

# Recommendations for Counting Results of English Language Learners (ELLs)

April 27, 2009

Pete Bylsma, EdD, MPA

*To improve the validity of accountability results, test scores of ELLs should be excluded in the first three calendar years of their enrollment in a U.S. public school, or until they reach Level 3 (advanced English) on the WLPT, whichever comes first. Although this policy has little impact on the index results and affects relatively few students, it improves the validity and fairness of the accountability system. To improve accountability even further, OSPI should begin reporting WLPT results on its Report Card in a way that allows educators, parents, and other stakeholders to monitor the progress ELLs make in terms of learning English and meeting state content standards. OSPI should also determine the extent to which ELLs are on track to meet state standards by analyzing their combined state assessment results (WASL and WLPT).*

## **BACKGROUND**

The accountability index is based on how students perform on a set of outcomes, including four state assessments: reading, writing, mathematics, and science. Federal accountability regulations require states to include the reading and math results of English language learners (ELLs) who are in their second year of enrollment in a U.S. public school, regardless of their English ability, when determining adequate yearly progress (AYP).

To increase the validity of federal accountability results, OSPI repeatedly requested that ELL test results not be included during a student's first three years of enrollment in a U.S. public school or until the student achieves an advanced level of English proficiency (Level 3) on the Washington Language Proficiency Test (WLPT), whichever comes first.<sup>1</sup> The requests were made because (a) research shows it takes many years for ELLs to acquire "academic" proficiency in English, (b) ELLs must be able to read and write English to understand and respond to test items,<sup>2</sup> (c) analyses of ELL performance found that more than half passed the WASL by the end of their third year in the program,<sup>3</sup> and (d) testing these students in English violates widely-adopted testing standards and ethics due to threats to validity and mistreatment of human subjects. However, the U.S. Education Department has denied OSPI's requests to use this policy.

In January 2009 OSPI took a different approach and submitted a proposal to the Department as part of its Title III plans. OSPI proposed using the ELL results in the second year of U.S. enrollment when determining AYP, but the definition of proficiency would be based on a "sliding

---

<sup>1</sup> All ELLs take a *placement* test to determine *initial* eligibility in the state program. ELLs must then take an *annual* language proficiency test (WLPT-II), to determine *continued* eligibility. The composite score from the annual test, which reflects proficiency in reading, writing, speaking, and listening, would determine if the ELLs' WASL/WAAS results are included in accountability calculations that year. Per federal requirements, ELL students are not required to take the reading test in their first year of enrollment in a U.S. public school, but they are required to take the math test, even though the results are not counted in the first year.

<sup>2</sup> The WASL math and science assessments are available in Spanish and Russian for the first time in 2009. However, all students must still respond in English on open-ended items rather than in their primary language. Responses in languages other than English are given no credit.

<sup>3</sup> Research has shown it may take longer than three years to acquire proficiency in English in an *academic* setting, depending on the knowledge and skills a student has when entering the country.

scale” according to a student’s level of English proficiency as measured on the WLPT. Specifically, OSPI proposed counting students as proficient using a lower scale scores on the WASL for each grade and subject, depending on how the student performed on the WLPT. The required scale scores were established using statistical techniques that estimated the score needed for ELLs to be “on track” to meet standard by the time they transitioned out of the program.<sup>4</sup> For example, a grade 3 ELL student scoring in Level 2 on the WLPT (intermediate English) would be counted as proficient in reading with a WASL scale score of 359; a grade 3 student in Level 3 on the WLPT (advanced English) would count as proficient with a scale score of 388. The required scale scores were lower for math. No required scale scores were generated for writing and science because they are not part of the federal accountability system. Theoretically, the required scale scores would change each year because the overall test results change each year. The Department did not accept this proposal, just as it had rejected a similar concept related to students with disabilities (some students would have been considered proficient if they reached Level 2 on the WASL, based on the requirements in their Individualized Education Program).

### Stakeholder Views

Various stakeholder groups and technical advisors discussed how best to use ELL results when calculating the Accountability Index. They considered the merits and disadvantages of both approaches—the “extended exemption” and the “sliding scale”—as well as other options. Analyses of both approaches found that both had little impact on the index results.<sup>5</sup>

- The 3-year exclusion does not exclude many ELL results because (a) most ELLs have been enrolled for 3 years by the time they first take a state assessment in grade 3, and (b) most ELLs in grades 3-10 have achieved advanced level on the WLPT (see Figure 1).<sup>6</sup>
- The slide scale method increases the number of students considered proficient by a small amount.

While the effect of both approaches is small, the minor gains are nevertheless important to educators who support a new approach to increase the fairness and validity of accountability results. Stakeholders agreed that if the sliding scale option were used, it needed to be simpler than what OSPI had proposed to the Department. While there was support for both approaches, ultimately a majority of stakeholders preferred the “extended exclusion” approach because it is simpler to understand and communicate. Appendix A discusses the merits and limitation of the sliding scale option.

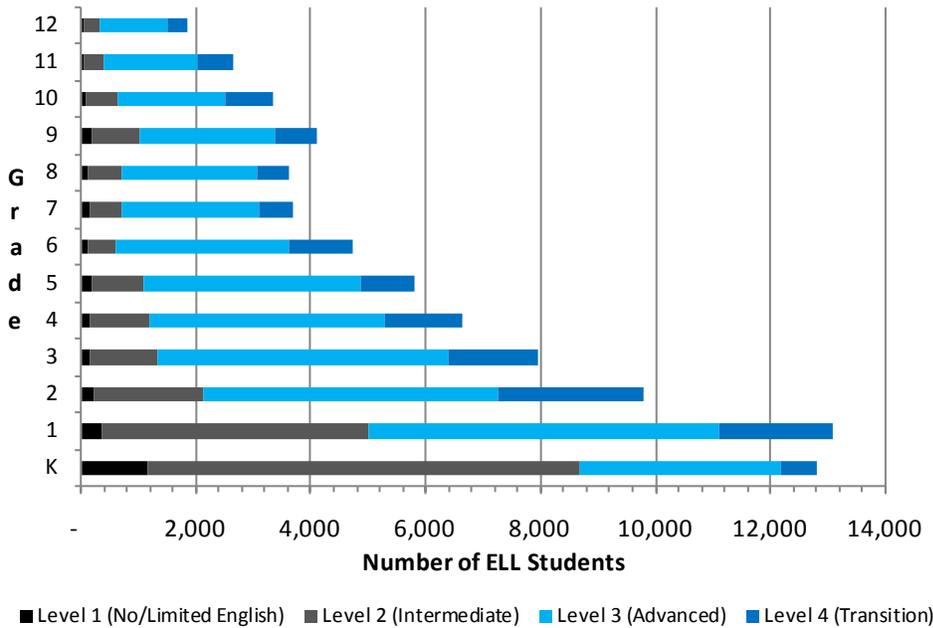
---

<sup>4</sup> OSPI identified the proportion of language proficient students (Level 4 on the WLPT) who were at or exceeded the scale score on the WASL by a standard error of measurement (SEM), for each grade level in both reading and math. It then found the scale score corresponding to the same proportion of students at each of the language proficiency levels (Level 2 and 3 of the WLPT). For example, the standard scale score of 400 plus the SEM (13) was 413 for Grade 3 reading in 2008, and 42% of all ELLs in Level 4 had a scale score of at least 413. The corresponding scale score for the top 42% of the students who were in Level 2 of the WLPT was 359; the corresponding scale score for top 42% of the students who were in Level 3 of the WLPT was 388.

<sup>5</sup> Analyses of a district with a high level of ELL students found that both methods improved the percentage of non-low income students meeting standard in the various content areas by 1 percentage point and the percentage of low income students meeting standard by about 7 percentage points (ELL students are more likely to be low income). The overall index for this district improved by 0.1 on a 7-point scale using both methods.

<sup>6</sup> About 70% of all ELL students enter school in kindergarten, and they will have attended school for three years before taking the state assessment for the first time in grade 3. Of the ELL students who were enrolled in grades assessed by the WASL/WAAS (grades 3-8 and 10), more than 81% had reached the advanced level of the WLPT in 2008 and would have their scores included in the accountability calculations.

**Figure 1: WLPT Results in 2008, by Grade and Language Proficiency Level**



**RECOMMENDATIONS**

- 1. WASL/WAAS results should be excluded from the accountability calculations for ELLs who are in the first three calendar years of U.S. public school enrollment or until Level 3 is reached on the WLPT, whichever comes first.**

This “extended exemption” policy has been discussed with stakeholders around the state and has been well received.<sup>7</sup> It is the same approach repeatedly taken by OSPI in its proposal to the Department, and it is consistent with the empirical evidence about how long it takes students to acquire enough language skills to meet state standards. Moreover, it is simple to understand and lowers the stakes for students who have not yet acquired the necessary English language skills to show what they know and can do on the WASL. While it has relatively little impact on the index results, it still increases the validity and perceived fairness of the results. ELLs would still continue to take the state tests and their results would be reported after the first year of enrollment. If they meet standard before the exclusion period ends, their results would be included in the index results. This provides an incentive to help ELLs acquire the necessary language and content skills they need as soon as possible. Finally, the use of the “calendar year” is the easiest method for determining the length of enrollment.<sup>8</sup> Some ELLs enroll for part of a year, then move and may return later. This means some ELLs may not have been served by a U.S. public school for three full academic years during the three calendar years. However, keeping track of the number of days of enrollment before counting the test results for accountability purposes would impose more complicated reporting requirements on schools and districts.

<sup>7</sup> Some stakeholders believe the exclusion should be for a longer period of time (e.g., 4-5 years), while others believe that any exclusion beyond the first year implies the state is less interested in the performance of these students.

<sup>8</sup> The date of entry into a U.S. school is captured in the home language survey related to the ELL program, but the accuracy of the data is unknown.

## **2. WLPT results should be made public on OSPI's on-line Report Card.**

OSPI currently reports ELL results from the WASL on its Report Card, and it posts an annual report in PDF format with some WLPT results on the state “transitional bilingual education program.” However, no WLPT results are reported on the Report Card for the state, districts, or schools. While the information is available, OSPI says it has not had the resources to post the results on the Report Card in the same way that WASL and WAAS results are posted. WLPT data files are not available for downloading like the other sets of data.

To improve accountability for helping ELLs, at a minimum OSPI should begin reporting the percentage of students achieving at each WLPT level in each subject and grade, data on the length of time students are enrolled in the program, and the Annual Measurable Achievement Objectives (AMAOs) required by the U.S. Education Department as part of Title III. Districts are required to publish their AMAO results, so having OSPI publish the results would reduce districts’ reporting burden. Making the results public and accessible will provide more transparency and accountability for helping ELLs. Often, simply making the results public has a positive impact on student outcomes.

## **3. OSPI should provide information to help districts and school know if ELLs are on track to meet standard based on their WLPT and WASL results.**

This recommendation is consistent with the “sliding scale” proposal that OSPI submitted to the Department. Many stakeholders saw the benefit of this information and believe it would be helpful when determining if ELLs are on track to meet state standards based on their English language ability. OSPI could determine if it wanted to use a fixed set of scale scores to improve simplicity or calculate the results each year for each grade and subject, as was done for its initial proposal. State, district, and school results on the percentage of students who are “on track” to meet standard could be posted on the OSPI Report Card. These and the other WLPT results would be used in OSPI’s annual evaluation of the state program, and they would be reviewed as part of the analysis of schools and districts that do not make AYP two years in a row. Appendix A provides more information about this approach.

## APPENDIX A

### **Merits and Limitations of the “Sliding Scale” Approach**

***Sliding Scale Approach** Count as proficient the ELLs who are in their second year of U.S. public school enrollment who meet a lower scale score, based on their WLPT level.*

This approach was viewed favorably by a number of stakeholders. Under this approach, ELLs would continue to take the WASL and have their results reported and included in accountability calculations in their second year. It sends the message that the education system should support ELLs movement toward proficiency on the state content standards, even while their English ability is still being developed. It also gives educators an indication about whether ELLs are on track to meeting standard.

This option is less transparent and much more difficult to explain to the public. OSPI used a complicated method to determine the needed scale scores that, when reached, implies ELLs are “on track” to meet the standard by the time they reached Level 4 of the WLPT. The different scale scores that were set for each grade, subject, and WLPT level further hinders transparency and simplicity. Additional complexity occurs because writing has a different set of scale scores. Moreover, the required scores could change each year because the results change each year (this is partially a norm-referenced approach). Some believe this policy implies the state has lowered its standards for ELL students.

To remove much of the complexity, stakeholders suggested using the same scale scores for all grades and subjects. Table 1 shows a set of consistent scale scores that could be used.<sup>9</sup> Different cut scores are used for writing because the scales are different than the other subjects and because grade 10 uses a different scale than grades 4 and 7.

- ELLs scoring in *Level 4 of the WLPT* are required to reach the scale score that non-ELL students must reach to meet standard.
- The scale score needed for ELLs scoring in *Level 3 of the WLPT* is 10 points below the current scale score needed to meet standard in reading, math, and science (390 vs. 400).
- The scale score needed for ELLs scoring in *Level 2 of the WLPT* is 10 points below the current scale score needed to “partially meet standard” in reading, math, and science (365 vs. 375).
- ELLs scoring in *Level 1 of the WLPT*—those with no/limited English—would not have their WASL results counted in the accountability index calculations. (This also means a student could stay in Level 1 for a long time before the results are counted in the index.)

Table 2 shows the average scale score in the OSPI proposal. The average reading scores are almost identical to those in Table 1, but the average math scores in the OSPI proposal are 5-11 points below the reading average. No study has been done about the desired cut scores for writing and science.<sup>10</sup> The scale scores listed in Table 1 will be more challenging to attain in math and science. So if consistent scores are used, they will not always correspond to the performance level

---

<sup>9</sup> Since the test results change each year, the needed scale score could be determined using 3-year averages. However, for simplicity, it helps if the scores do not vary across grades and subjects.

<sup>10</sup> Writing and science are only tested in three grades. Only students in grade 10 are tested in all four subjects.

ELLs need to be “on track” to meet standard. In OSPI’s proposal, the required math scale scores are lower than the reading scores (the scale scores needed in science will be even lower than math, while the scores needed in writing will likely be comparable to the reading scores).

**Table 1: Proposed Scale Scores Needed for ELLs to be Considered Proficient**

WLPT Score	WASL scale score needed for “proficiency”		
	Reading, Math, and Science	Writing Grade 4,7	Writing Grade 10
Level 4*	400*	9*	17*
Level 3	390	8	15
Level 2	365	6	11
Level 1	—	—	—

\*Considered meeting standard for non-ELL students

**Table 2: Average Scale Scores in the OSPI Title III Proposal**

WLPT Level	Average WASL scale score for all grades in OSPI proposal		
	Reading	Math	Writing/Science
Level 4*	400*	400*	Not computed
Level 3	390	385	Not computed
Level 2	366	355	Not computed
Level 1	—	—	Not computed

\*Considered meeting standard for non-ELL students

Note: In *reading*, the needed scale scores in the seven tested grades in OSPI’s proposal ranged from 388 to 394 for Level 3 (a 6 point range) and from 359 to 374 for Level 2 (a 15 point range). In *math*, the needed scale scores in the seven tested grades ranged from 379 to 388 for Level 3 (a 9 point range) and from 346 to 362 for Level 2 (a 16 point range).