you come to your decision? How does the language in the portfolio relate to the PL descriptors?
- Round 2: In light of your discussion, do you want to adjust your judgments?
  - Listen to the portfolio again if needed
  - Record your final results in Round 2 on the note sheets
  - Record your final results in the online form

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 23

Recording your judgment in the online data sheet

Portfolio 1

Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60					
Grade 11	30	70					
Grade 10	20	80					
Grade 9		100					
Grade 8		90	10				

This updates automatically

Don't delete these 0s!

Copy here

Portfolio 1

Round 1							
	Proficiency Level						
Grade	1	2	3	4	5	6	Total
Grade 12	40	60	0	0	0	0	100
Grade 11	30	70	0	0	0	0	100
Grade 10	20	80	0	0	0	0	100
Grade 9	0	100	0	0	0	0	100
Grade 8	0	90	10	0	0	0	100

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 21

Process

Remember:

- Do one portfolio at a time
- Do not move on until everyone in your group is ready
- The key question:

Given your understanding of the **WIDA ELD Standards**, which WIDA ELD proficiency level (1–6) does the performance you hear most reflect for that grade level?

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 24

Final Steps

- Facilitator will show you:
  - Cut scores resulting from your group's judgments
  - Impact data of those cuts (i.e., what percentage of students taking Operational Series 400 would have reached each proficiency level if those cuts had been in place)
- Remember: these are **your recommendations**
- Fill out evaluation form expressing your opinions about the process and your confidence in the outcome for that domain

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 25

Cheat Sheet

	Proposed PL Score	1	2	3	4	5	6
Borderline PL3/PL4 Student →	4.0	0	0	50	50	0	0
	4.1	0	0	40	60	0	0
	4.2	0	0	30	70	0	0
	4.3	0	0	20	80	0	0
	4.4	0	0	10	90	0	0
Solid PL4 Student →	4.5	0	0	0	100	0	0
	4.6	0	0	0	90	10	0
	4.7	0	0	0	80	20	0
	4.8	0	0	0	70	30	0
	4.9	0	0	0	60	40	0
Borderline PL4/PL5 Student →	5.0	0	0	0	50	50	0

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 26

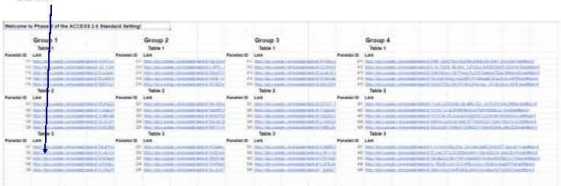
Procedures (Writing)

- Same as Speaking but
- We have created proposed cut scores, based on:
  - Previous cut scores
  - Results from Phase 1

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 26

Next Steps (cont'd)

Find your ID and click on the link



CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 29

How to use Proposed PL Score

If our proposed cuts were in effect, this column shows the PL score a student would receive for the portfolio (depending on grade level)

Use this proposal as a starting guide for your judgment, but feel free to confirm or adjust

Portfolio 4								
Round 1								
		Proficiency Level						
Grade	Proposed PL Score	1	2	3	4	5	6	Total
Grade 10	3.3	0	0	0	0	0	0	0
Grade 9	3.5	0	0	0	0	0	0	0
Grade 8	2.6	0	0	0	0	0	0	0
Grade 7	2.9	0	0	0	0	0	0	0
Grade 6	3.3	0	0	0	0	0	0	0

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 27

Next Steps (cont'd)

Check that this matches your ID

Panelist:	132				
Group:	1				
Table:	3				
ID:	2				
Grades:					
	5	4	3	2	1

Check the grade levels

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 30

Next Steps (cont'd)

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 91

Our roles and responsibilities

- Keep process moving
- Facilitate discussion
- Clarify procedures
- Answer questions

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 92

Groups

Group	Grade Levels	Facilitator	Helper
1	1-5	Melissa	
2	4-8	Tanya	
3	8-10	Cary	
4	8-12	Dorry	

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 93

Domain Screen

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

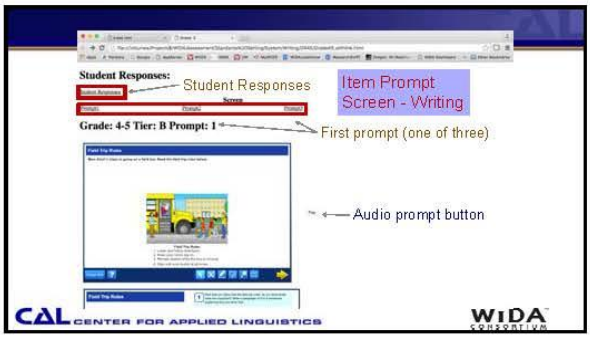
Your Roles and Responsibilities

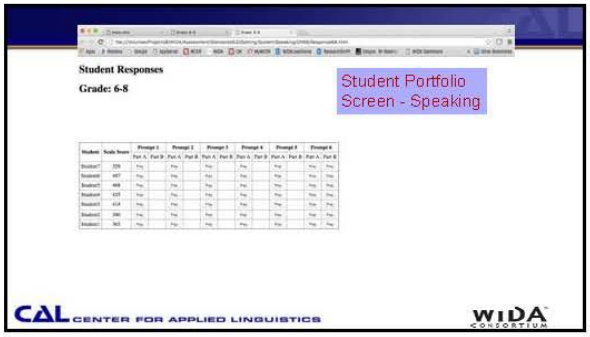
- **Table leader:**
  - Keep table on task
  - Ensure that all voices are heard
  - Report out from table discussions to whole group
  - Participate as panelist
- **Table note taker (rotating):**
  - Take clear notes of table discussions
  - Participate as panelist
- **Panelist:**
  - Understand the process
  - Ask clarification questions as needed
  - Participate in discussions
  - Make your decisions independently, based on your expertise (especially with the WIDA ELD Standards)

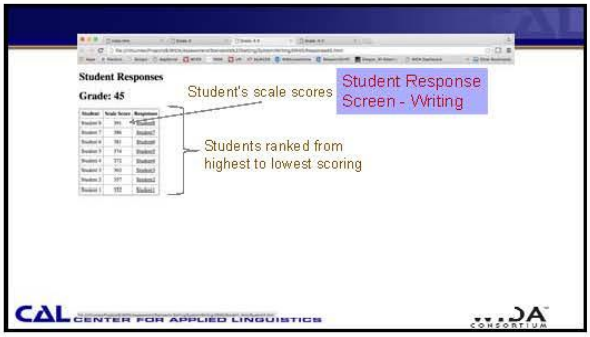
CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM 95

Cluster Screen

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM







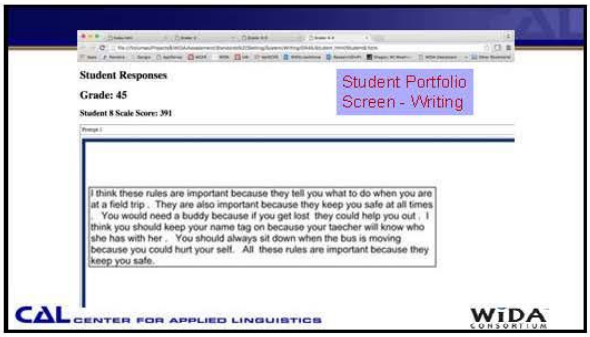
Process

Remember:

- Do one portfolio at a time
- Do not move on until everyone in your group is ready
- The key question:

Given your understanding of the **WIDA ELD Standards**, which WIDA ELD proficiency level (1–6) does the performance you hear most reflect for that grade level?

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM »



Process

Remember:

- Do one portfolio at a time
- Do not move on until everyone in your group is ready
- The key question:

Given your understanding of the **WIDA ELD Standards**, which WIDA ELD proficiency level (1–6) does the performance you read most reflect for that grade level?

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

The ACCESS Scale

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Procedures for reading and listening

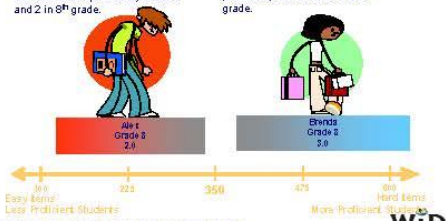
- Consider the ***borderline student*** at the ***highest grade level*** in the grade level cluster under consideration

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Borderline students at the highest grade level

For example, Alex is currently functioning as a borderline student between WIDA proficiency levels 1 and 2 in 8th grade.

Brenda is currently functioning as a borderline student between WIDA proficiency levels 2 and 3 in 8th grade.



CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

Focus on borderline student: R and L

At each grade, toward the end of a given level of English language proficiency, and with instructional support, English learners will process...

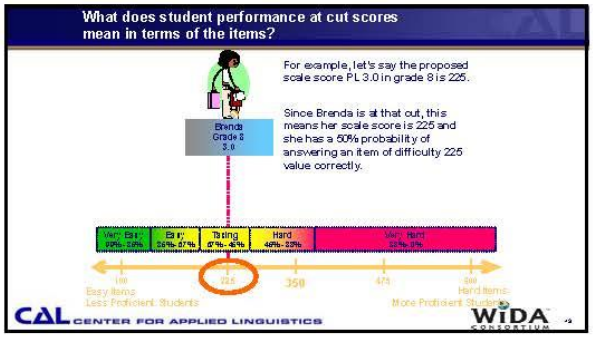
Level 4 Expanding	▪ Connected discourse with a variety of sentences ▪ Expanded related ideas	▪ A variety of complex grammatical constructions ▪ Sentence patterns characteristic of particular content areas	▪ Specific and some technical content-area language ▪ Words or expressions with multiple meanings across content areas

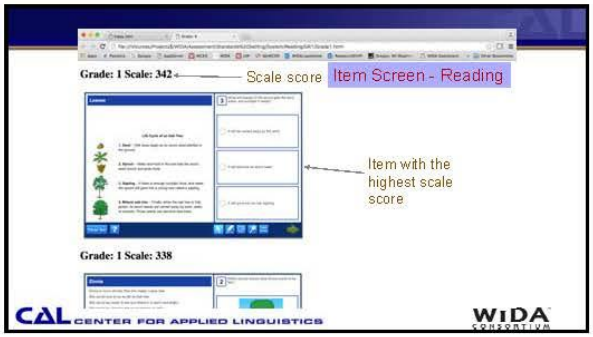
- By this definition, a student at PL 5.0 should be able to **process** language with all of these characteristics

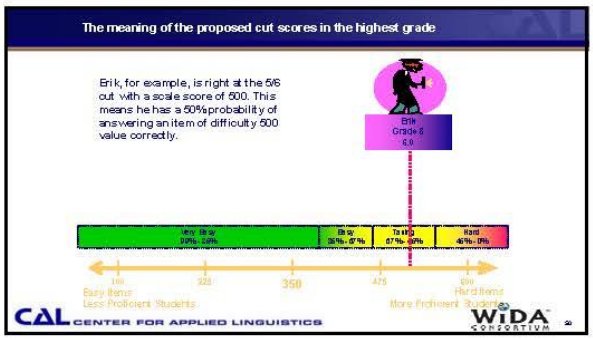
CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM

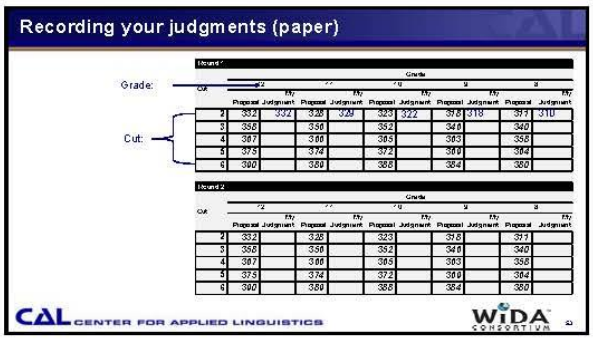
Think of borderline students at each cut point at the highest grade of the cluster

CAL CENTER FOR APPLIED LINGUISTICS — WIDA CONSORTIUM









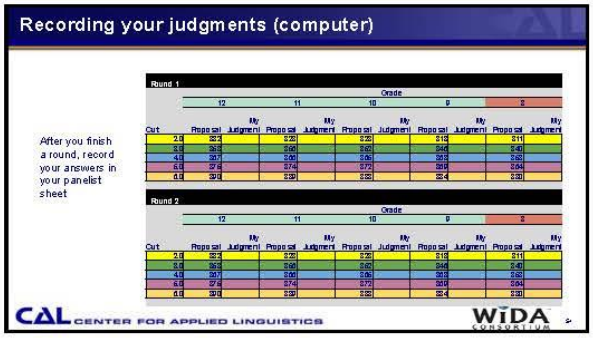
Procedures for reading and listening (cont'd)

- Thinking of students at the highest grade level at each of the cut points, review the proposed 1/2 cut point at the highest grade level (i.e., PL Score 2.0)
- Look at items at or near that cut score
- Ask yourself:

At what point would the borderline student no longer have a 50% or higher chance of getting an item correct?

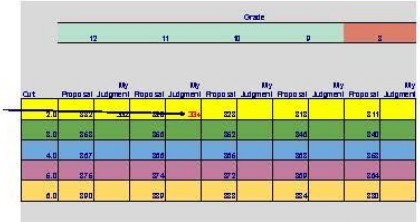
- **Confirm** our proposal, or
- **Adjust** by proposing a new cut score

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM



Recording your judgments (computer)

Note that if you set a higher cut for a lower grade, that cut will be in red font

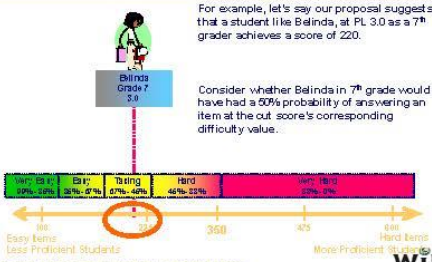


CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Example: Belinda as a 7th grader

For example, let's say our proposal suggests that a student like Belinda, at PL 3.0 as a 7th grader achieves a score of 220.

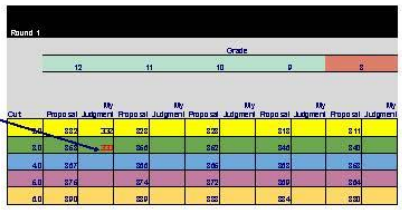
Consider whether Belinda in 7th grade would have had a 50% probability of answering an item at the cut score's corresponding difficulty value.



CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Recording your judgments (computer)

Also, if you set a lower cut for a higher Proficiency Level, that cut will be in red font



CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Procedures for Reading and Listening (cont'd)

- After everyone in the group has finished Round 1 for the 1/2 cuts, average group cut score will be reported
- If necessary, discuss
- Conduct round 2
- Then continue with all of the remaining cut scores for all of the grade levels
- After everyone in the group has finished Round 1 for the other cuts, average group cut scores will be reported
- If necessary, discuss
- Conduct round 2

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Procedures for reading and listening (cont'd)

- Repeat the process for the next lowest grade level in the cluster
- Keep going until you've finished all the grades in your group

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

Final Steps

- Facilitator will show you:
  - Cut scores resulting from your group's judgments (with a mathematical adjustment so that the cuts will be at the 67% probability level)
  - Impact data of those cuts (i.e., what percentage of students taking Operational Series 400 would have reached each proficiency level if those cuts had been in place)
- Remember: these are **your recommendations**
- Fill out evaluation form expressing your opinions about the process and your confidence in the outcome for that domain

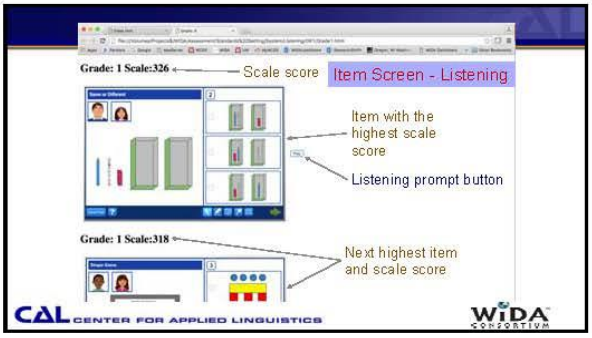
CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

The Key Question

- Start with two or three items easier than our proposed cut and move forward
- Review each item and ask yourself:

  At what point would the borderline student no longer have a 50% or higher chance of getting an item correct?
- When you reach that point, keep going for a few items to confirm your judgments

CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM



The Key Question

- Start with two or three items easier than our proposed cut and move forward
- Review each item and ask yourself:

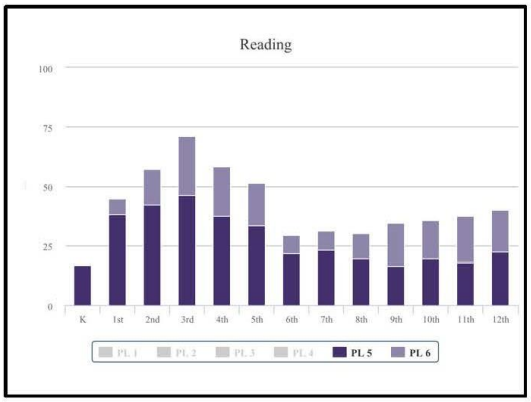
  At what point would the borderline student no longer have a 50% or higher chance of getting an item correct?
- When you reach that point, keep going for a few items to confirm your judgments

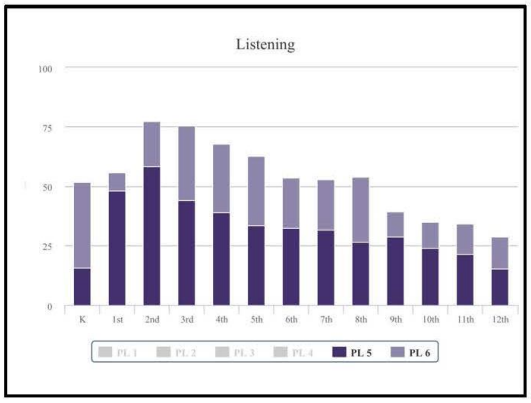
CAL CENTER FOR APPLIED LINGUISTICS WIDA CONSORTIUM

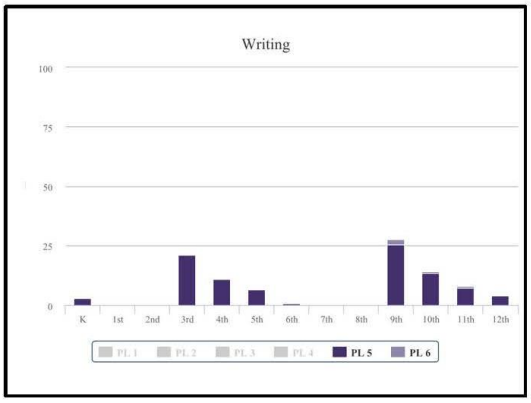
Appendix N

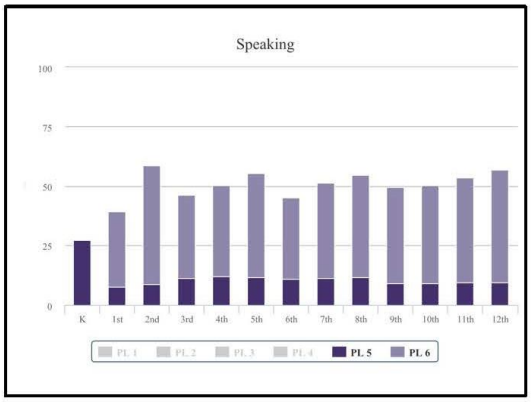
Phase 1 Impact Information

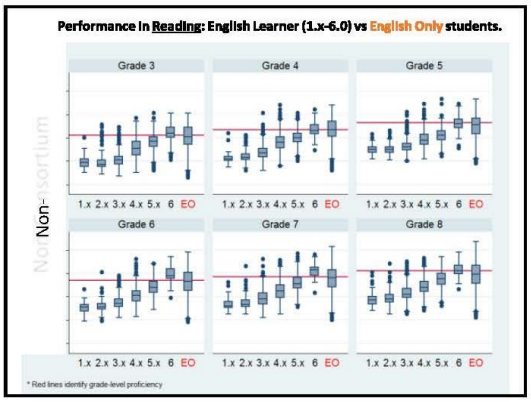
Stacked Bar Charts of 2015 Domains

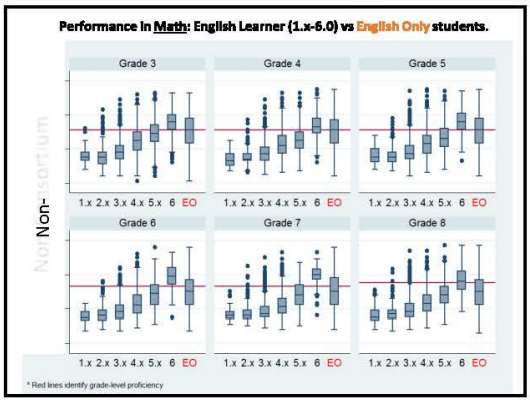


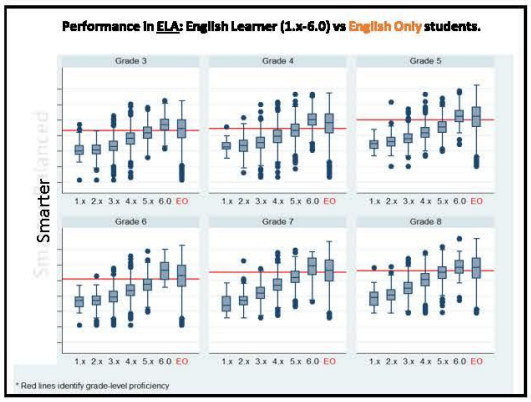


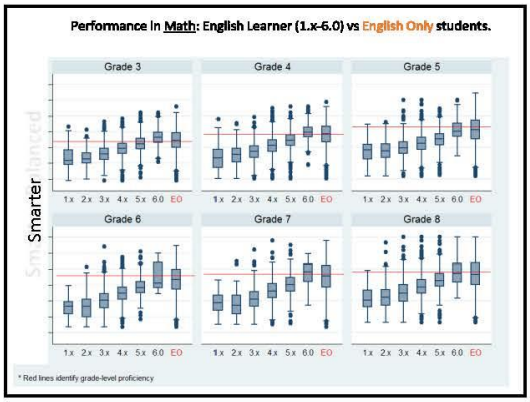


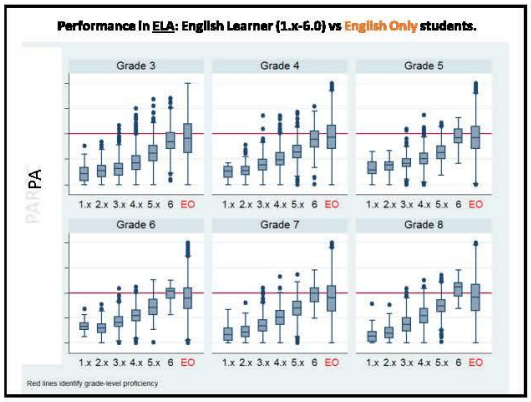


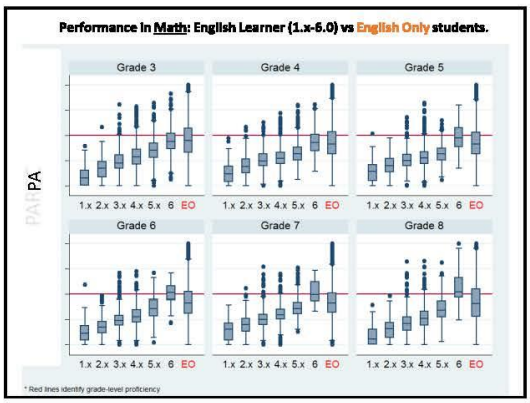



Performance in Math: English Learner (1.x-6.0) vs English Only students.
Non-
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
1.x 2.x 3.x 4.x 5.x 6 EO
* Red lines identify grade-level proficiency


Performance in ELA: English Learner (1.x-6.0) vs English Only students.
Smarter
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
1.x 2.x 3.x 4.x 5.x 6.0 EO
* Red lines identify grade-level proficiency


Performance in Math: English Learner (1.x-6.0) vs English Only students.
Smarter
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
1.x 2.x 3.x 4.x 5.x 6.0 EO
* Red lines identify grade-level proficiency


Performance in ELA: English Learner (1.x-6.0) vs English Only students.
PA
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
1.x 2.x 3.x 4.x 5.x 6 EO
Red lines identify grade-level proficiency


Performance in Math: English Learner (1.x-6.0) vs English Only students.
PA
Grade 3
Grade 4
Grade 5
Grade 6
Grade 7
Grade 8
1.x 2.x 3.x 4.x 5.x 6 EO
* Red lines identify grade-level proficiency

Appendix O
Phase 2 Data Collection Sheets

Speaking Panelist Number:_____

Speaking

Practice							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Portfolio 1							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Speaking Panelist Number:_____

Portfolio 2							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Portfolio 3							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Speaking Panelist Number:_____

Portfolio 4							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Portfolio 5							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Speaking Panelist Number:_____

Portfolio 6							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Portfolio 7							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Speaking Panelist Number:_____

Portfolio 8							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Portfolio 9							
Round 1							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							
Round 2							
	Proficiency Level						
Grade	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 5							
Grade 4							
Grade 3							
Grade 2							
Grade 1							

Panelist Number:_____

Writing

Portfolio 1								
Round 1								
		Proficiency Level						
Grade	Proposed PL	1	2	3	4	5	6	Total
Grade 12	1.5							
Grade 11	1.6							
Grade 10	1.7							
Grade 9	1.8							
Grade 8	1.7							
Round 2								
		Proficiency Level						
Grade	Proposed PL	1	2	3	4	5	6	Total
Grade 12	1.5							
Grade 11	1.6							
Grade 10	1.7							
Grade 9	1.8							
Grade 8	1.7							

Portfolio 2								
Round 1								
		Proficiency Level						
Grade	Proposed PL	1	2	3	4	5	6	Total
Grade 12	1.8							
Grade 11	1.9							
Grade 10	2.2							
Grade 9	2.4							
Grade 8	1.9							
Round 2								
		Proficiency Level						
Grade	Proposed PL	1	2	3	4	5	6	Total
Grade 12	1.8							
Grade 11	1.9							
Grade 10	2.2							
Grade 9	2.4							
Grade 8	1.9							

Writing Panelist Number:_____

Portfolio 3								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	2.6							
Grade 11	2.9							
Grade 10	3.1							
Grade 9	3.3							
Grade 8	2.2							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	2.6							
Grade 11	2.9							
Grade 10	3.1							
Grade 9	3.3							
Grade 8	2.2							

Portfolio 4								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	2.9							
Grade 11	3.1							
Grade 10	3.3							
Grade 9	3.5							
Grade 8	2.6							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	2.9							
Grade 11	3.1							
Grade 10	3.3							
Grade 9	3.5							
Grade 8	2.6							

Writing Panelist Number:_____

Portfolio 5								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.0							
Grade 11	3.2							
Grade 10	3.5							
Grade 9	3.7							
Grade 8	3.0							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.0							
Grade 11	3.2							
Grade 10	3.5							
Grade 9	3.7							
Grade 8	3.0							

Portfolio 6								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.5							
Grade 11	3.7							
Grade 10	3.9							
Grade 9	4.1							
Grade 8	3.3							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.5							
Grade 11	3.7							
Grade 10	3.9							
Grade 9	4.1							
Grade 8	3.3							

Writing Panelist Number:_____

Portfolio 7								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.7							
Grade 11	3.9							
Grade 10	4.2							
Grade 9	4.5							
Grade 8	3.5							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	3.7							
Grade 11	3.9							
Grade 10	4.2							
Grade 9	4.5							
Grade 8	3.5							

Portfolio 8								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	4.1							
Grade 11	4.4							
Grade 10	4.7							
Grade 9	4.9							
Grade 8	3.7							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	4.1							
Grade 11	4.4							
Grade 10	4.7							
Grade 9	4.9							
Grade 8	3.7							

Writing

Panelist Number:_____

Portfolio 9								
Round 1								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	4.6							
Grade 11	4.9							
Grade 10	5.2							
Grade 9	5.5							
Grade 8	4.0							
Round 2								
		Proficiency Level						
Grade	**Proposed PL**	**1**	**2**	**3**	**4**	**5**	**6**	**Total**
Grade 12	4.6							
Grade 11	4.9							
Grade 10	5.2							
Grade 9	5.5							
Grade 8	4.0							

Reading

Round 1

Cut	Grade 10 Proposal	Grade 10 My Judgment	Grade 9 Proposal	Grade 9 My Judgment	Grade 8 Proposal	Grade 8 My Judgment	Grade 7 Proposal	Grade 7 My Judgment	Grade 6 Proposal	Grade 6 My Judgment
2	333		328		320		313		305	
3	364		358		351		343		336	
4	378		376		371		364		357	
5	386		382		377		371		364	
6	402		398		393		387		380	

Round 2

Cut	Grade 10 Proposal	Grade 10 My Judgment	Grade 9 Proposal	Grade 9 My Judgment	Grade 8 Proposal	Grade 8 My Judgment	Grade 7 Proposal	Grade 7 My Judgment	Grade 6 Proposal	Grade 6 My Judgment
2	333		328		320		313		305	
3	364		358		351		343		336	
4	378		376		371		364		357	
5	386		382		377		371		364	
6	402		398		393		387		380	

Listening

Round 1

Cut	Grade 10 Proposal	Grade 10 My Judgment	Grade 9 Proposal	Grade 9 My Judgment	Grade 8 Proposal	Grade 8 My Judgment	Grade 7 Proposal	Grade 7 My Judgment	Grade 6 Proposal	Grade 6 My Judgment
2	296		287		278		270		261	
3	332		327		321		315		307	
4	360		356		352		347		339	
5	386		381		376		369		361	
6	408		405		402		397		389	

Round 2

Cut	Grade 10 Proposal	Grade 10 My Judgment	Grade 9 Proposal	Grade 9 My Judgment	Grade 8 Proposal	Grade 8 My Judgment	Grade 7 Proposal	Grade 7 My Judgment	Grade 6 Proposal	Grade 6 My Judgment
2	296		287		278		270		261	
3	332		327		321		315		307	
4	360		356		352		347		339	
5	386		381		376		369		361	
6	408		405		402		397		389	

Appendix P
ACCESS 2.0 Standard Setting Study Training Evaluation Form

The same Training Evaluation form was used for Phase 2, but the meeting title (highlighted section) was modified.

WIDA CONSORTIUM

ACCESS 2.0 Phase 1 Standard Setting Study
Training Evaluation Form

What is your opinion about the standard setting training?

	Excellent	Good	Fair	Poor
The **purpose** of the standard setting meeting was communicated clearly	O	O	O	O
The standard setting training was **organized** and **easy to follow**	O	O	O	O
I understand my **role and responsibilities** at this standard setting meeting	O	O	O	O
The presenter **responded effectively** to audience questions and comments	O	O	O	O
I feel **prepared** to begin the standards setting process	O	O	O	O

Any comments?

Appendix Q
ACCESS 2.0 Standard Setting Study Panelist Background Form

The same Panelist Background Form was used for Phase 2, but the meeting title (highlighted section) was modified.

Panelist Group: ☐ 1 ☐ 2 ☐ 3 ☐ 4 Panelist Name: _____

ACCESS 2.0 Phase 1 Standard Setting Study
Panelist Background Form

This form collects information on the qualifications of the panelists who serve in the WIDA ACCESS 400 Standard Setting Study. This information will be tabulated across all panelists and summarized in a technical report about the development of Grades 1-12 ACCESS 400.

1) What is your gender?

   O Male
   O Female

2) What is your race? (Please select all that apply.)

   O American Indian/Alaskan Native
   O Asian
   O Black/African American
   O Pacific Islander/Native Hawaiian
   O Hispanic or Latino
   O White

3) What is your native language?

   O English
   O Spanish
   O Other: _____

4) What is the highest level of education that you have completed?

   O Bachelor's degree
   O Some graduate study
   O Master's degree
   O Some doctoral study
   O Doctoral degree

5) What type of teaching credential do you have?

6) What types of additional certification do you have?

7) For how long have you been an educator?

O Less than 2 years
O 2-5 years
O 6-10 years
O 11-20 years
O 21 or more years

8) Where do you currently work?
School district/Employer: _____
City: _____ State: _____

9) For how long have you worked for your current school district/employer?

O Less than 2 years
O 2-5 years
O 6-10 years
O 11-20 years
O 21 or more years

10) What is your current position or job title?

11) If you teach, please indicate which grade level(s) you currently teach, and please indicate with a X for how long you have you been teaching the grade level(s). ***Please select all that apply.***

Grade	**Number of years**					**NA**
	Less than 2	2-5	6-10	11-20	21 or more	
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

12) If you teach, which language services do you provide instruction in at your school? (Please select all that apply.)

- O Content area tutoring
- O Content-based ESL
- O Developmental bilingual programs (two-way bilingual programs)
- O Heritage language
- O Pull-out ESL
- O Sheltered English instruction
- O Structured English Immersion / SDAIE
- O Transitional bilingual
- O Dual language two-way immersion
- O Other: _____
- O No services

13) How familiar are you with the WIDA ELD Standards?

- O Very familiar
- O Somewhat familiar
- O Not familiar

14) How familiar are you with the six WIDA proficiency levels?

- O Very familiar
- O Somewhat familiar
- O Not familiar

15) How familiar are you with the WIDA ACCESS for ELLs test?

- O Very familiar
- O Somewhat familiar
- O Not familiar

Appendix R
ACCESS 2.0 Standard Setting Study Evaluation Form: All Domains

The same Evaluation Form: All Domains was used for Phase 2, but the meeting title (highlighted section) was modified.

WIDA ACCESS 400 Phase 1 Standard Setting Study
Evaluation Form: All Domains

Reading

*The purpose of this form is to obtain your opinions about the ACCESS 400 Standard Setting Study that was just completed for **Reading**. Your opinions will help us to evaluate the process and will provide information that may be summarized in a technical report about the Standard Setting Study.*

1. What is your opinion about the adequacy of the following components of the **Reading** standard setting?

	Excellent	Good	Fair	Poor
Usefulness of the **Borderline student** discussion and descriptions	O	O	O	O
Clarity of the **WIDA performance definitions**	O	O	O	O
Facilitator and/or Table Leader support on establishing the cut scores	O	O	O	O
Materials used in establishing the cut scores	O	O	O	O
Process used in establishing the cut scores	O	O	O	O
Time allotted to establishing the cut scores	O	O	O	O
Usefulness of the impact information in making my decisions	O	O	O	O

Briefly, what process did you use to make your cut score determinations?

Any further comments?

2. What is your confidence in the ability of the group's recommended cut scores for **Reading** to adequately and appropriately classify students as borderline English proficient?

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Any comments?

Listening

The purpose of this form is to obtain your opinions about the ACCESS 400 Standard Setting Study that was just completed for ***Listening****. Your opinions will help us to evaluate the process and will provide information that may be summarized in a technical report about the Standard Setting Study.*

3. What is your opinion about the adequacy of the following components of the **Listening** standard setting?

	Excellent	Good	Fair	Poor
Usefulness of the **Borderline student** discussion and descriptions	O	O	O	O
Clarity of the **WIDA performance definitions**	O	O	O	O
Facilitator and/or Table Leader support on establishing the cut scores	O	O	O	O
Materials used in establishing the cut scores	O	O	O	O
Process used in establishing the cut scores	O	O	O	O
Time allotted to establishing the cut scores	O	O	O	O
Usefulness of the impact information in making my decisions	O	O	O	O

Briefly if different from the previous domain, what process did you use to make your cut score determinations?

Any comments?

4. What is your confidence in the ability of the group's recommended cut scores for **Listening** to adequately and appropriately classify students as borderline English proficient?

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Any comments?

Writing

*The purpose of this form is to obtain your opinions about the ACCESS 400 Standard Setting Study that was just completed for **Writing**. Your opinions will help us to evaluate the process and will provide information that may be summarized in a technical report about the Standard Setting Study.*

5. What is your opinion about the adequacy of the following components of the **Writing** standard setting?

	Excellent	Good	Fair	Poor
Usefulness of the **Borderline student** discussion and descriptions	O	O	O	O
Clarity of the **WIDA performance definitions**	O	O	O	O
Facilitator and/or Table Leader support on establishing the cut scores	O	O	O	O
Materials used in establishing the cut scores	O	O	O	O
Process used in establishing the cut scores	O	O	O	O
Time allotted to establishing the cut scores	O	O	O	O
Usefulness of the impact information in making my decisions	O	O	O	O

Briefly if different from the previous domain, what process did you use to make your cut score determinations?

Any comments?

6. What is your confidence in the ability of the group's recommended cut scores for **Writing** to adequately and appropriately classify students as borderline English proficient?

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Any comments?

Speaking

*The purpose of this form is to obtain your opinions about the ACCESS 400 Standard Setting Study that was just completed for **Speaking**. Your opinions will help us to evaluate the process and will provide information that may be summarized in a technical report about the Standard Setting Study.*

7. What is your opinion about the adequacy of the following components of the **Speaking** standard setting?

	Excellent	Good	Fair	Poor
Usefulness of the **Borderline student** discussion and descriptions	O	O	O	O
Clarity of the **WIDA performance definitions**	O	O	O	O
Facilitator and/or Table Leader support on establishing the cut scores	O	O	O	O
Materials used in establishing the cut scores	O	O	O	O
Process used in establishing the cut scores	O	O	O	O
Time allotted to establishing the cut scores	O	O	O	O
Usefulness of the impact information in making my decisions	O	O	O	O

Briefly if different from the previous domain, what process did you use to make your cut score determinations?

Any comments?

8. What is your confidence in the ability of the group's recommended cut scores for **Speaking** to adequately and appropriately classify students as borderline English proficient?

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Grade ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10 ☐ 11 ☐ 12					
	Very High	**High**	**Medium**	**Low**	**Very Low**
Borderline English Proficient	O	O	O	O	O

Any comments?

Appendix S
Phase 1 Panelist Rating Form

Panelist Name:

Grade:	
Reading	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Listening	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Writing	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Speaking	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Overall Composite	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Reading	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Listening	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Writing	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Speaking	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Overall Composite	
Ranking	
Round 1	
Round 2	
Round 3	

Panelist Group and Table:

Grade:	
Reading	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Listening	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Writing	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Speaking	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Overall Composite	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Reading	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Listening	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Writing	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Speaking	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Overall Composite	
Ranking	
Round 1	
Round 2	
Round 3	

Panelist Number:

Grade:	
Reading	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Listening	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Writing	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Speaking	
Ranking	
Round 1	
Round 2	
Round 3	

Grade:	
Overall Composite	
Ranking	
Round 1	
Round 2	
Round 3	

Appendix T
WIDA Performance Level Definitions for Levels 4 Through 6

WIDA Performance Definitions – Listening and Reading Grades K-12			
	Discourse Dimension	Sentence Dimension	Word/Phrase Dimension
	Linguistic Complexity	Language Forms and Conventions	Vocabulary Usage
At each grade, toward the end of a given level of English language proficiency, and with instructional support English learners will process…			
Level 6 Reaching	• Text and speech from a wide range of perspectives and registers (e.g., sarcasm, humor) • Abstract and often nuanced language of each content area	• Dense complex sentences indicative of particular text types and genres • A full range of grade-level sentence patterns associated with each content area	• Extensive technical, specific and general language of content area topics and themes • Subtle content area words, terms, and expressions
Level 5 Bridging	• Rich descriptive discourse with complex sentences Cohesive and organized related ideas	• Compound, complex grammatical constructions (e.g., multiple phrases and clauses) A broad range of sentence patterns characteristic of particular content areas	• Technical and abstract content-area language Words and expressions with shades of meaning across content areas
Level 4 Expanding	• Connected discourse with a variety of sentences • Expanded related ideas	• A variety of complex grammatical constructions • Sentence patterns characteristic of particular content areas	• Specific and some technical content-area language • Words or expressions with multiple meanings across content areas

WIDA Performance Definitions – Speaking and Writing Grades K-12			
	Discourse Dimension	**Sentence Dimension**	**Word/Phrase Dimension**
	Linguistic Complexity	**Language Forms and Conventions**	**Vocabulary Usage**
At each grade, toward the end of a given level of English language proficiency, and with instructional support English learners will produce…			
Level 6 Reaching	• Text and speech with a tightly woven flow of sentences matched to purpose, situation, and audience • Linguistically dense text and speech to communicate complex ideas associated with each content area	• A variety of grammatical structures indicative of particular genres • A full range of sentence patterns (e.g., embedded clauses and complex noun phrases) and conventions associated with each content area	• Technical, specific, and general language of content area topics and themes • Words, terms, and expressions with precise content area meaning
Level 5 Bridging	• Multiple, complex sentences • Organized, cohesive, and coherent expression of ideas	• A variety of grammatical structures matched to purpose • A broad range of sentence patterns characteristic of particular content areas	• Technical and abstract content-area language, including content-specific collocations • Words and expressions with precise meaning across content areas
Level 4 Expanding	• Short, expanded, and some complex sentences • Organized expression of ideas with emerging cohesion	• A variety of grammatical structures • Sentence patterns characteristic of particular content areas	• Specific and some technical content-area language • Words and expressions with expressive meaning through use of collocations and idioms across content areas

Appendix U
Study Evaluation Summary

Panelists were asked to complete an anonymous *Standard Setting Study Evaluation Form: All Domains* for Phase 1 and 2 (see Appendix R). The evaluation form was organized into domains (Reading, Listening, Writing, and Speaking). In each domain, panelists were asked to respond to the following two questions:

1. **What is your opinion about the adequacy of the following components of the [Domain] standard setting?** (Likert scale). Followed by a list of seven components.
   a. Describe briefly the process you used to make your cut score determinations. (For subsequent domains this was asked "if different from the previous domain.")(Open ended).
   b. Any comments? (Open ended.)

2. **What is your confidence in the ability of the group's recommended cut scores for [Domain] to adequately and appropriately classify students as borderline English proficient?** (Likert scale). Followed by list of grades.
   a. Any comments? (Open ended.)

In their comments panelists tended to treat Questions 1a and 1b as the same question, so we have combined the discussion of those subsections in this report. In Phase 2, Question 2, some respondents modified the evaluation form, and then answered, and commented on, their modified questions. Those responses were not included in this description of the feedback. Comments quoted in this summary are representative of common themes in the feedback. A complete list of panelist comments is available upon request.

In both Phase 1 and Phase 2 panelists, on average, rated the adequacy of the study components as a little better than *good* in all four domains. This rating average increased slightly in Phase 2. Table U1 shows the average rating panelists gave on the adequacy of the seven components of the standard setting study in all domains and in both Phase 1 and 2.

Table U1. *Phase 1 and Phase 2 Panelist Average Opinion Rating on the Adequacy of the Seven Components of the Standard Setting Study for Reading, Listening, Writing, and Speaking (4 = Excellent; 3 = Good, 2 = Fair, 1 = Poor)*

Components	Reading		Listening		Writing		Speaking	
	n	**Ave.**	***n***	**Ave.**	***n***	**Ave.**	***n***	**Ave.**
Phase 1								
Usefulness of the **Borderline student** discussion and descriptions	59	3.1	59	3.0	59	2.9	54	3.0
Clarity of the **WIDA Performance Definitions**	58	3.4	59	3.4	59	3.3	54	3.4
Facilitator and/or Table Leader support on establishing the cut scores	59	3.4	59	3.4	58	3.4	54	3.3
Materials used in establishing the cut scores	58	3.2	59	3.3	59	3.1	54	3.1
Process used in establishing the cut scores	58	2.9	58	3.2	59	3.1	53	3.2
Time allotted to establishing the cut scores	55	3.0	59	3.1	59	2.9	53	2.8
Usefulness of the impact information in making my decisions	58	3.2	59	3.4	59	3.3	45	3.3
Phase 2								
Usefulness of the **Borderline student** discussion and descriptions	53	3.4	51	3.2	52	3.1	53	3.2
Clarity of the **WIDA Performance Definitions**	54	3.4	54	3.3	55	3.3	55	3.2
Facilitator and/or Table Leader support on establishing the cut scores	54	3.5	55	3.5	55	3.4	55	3.3
Materials used in establishing the cut scores	55	3.4	55	3.3	55	3.1	56	3.5
Process used in establishing the cut scores	54	3.4	55	3.3	54	3.2	55	3.4
Time allotted to establishing the cut scores	55	3.4	55	3.2	55	2.8	56	3.0
Usefulness of the impact information in making my decisions	50	3.4	50	3.3	47	3.2	54	3.3

When asked to briefly describe the process they used to make their cut score determinations, panelists described using a variety of tools; including bookmarking, the 67% rule, the borderline student definition, the Performance Definitions, group discussions, impact data, and personal experience.

> *Read items and decide individually. Review min/max/median and discuss as a group, revise scores as needed. Review impact data and revise scores. Thought specifically about students I know, ELD and Child Development to inform decisions* [Phase 1, Reading]
>
> *Reference to performance definitions and focusing on the abilities of a "borderline" student that I've had. Personal experience with "borderline" abilities helps a lot.* [Phase 1, Reading]
>
> *I chose all cut scores based on a couple of factors. 1) description written by/with groups 2) If I thought the student would have the ability to succeed in the main stream classroom with the skills shown 3) the ability of the student to*

succeed at their grade level 67% of the time with the skills exhibited. [Phase 1, Reading]

Bookmarking- read and re-read different passages and analyzed through a discussion with table our justifications. [Phase 1, Reading]
I listened to the samples and compared them to the performance definition expectations. I also used my experience with students at the various proficiency levels for comparison. [Phase 2, Listening]
The bookmarking process was a great way to assess cut scores for the listening. [Phase 2, Listening]

Used our developed definitions and conversations. [Phase 1, Writing]
We discussed as a group after we had made individual decisions. We helped each other with breaking apart the splits: 20-80 or 30-70. I think usefulness of the impact information came after decision making, not before. [Phase 2, Writing]

Comparing to the borderline student profile and highest scale score example. [Phase 1, Speaking]

By far the most commented on component was time allotted to establishing the cut scores. In both Phase 1 and Phase 2 panelists expressed frustration that their groups worked at the pace of their slowest member, resulting in the majority of panelists having to wait between rounds or domains. This issue was addressed to a degree during the actual meetings. During the first day of both meetings facilitators started setting time limits for completing tasks to reduce the amount of downtime for panelists. WIDA and CAL staff also troubleshot technical issues that slowed entry of scores.

In Phase 1 a third of the comments about timing were specifically about time allotted to the discussion and description of a borderline student. These comments were very even, in that for every panelist that remarked that too much time was spent on the borderline student definition, another panelist felt too little time was spent on this component. According to their comments and the average rating (see Table U1) Phase 1 panelists felt the borderline student discussion was *good*, but realized that they ran out of time or had to rush to complete all the domains.

More moderation and time-keeping needed for the borderline student discussion. This should not have taken longer than an hour or so and people were focused on word-smithing problems with tallying scores slowed process down! [Reading]

The discussion to define the "borderline" student took much longer than anticipated. It is a valuable discussion and should not be short changed;

however, an hour does not provide adequate time. …Reading scores felt "rushed" because we were behind schedule due to the time needed to define "borderline". [Reading]

Could have used more discussion on borderline students here. We spent a lot of time on creating a definition but we sacrificed as much of time and saw the student portfolios. [Writing]

Because so much time was wasted on the borderline student discussion, our time was impacted for determining cut scores. [Listening]

In Phase 1 and Phase 2 panelists, on average, rated their confidence in the ability of the group's recommended cut scores to adequately and appropriately classify students as borderline English proficient as 3.8, where (3 = Medium, 4 = High, and 5= Very High). In Phase 2 that average rating was 4.2. Table U2 shows the average rating panelists gave of their confidence in the ability of the group's recommended cut scores to adequately and appropriately classify students as borderline English proficient in all domains and in both Phase 1 and 2.

Table U2. *Panelist Average Rating of Their Confidence in the Ability of the Group's Recommended Cut Scores to Adequately and Appropriately Classify Students as Borderline English Proficient in Reading, Listening, Writing, and Speaking (5 = Very High , 4 = High; 3 = Medium, 2 = Low, 1 = Very Low)*

Grade	Reading		Listening		Writing		Speaking	
	n	Average	*n*	Average	*n*	Average	*n*	Average
Phase 1								
1	14	3.9	15	4.3	12	3.5	13	3.3
2	14	3.9	14	4.1	12	3.6	13	3.3
3	14	4.1	14	4.2	12	3.6	13	3.5
4	28	4.1	28	4.0	24	3.7	24	3.2
5	28	4.0	28	4.1	24	3.7	21	3.4
6	29	3.5	28	4.0	26	3.6	17	3.8
7	29	3.6	28	3.9	26	3.6	17	3.8
8	44	3.7	42	3.8	40	3.5	32	3.7
9	30	3.8	28	4.0	28	3.4	24	3.6
10	30	3.9	28	4.0	28	3.5	24	3.8
11	15	3.7	14	4.0	14	3.6	15	3.7
12	15	3.8	14	4.1	14	3.6	15	3.6
Phase 2								
1	15	4.3	14	4.3	14	3.9	15	4.3
2	15	4.3	14	4.4	14	4.0	14	4.1
3	15	4.3	14	4.4	14	4.1	14	4.2
4	18	4.1	17	4.0	29	4.1	29	4.1
5	19	4.2	17	4.0	29	4.1	29	4.1
6	27	4.0	27	4.0	28	4.0	28	4.1
7	27	4.0	27	4.0	28	4.0	28	4.1
8	39	4.3	40	4.1	34	3.9	40	4.2
9	24	4.3	25	4.2	26	4.2	26	4.4
10	24	4.3	25	4.2	26	4.1	26	4.3
11	12	4.3	13	4.2	13	4.0	13	4.4
12	12	4.2	13	4.2	13	4.0	13	4.5

One area where some panelists expressed concerns in the comments was regarding middle school cut scores.

> *While I recognize that MS exit rates are generally lower, the impact numbers presented (5%, 7%, 8% if I recall correctly) seemed way too low. I maintain that based on the skill descriptors we wrote that the minimum values recorded would be sufficient, or at least much closer to accurately indicating the "borderline" student.* [Phase 1, Reading]

> *Many were concerned with the low percent of students meeting the median cut score at grades 6-8. It would have been helpful to have more info from WIDA about middle school and high school scoring trends.* [Phase 1, Reading]

9-12 cluster test and 6-8 cluster test are two different animals- therefore, couldn't we expect the scale scores to differ? [Phase 1, Listening]

I feel a little bothered that I heard some people rated scores for 8th grade lower than 9th grade as they believe 8th must be logically lower. Whereas other conversations I heard they say they would prefer to not have scored in that manner for their gut said the test questions were easier for 8th. [Phase 1, Listening]

I don't think I understand the 8th grade brain because either that's the problem and 8th graders are cognitively much lower than I think or the 8th grade item was much easier. [Phase 1, Listening]

I strongly feel the median scores for 6, 7, 8 were too high. [Phase 1, Writing]

Too big gap ms gr 6/7. Think it is a bit too high at median [Phase 1, Speaking]

I am still grappling with the concept of different cut scores for 6-8 and 9-10. If the standard is the same and test is the same, why would scores be different? This may be something I need to just accept. [Phase 2, Reading]

Not entirely confident given we were so rushed w/ little time for discussion. People have high school expectations for middle school children. More focus on the middle school age group for the test. [Phase 2, Reading]

It should be noted that there were less than half the amount of comments on panelist confidence in the cut scores in Phase 2 compared to Phase 1. This is probably because there is evidence that panelists became more confident with the process as the work progressed. In addition, 19 panelists participated in both phases of the study and had experience with the process.

We were all very consistent in our scoring which shows that the group is confident in its recommendations. [Phase 1, Reading]

Confident that we are meeting EL's needs. [Phase 1, Listening]

I think our group is representative of educators who understand student needs and the CCSS in depth, so they know what proficiency students need to be successfully exited. [Phase 1, Reading]

Since this was the second domain, it made it easier. [Phase 1, Listening]

At first, I couldn't figure out what I needed to do but I finally figured it out. It's a great way to determine cut scores thinking about students in my school district. [Phase 2, Reading]

It took a little while to get the hang of this process. Bookmarking is new to me. Because of the limited number of items, some of the cut scores felt subjective.

However, I do feel that we established some accurate cut scores today. [Phase 2, Listening]

I am glad we got a chance to "reboot" this on Friday morning - much more confident now. [Phase 2, Listening]

This process was extremely eye opening. After seeing this process, I am very confident about the validity of the test and my students' scores. [Phase 2, Listening]

Once we did listening, reading was easier. [Phase 2, Reading]

I feel pretty confident about the cut scores recommended by my group. [Phase 2, Speaking]