



# Washington OSPI WIDA ACCESS Exit Criteria Standard Setting Report

May 2022

Susan Davis-Becker, Ph.D.  
Chad Buckendahl, Ph.D.  
Kelley Wheeler, M.Ed.

## Overview

The Washington Office of Superintendent of Public Instruction (OSPI) contracted with ACS Ventures, LLC (ACS) to establish exit criteria for the WIDA ACCESS for kindergarten and the WIDA ACCESS. ACS worked with OSPI leadership to design and facilitate a study by which Washington subject matter experts could make judgments as to what level(s) of performance would indicate that a student had met English language (EL) proficiency and is ready to Exit EL services. This report documents the process and results of these efforts.

Scores from the WIDA ACCESS reflect measures of four domains: Reading, Listening, Writing, and Speaking. The Reading and Listening domains include multiple-choice items (scored by the computer) whereas the Writing and Speaking domains include performance tasks (scored against a scoring rubric). Tasks and items are selected for a student based on their prior performance (i.e., students who do well receive more challenging tasks).

Students receive a score report from the WIDA ACCESS end of year assessment with a total of seven scores (three composite, four domain):

- Composite Scores
  - Overall: Reading (35%) + Writing (35%) + Speaking (15%) + Listening (15%)
  - Oral Language: Speaking (50%) + Listening (50%)
  - Literacy: Reading (50%) + Writing (50%)
  - Comprehension: Reading (70%) + Listening (30%)
- Domain Scores
  - Reading
  - Writing
  - Speaking
  - Listening

Each score is reported on the WIDA performance scale (1-6<sup>1</sup>, whole integers for domain scores, half points possible for composite scores):

- 1) Entering
- 2) Emerging
- 3) Developing
- 4) Expanding
- 5) Bridging
- 6) Reaching

The expectations for each level are described generally in the *Performance Level Definitions* (PLDs) and more specifically for each grade level in the *Can Do Descriptors* provided by WIDA.

Each state that uses the WIDA ACCESS assessments is responsible for establishing Exit criteria that indicate when students are ready to stop receiving EL services. To complete this activity, ACS designed and facilitated a judgmental process by which Washington subject matter experts could recommend what levels of performance on the WIDA ACCESS indicate that a student is ready to exit EL services.

---

<sup>1</sup> Because the WIDA ACCESS for kindergarten does not assess advanced reading and writing skills, kindergartners cannot earn a Reading proficiency level above 5.0, or a Writing proficiency level above 4.5



## Meeting Process

OSPI selected thirty subject matter experts to serve as panelists in this study who all worked with English language learners in Washington. Based on their experience, the panelists were organized into six panels to reflect the different grade levels of the WIDA ACCESS assessment: K, 1, 2-3, 4-5, 6-8, 9-12. Table 1 below shows the experience of each panel.

**Table 1. Subject Matter Expert Background**

Panel	Count	Experience Teaching English Development			
		% < 3 years	% 3-7 years	% 7-10 years	% > 10 years
K	5	20%	40%	40%	--
1	5	20%	20%	60%	--
2-3	5	--	--	60%	40%
4-5	5	--	20%	80%	--
6-8	5	--	40%	60%	--
9-12	5	--	20%	60%	20%

To complete this task, ACS designed and facilitated a series of meetings April 12-14, 2022, where the panelists had the opportunity to discuss general expectations as well as more specific grade level criteria while iterating on their judgments. This iterative judgmental process focused on the opinion of the expert panelists but included review and consideration of:

- **Expectations by Performance Level** – panelists were provided copies of the general *WIDA PLDs* (expectations for K or grades 1-12 by domain) and the more specific *Can Do Descriptors* which are organized by grade band (K, 1, 2-3, 4-5, 6-8, 9-12), domain, and purpose of communication.
- **Student Performance** – Washington student data was not available at the time of the study. However, WIDA provided data that ACS used to create performance profiles. These profiles showed the average domain scores for students who earned a given overall score (4.0, 4.5, 4.8, 5.0). This information could be used in two ways. First, the panelists could determine, for students who achieved an overall domain score of 4.5 (for example), which of their domain scores were above this score, at this score, or below this score. Second, the panelists could compare domain scores for multiple overall scores (e.g., how different is the average Reading score for students who earned an overall 5.0 compared to an overall 4.5?).
- **Exit criteria set by other states** – the panelists reviewed a summary of these expectations including the type of scores used in the Exit criteria (overall, composite, domain) as well as the minimum scores expected for each.

The process followed is outlined below by meeting day.

### 1. Meeting Day 1: Process Training and General Discussion

The first meeting was conducted as a general session with all panelists. ACS conducted training for the entire process that would be followed (goals and activities for each meeting day), the WIDA ACCESS (tasks, scoring, reporting, PLDs), and how they would make their judgments that would eventually lead to the panel's final Exit criteria recommendations. ACS facilitated three brainstorming activities. Specifically, the panel discussed:

- What knowledge and skills signals that a student is ready to exit EL services?
- What evidence within the WIDA PLDs represent the signals discussed in the first activity?



- Which score(s) from the WIDA ACCESS should be used in setting the Exit Criteria? [Overall score, Composite scores, Domain scores].

## 2. Meeting Day 2: Panel-Level Group Discussions

ACS began the second meeting with a general session that introduced the task of panelists making judgments as to the Exit criteria for their grade level based on the PLDs and Can Do statements. Panelists then moved into their grade level panels (K, 1, 2-3, 4-5, 6-8, 9-12). Within each panel, facilitators guided panelists to work through the following steps:

- Review the grade level Can Do statements and discuss included expectations that align with the general signals discussed during the previous meeting day.
- Provide initial judgments as to the Exit criteria (overall expectation, any domain/composite score expectations in relation to the judgments made the day before)
- Discuss the initial judgments from all grade level groups to determine where the expectations align and where they differ. Determine if, and how, their judgments should be adjusted based on those of the other grade level groups.
- Provide second round of judgments as to the Exit criteria.

## 3. Meeting Day 3: Final Articulation

For the final meeting, one panelist from each grade level was invited to a general session and asked to share the perspectives of their group. The purpose of this meeting was to offer the groups the opportunity to come to consensus on their recommendations. The group was presented with the results from the second round of ratings and each representative shared their group’s perspectives while ACS facilitated a discussion by highlighting the common and differing perspectives.

After the final meeting, the panelists complete an evaluation of the process.

## Meeting Results

The panel was tasked with making two types of recommendations as a part of this project:

1. Which score(s) from the WIDA ACCESS should be used in setting the Exit criteria?
2. What level of performance, on each of the relevant scores, should indicate a student is ready to exit EL services?

The results of each of these recommendations are described below.

Tables 2 and 3 display the results of the two rounds of materials review and discussion within each grade level panel, specifically, the recommended provided by the panelists via online forms. For each WIDA ACCESS score, the results in Tables 2 and 3 indicate the number of panelists who recommended the use of the score as an Exit criterion (N) as well as the average score that would indicate student is ready to Exit (Avg).

**Table 2. Round 1 Exit Expectations**

Panel	Composite Score								Domain Score							
	Overall Score		Oral Proficiency		Literacy		Comprehension		Reading		Writing		Listening		Speaking	
	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg
<b>K</b>	5	4.1	1	5.0	1	3.5	3	3.8	4	3	2	3	4	4	4	4
<b>1</b>	4	4.1	0	--	2	4.3	1	4.0	2	4	5	4	2	4	2	4



<b>2-3</b>	1	4.8	2	4.5	2	4.4	1	4.8	5	4	5	4	5	5	5	4
<b>4-5</b>	3	4.5	1	4.0	1	4.0	1	4.0	2	5	1	4	2	4	1	4
<b>6-8</b>	3	4.7	0	--	0	--	0	--	2	4	2	4	2	4	2	4
<b>9-12</b>	3	4.5	1	4.5	1	4.5	3	4.5	1	5	1	4	1	5	0	--

**Table 3. Round 2 Exit Expectations**

Panel	Composite Score								Domain Score							
	Overall Score		Oral Proficiency		Literacy		Comprehension		Reading		Writing		Listening		Speaking	
	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg	N	Avg
<b>K</b>	0	--	5	5.0	2	3.5	5	3.5	1	4	0		1	4	1	4
<b>1</b>	0	--	0	--	0	--	0	--	5	4	5	4	5	4	5	4
<b>2-3</b>	3	4.8	0	--	0	--	0	--	4	4	4	4	4	4	4	4
<b>4-5</b>	2	4.8	0	--	0	--	0	--	1	4	1	5	1	5	1	4
<b>6-8</b>	4	4.7	0	--	0	--	0	--	2	4	2	4	2	4	2	4
<b>9-12</b>	4	4.4	1	4.5	2	4.5	2	4.5	2	4	2	4	2	4	1	4

**Summary of Grade Level Recommendations**

Table 4 provides a summary of the round 2 results (detailed in Table 3) by identifying which WIDA ACCESS scores were supported as appropriate Exit criteria by the majority (or all) of the panel agreed upon a recommendation (most/all), half the panel (half), or none (or a few) of the panelists (blank) along with the average recommended scores that would indicate a student is ready to Exit. As shown in these results, the kindergarten and grade 1 panels had reached consensus that exclusively using an Overall score would not be appropriate for determining the Exit criteria. This change was also reflected in the group discussions between rounds. The results for the panels representing grades 2-12 also demonstrated a consolidation of responses, however these panels reached a different consensus that the Overall score should contribute to the Exit criteria.

**Table 4. Exit Criteria Recommendations: Criteria (and Scores)**

Panels	Overall	Composite: Oral Proficiency	Composite: Literacy	Composite: Comprehension	Domain Scores: Reading, Writing, Listening, Speaking
<b>Kindergarten</b>		Most/all (5.0)		Most/all (3.5)	
<b>Grade 1</b>					Most/all (4,4,4,4)
<b>Grade 2-3</b>	Most/all (4.8)				Most/all (4,4,4,4)
<b>Grades 4-5</b>	Most/all (4.8)				Half (4,5,5,4)
<b>Grades 6-8</b>	Most/all (4.7)				Half (4,4,4,4)
<b>Grades 9-12</b>	Most/all (4.7)		Half (4.5)	Half (4.5)	Half (4,4,4,4)

## Vertical Articulation

During the vertical articulation process, the panelists were given an opportunity to explain their rationale for the differing recommendations between grades (see Appendix A for the full set of notes). The summarizes the justification provided for the difference between the final recommendations (Table 5):

- Kindergarten: This panel was originally supportive for using Overall score in the Exit criteria but disagreed with the importance placed on the different domains based on the weighting (percent represented by each domain) that contributed to the Overall score. The panel did not believe the weighting was developmentally appropriate for kindergarten.
- 1<sup>st</sup> Grade: This panel had similar concerns as the kindergarten panel due to the developmental similarities of the students. The panel believed that the individual Domain scores should be the focus of the instead of the Overall score.
- Overall: The other panels agreed that using an Overall score would be appropriate if there was a minimum score required on each Domain to exit. The panels agreed that having a minimum Domain score of 4 would ensure the domains were being appropriately emphasized if the Overall score was higher. The group recommended that the Overall score should be close to 5 to protect students from exiting the program too early and missing services they might still require to be successful in the academic content.

**Table 5. Final Exit Criteria Recommendations**

Panel	Overall Score Recommendation	Domain Score Recommendation
Kindergarten	<i>None</i>	4.0 – Reading, Writing, Speaking, and Listening
Grade 1	<i>None</i>	4.0 – Reading, Writing, Speaking, and Listening
Grades 2-3	4.7	4.0 – Reading, Writing, Speaking, and Listening
Grades 4-5	4.7	4.0 – Reading, Writing, Speaking, and Listening
Grades 6-8	4.7	4.0 – Reading, Writing, Speaking, and Listening
Grades 9-12	4.7	4.0 – Reading, Writing, Speaking, and Listening

## Evaluation

The results of the online, anonymous evaluation are shown in Table 6 below. The panelists indicated they found value in the orientation activities. For their initial judgments on day 2, the panelists indicated they used all of the available information with lower emphasis on the exit criteria from other states whereas when making their final recommendations they placed more emphasis on the group discussions and performance profiles. For the final day, the panel felt they had the opportunity to share their grade group results. With respect to timing, the panel largely felt there was sufficient time for each activity although some panelists would have liked more time for the general session and grade group session. Additional comments provided in the evaluation can be found in Appendix.



**Table 6. Study Evaluation Results**

Session & Evaluation Questions	Responses by Category			
<b>GENERAL SESSION (DAY 1)</b>				
	<b>Not at all Valuable</b>	<b>Minimally Valuable</b>	<b>Somewhat Valuable</b>	<b>Valuable</b>
<i>Value of Orientation Activities</i>				
Information about WIDA ACCESS	4%	7%	18%	71%
Discussion #1: Signals that indicate a student is ready to Exit EL services	0%	4%	22%	74%
Discussion #2: Review of general PLDs as a large group	0%	7%	22%	70%
Discussion #3: Value of different WIDA scores in determining readiness to Exit.	0%	0%	19%	81%
<b>GRADE GROUP SESSION (DAY 2)</b>				
	<b>Not at all</b>	<b>Somewhat</b>	<b>Strongly</b>	
<i>Influences on <u>initial</u> recommendations</i>				
General Discussions from Day 1	7%	32%	61%	
Exit criteria from other states	11%	57%	32%	
Grade group discussions	7%	32%	61%	
WIDA Grade level Can Do statements	7%	32%	61%	
Performance Profiles	7%	32%	61%	
<i>Influences on <u>final</u> recommendations</i>				
General Discussions from Day 1	7%	50%	43%	
Exit criteria from other states	14%	57%	29%	
Grade group discussions	0%	21%	79%	
WIDA Grade level Can Do statements	4%	39%	57%	
Performance Profiles	7%	29%	64%	
Exit criteria from other grade groups	18%	46%	36%	
<b>ARTICULATION MEETING (DAY 3)</b>				
	<b>No</b>	<b>Somewhat</b>	<b>Yes</b>	
Did you have sufficient opportunity to share group's recommendation	0%	29%	71%	
<b>TIMING OF DISCUSSIONS</b>				
	<b>Not enough time</b>	<b>The right amount of time</b>	<b>Too much time</b>	
Time dedicated to general session	14%	82%	4%	
Time dedicated to grade group discussions	36%	64%	0%	
Time dedicated to articulation discussion	0%	100%	0%	



## Summary and Next Steps

As documented in this report, Washington subject matter experts engaged in a series of discussions and information sharing to formulate recommendations as to what level of performance on the WIDA ACCESS indicates that a student is ready to exit EL services. Through iterative rounds of judgments, each grade level panel established a general understanding of what knowledge and skills students should demonstrate at their respective grade level. The cross-panel discussions yielded some consistency in recommendations (with a few slight differences). Overall, the recommended expectations were slightly different for kindergarten and 1<sup>st</sup> grade students than for the rest of the students due to developmental needs.

These results and recommendations are presented to the Washington OSPI for their consideration and decision making.





## Appendix A: Vertical Articulation

The notes below represent the final perspectives of each grade-level group as captured during the vertical articulation discussion and by the group facilitator.

- Not all panelists completed form for some groups because those panelists agreed with the responses being given by other members of the group.
  - In another group, one panelist could not complete the first survey due to storms.
  - Some groups had a wide array of opinions on what should be emphasized the most.
- Kindergarten: Originally unanimous for using overall, but eventually did not like the weighting and thought that comprehension was more in line the ultimate goal for kindergarten.
  - The Exit Criteria should be different in kindergarten than in other due to the different student experiences and contexts of the domains.
- 1<sup>st</sup> Grade: Domain scores should be the focus of the Exit Criteria
  - Did not like the weighting of the overall scores
  - Focusing on overall would not emphasize writing enough due to the weighting (%) distribution of the overall score
  - Recommendations for an Exit Criteria different for 1st grade compared to other grades
- Grades 2-3: Just using an overall score of and 4 might not be appropriate but felt there was a large jump between the expectations of the scores of 4 and 5.
- Grades 6-8: Agrees that domain and overall scores should be weighted equally.
- HS: Discussion focused on the need to have comprehension weighted heavier due to the cultural nature of the exams. Differences in the full group due to the disagreement on how much to weight comprehension compared to reading.
- Overall:
  - There should at least be a domain minimum for certain scores even if overall score is used.
  - Speaking domain recommended lower weighting due to the assessment being the most intimidating portion for students.
  - Domain scores should be used equally with the composite to determine Exit Criteria because it helps set goals for the students vs just using the composite/overall scores.
    - Some concerns that just an overall score would be hard to explain to students
  - Groups would be comfortable having a 4 as a floor score for domains if the overall score was higher.
    - Maybe approaching 5 for an overall score
  - The expectations detailed in the Can Do statements might be too high in context of learning after COVID, so the Exit Scores should take that into consideration.
  - Happy with using the overall score as long as a 4 is used as a floor score for the domains.



## Appendix B: Evaluation Comments

- I appreciated setting the groundwork on Day 1. Conversations were rich and really helped.
- This was a great process to participate in. I really appreciated how the facilitators worked really hard to rephrase what we said for understanding. It was great learning. Thank you!
- As someone who didn't participate in Day 3, I'm curious how our input was conveyed to the group, and then how it was/will be included in state decision making.
- I think this is something that we/OSPI should come back to in about a year or two when we have more data to see if we need to adjust. I think our team really struggled with finding the line between "exiting too early" and "beyond what we expect at grade-level performance." Somewhere between a four and a five, students are meeting grade-level expectations. More data might help that picture become more clear.
- Some of the participants seemed to have a poor understanding of Applied Linguistics or a deficit orientation to the work. In some cases, I felt desperate about the amount of time they derailed the discussions or tried to silence other group members. One of these persons had a poor understanding of what happens when students are exited from care too early. I wish the conversation had included the purpose of this assessment. Trying to move forward without understanding the Castaneda Standards or disparate impact was problematic since two of our group members did not seem to have an understanding of the legal purpose of this assessment.
- I felt like on 2nd day, with the small group discussion, that there was at least one person who had very strong opinions and we kind of ended up going with their thoughts/they kind of influenced the group. I felt like I didn't always have a chance to say my opinion. I did feel as if I was able to do it on the survey, but then once we discussed the survey I felt like once again the one person's opinions influenced, what the person for the 3rd day group articulation was going to state.
- I thought the whole process was very well done and that we achieved our goal. I say this while at the same time greatly disliking the WIDA test and wishing the state hadn't adopted it.
- Thank you for the opportunity to participate. The facilitator did an excellent job capturing individual ideas and presenting them in a clear and concise manner. Additionally, the transition from individual ideas to group consensus flowed smoothly at a good pace.
- The facilitator did a wonderful job.
- It was really great to participate in the fruitful discussions with the groups and the facilitators during our meetings. I really like how it was organized to meet in large and then small groups, that led to overall great discussions. I believe such decision could have an impact on many ELLs, thus I would suggest another day perhaps could help us better look more deeply on the Can Do Descriptors with each key use for each language domain. These descriptors are really important since they focus on the academic language necessary for ELLs to move from one level to another and to exit the program. Such discussions could lead us to refer back to standards, look at other factors such as testing windows (this is critical for K group), and hence our recommendation could be different. Overall, I am just curious about misplacing students into the program and this made me think of any other possible factor that could help us come to the best possible decision. I really enjoyed the discussion with all and I am confident though that we did our best to come into the final recommendation for our grade level. Thank you all for facilitating such great discussions to help our English language learners all over the state of Washington.
- I think that this process was well defined
- Thanks for doing this important work!
- This was a valuable experience. I appreciate the opportunity to be involved.
- It would have been helpful to discuss the purpose of each kind of score according to the test specifications.

