# WIDA Technical report

## Generating Imputed* Overall Composite Scale Scores for English Learners with Disabilities Who Are Missing Domain Scores in the ACCESS for ELLs Assessment

Narek Sahakyan

September 23, 2020

* In the original, 2020 report these scores were called "Alternate overall composite scale scores". To avoid confusion with the Alternate Access assessment, in 2022 WIDA decided to rename these scores into "Imputed overall composite scale scores. The terminology in this report has been updated in 2022 to reflect this change.

## Contents



## Figures



## Tables

## Executive Summary

We present a new approach to generating Imputed overall composite scale scores (IOCSS) for those English Learners who are identified as students with disabilities and cannot participate in all of the four test domains - Reading, Speaking, Listening and Writing - on the annual ACCESS for ELLs assessment because of their disability. The approach is based on reweighting of the non-missing domains when calculating overall composite scale scores for students who are missing scores in one or two test domains. To evaluate the proposed method, we use census data from the 2018-2019 ACCESS for ELLs assessment, to simulate test scores and compare these to students' observed test scores. We then generate a random sample from the population of students missing domain scores in one or two domains. Next, we compute simulated Imputed overall composite scale scores generated by the proposed reweighting method. These reweighted scores are then compared to those produced by a method that WIDA has used since 2018, based on imputing missing scores using the relative rank of similar students' performance in the missing domain.[1] Finally, we examine differences in each method's properties and examine correlations between the two approaches. The results show that the proposed reweighting method, while relying on fewer assumptions and requiring fewer steps to implement, provides similar (highly correlated) outcomes in terms of generated Imputed overall composite scale scores (IOCSS) and associated Imputed overall composite proficiency levels (IOCPL). [2]

[1] Porter, T., Cook, H. G., & Sahakyan, N. (2019). Less than four domains: Creating an overall composite score for English learners with Individualized Education Plans. Retrieved from: https://wcer.wisc.edu/docs/working-papers/Working_Paper_No_2019_3.pdf

[2] In conjunction with this technical report WIDA's Technical Assistance department has developed a Microsoft Excel based tool/spreadsheet that may be used to easily generate Imputed overall composite scale scores for eligible students with missing domain scores under the proposed reweighting approach. The tool is available on the WIDA SEA Secure Portal. Please contact Narek Sahakyan for questions and comments at: sahakyan@wisc.edu

## Introduction

For English learners (ELs) taking the ACCESS for ELLs annual summative assessment (ACCESS), overall composite scale scores are generated based on a weighted formula. Students' observed scale scores in the Reading, Speaking, Listening and Writing domains are assigned the weights of 0.35, 0.15, 0.15 and 0.35, respectively. However, a simple application of a weighted average is impossible in calculating composite scores for those students who did not take all four test domains and are therefore missing test scores in one or more domains. The most common reason for missing test scores is related to student disabilities, Individualized Education Plan (IEP) designations and 504 plans. If a student cannot be fairly assessed in a given domain, the requirement that a student take that specific domain of ACCESS may be waived. Before 2018, no method of deriving overall composite scale scores and associated proficiency levels was provided by WIDA for students missing test scores in one or more of the four domains. In 2019, WIDA responded to member-states' requests for guidance by suggesting four options, three of which included recalculation of student scores. The most rigorous of these three options (model 3), is based on domain performance ranking through z-scores and imputation. We refer to this approach as the ranking method.[3] While the ranking method produced consistent scores, the technical difficulties and the number of steps required to create the scores were daunting. Given the current methods' computational complexity, in 2020 WIDA revisited the "less than four domains" (LT4D) approach.[4] This report describes a new approach to generating Imputed composite scale scores and associated imputed overall composite proficiency levels, which we call the reweighting method. Our analysis compares student outcomes that would be generated by the proposed reweighting method, first to observed student population composite scale scores (from ACCESS 2018-2019), and second, for a smaller subset of randomly selected students, to scores generated by the ranking method. The next section describes our proposed method for reweighting and producing imputed overall composite scale scores for students who did not take one or two of the four domains on ACCESS due to disabilities.

## The Reweighting Method

The reweighting method has a number of advantages, which include relying on fewer assumptions, requiring fewer steps to implement, and avoiding an involved manual process of scoring through by-grade and by-domain lookup tables that need annual updating.[5] Our approach is based on modifying WIDA's existing formula for calculating overall composite scale scores (i.e. Reading - 35%, Listening - 15%, Speaking - 15% and Writing – 35%) for those English learners who are missing test scores in one or two language domains. Generally, the reweighting method equally redistributes the weight of the missing domain(s) to the non-missing one(s). More specifically, the "equal reassignment of domain weights" rule applies for the cases when there is one missing domain, but not for all cases when there

[3] The fourth, standard setting approach, while the most difficult and costly to implement, is the preferred option as it would provide the most accurate and reliable method to assess the overall proficiency for students who do not take all domains of ACCESS for ELLs.

[4] Additionally, insightful external advice suggested that more parsimonious alternatives could be explored.

[5] Some of the assumptions underlying the ranking method are: a) the unweighted average of the student's relative rank (as measured by z-scores) in the non-missing domains is a good proxy for their would-be relative rank in the missing domain; b) students with a missing domain, had they been able to take it, would have a "similar" rank in the missing domain compared to students with no missing domains, c) the (averaged) z-score to scale score crosswalk is smooth (there are many possible z-values within grades with no large discontinuities).

are two missing domains. For example, in those cases when there are two domains missing being either Listening or Speaking, an equal reassignment of weights would result in a 60/40 redistribution, while our proposed approach recommends a 70/30 reassignment of domain weights. This slight modification is introduced to keep the structure of the reweighted Imputed composite scores identical to WIDA composite domain (i.e. Oral, Comprehension and Literacy) scores.[6] Table 1 provides the proposed redistribution rules for all ten possible one or two missing domains combinations.

**Table 1: Reweighting of domains for students with missing test scores in one or two domains.**

| Missing Domains | | Reading | Writing | Listening | Speaking |
|---|---|---|---|---|---|
| **None (no reweighting)** | | **35** | **35** | **15** | **15** |
| **One Missing Domain** | **Reading** *based on W, L and S* | N/A | 46 (*+11*) | 27 (*+12*) | 27 (*+12*) |
| | **Writing** *based on R, L and S* | 46 (*+11*) | N/A | 27 (*+12*) | 27 (*+12*) |
| | **Listening** *based on R, W and S* | 40 (*+5*) | 40 (*+5*) | N/A | 20 (*+5*) |
| | **Speaking** *based on R, W, and L* | 40 (*+5*) | 40 (*+5*) | 20 (*+5*) | N/A |
| **Two Missing Domains** | **Reading and Writing** *based on L and S: Oral* | N/A | N/A | 50 (*+35*) | 50 (*+35*) |
| | **Reading and Listening** *based on W and S* | N/A | 70 (*+35*) | N/A | 30 (*+15*) |
| | **Reading and Speaking** *based W and L* | N/A | 70 (*+35*) | 30 (*+15*) | N/A |
| | **Writing and Listening** *based R and S* | 70 (*+35*) | N/A | N/A | 30 (*+15*) |
| | **Writing and Speaking** *based R and L: Comprehension* | 70 (*+35*) | N/A | 30 (*+15*) | N/A |
| | **Listening and Speaking** *based R and W: Literacy* | 70 (*+35*) | N/A | 30 (*+15*) | N/A |

The numbers in bold represent the proposed modified weights that can be used to calculate a composite scale score when one or two domains are missing. The italicized numbers in green show the increase in the domain weight as compared to the original, non-reweighted formula. Given in the top row are the unmodified weights, which are applied to calculate the overall composite scale scores of about 95% of all English learners who take the ACCESS assessment and complete all four of the domains. In the row below, included within the "One Missing Domain" section, the imputed overall composite scale score for a student missing the Reading domain test would be calculated based on a

[6] Otherwise, had we applied the "equal redistribution of domain weights" rule for all of the possible cases, the WIDA composite domain scores (Oral, Comprehension and Literacy) of students who are not missing a domain would be weighted and scored differently compared to those students who have taken the same two domains, but are missing the other two. This would raise various theoretical and empirical concerns.

0.46 Writing, 0.27 Listening and 0.27 Speaking weight assignment. Similarly, if a student is missing two domains, e.g. Reading and Listening, the calculation of her imputed overall composite scale score would be based on the redistributed weighs in Writing (0.70) and Speaking (0.30).

The numbers in bold represent the proposed modified weights that can be used to calculate a composite scale score when one or two domains are missing. The italicized numbers in green show the increase in the domain weight as compared to the original, non-reweighted formula. Given in the top row are the unmodified weights, which are applied to calculate the overall composite scale scores of about 95% of all English learners who take the ACCESS assessment and complete all four of the domains. In the row below, included within the "One Missing Domain" section, the imputed overall composite scale score for a student missing the Reading domain test would be calculated based on a 0.46 Writing, 0.27 Listening and 0.27 Speaking weight assignment. Similarly, if a student is missing two domains, e.g. Reading and Listening, the calculation of her imputed overall composite scale score would be based on the redistributed weighs in Writing
(0.70) and Speaking (0.30).

## Methods and Data

We use descriptive statistics (means, standard deviations, z-scores and frequencies) and correlations (Pearson, Spearman and Kendall) to examine and compare observed and simulated test scores. To check the consistency of imputed overall composite scale scores and proficiency levels which would be generated by the proposed method (based on the reweighting rules shown in Table 1), we perform a two-part analysis. In the first part, we use individual-level population data from the 2018-2019 ACCESS for ELLs test administration. A total of 2,077,506 student records with scale scores reported in all the four domains were available. For these two million-or-so students, who had actually taken all four domains, we set individual domains, and each possible two-domain combinations to "missing", thereby simulating all scenarios in which students hypothetically did not take one or two domains. Imputed overall composite scale scores and proficiency levels are then generated for each of the 10 possible missing domain(s) scenarios as listed in Table 1. We then compute Pearson correlation coefficients between the observed overall composite scale scores and each one of the simulated Imputed overall composite scale scores. These correlations measure the overall magnitude of departure of the simulated scores from observed scores due to the redistribution of domain weights.

The second part of the analysis focuses specifically on those students who were in fact missing test scores in one or two domains in their 2018-2019 ACCESS for ELLs assessment. These students (approximately 20,000) made up about 1% of the population. About three quarters, or approximately 15,000 students were missing test scores in one of the four domains, while the remaining students were missing scores in two of the four domains. Next, we take a 3% random, stratified (by grade) sample from these students and generate Imputed overall composite scale scores using the average rank-imputation method and the proposed domain reweighting method. Finally, we compare the simulated Imputed overall composite scale scores and respective proficiency levels generated by the two methods using Pearson, Spearman and Kendall correlations. The next section presents our findings.

## Findings

In the first part of the analysis we generate imputed overall composite scale scores for each of the ten possible cases where one or two domains could be missing. Next, we estimate Pearson product-moment correlations between 2018-2019 ACCESS overall composite scale scores, and each one of the ten simulated imputed overall composite scale scores for all possible one or two missing domain combinations. These correlations are presented in Table 2.

**Table 2. Pearson product-moment correlations between observed and simulated Imputed overall Composite Scale Scores.**

| Missing Domains | Pearson ρ (*CSS, Alt CSS*$_{\sim Domain}$) |
|---|---|
| ~Listening | 0.99 |
| ~Reading | 0.97 |
| ~Speaking | 0.99 |
| ~Writing | 0.98 |
| ~Reading and Writing | 0.86 |
| ~Reading and Listening | 0.96 |
| ~Reading and Speaking | 0.97 |
| ~Writing and Listening | 0.97 |
| ~Writing and Speaking | 0.97 |
| ~Listening and Speaking | 0.98 |

Table 2 provides evidence that the simulated imputed overall composite scale scores are very closely aligned with observed student test scores (i.e. 2018-2019 ACCESS overall composite scale scores). Pearson correlation coefficients given in Table 3 were very high ($\rho > 0.95$) but for the case when Reading and Writing were not taken and simulated overall composite scale scores were based on the two domains of Speaking and Listening ($\rho = 0.86$). [7] The high correlations reported between observed and simulated scores in Table 2 support the application of the reweighting method in generating imputed overall composite scale scores for students who are missing one or two domains. [8]

In the second part of the analysis we take a 3% random, stratified (by grade) sample from the population of 20,000 students who had scale scores missing in either one or two domains. For this random sample of 665 students, we generate simulated Imputed overall composite scale scores using both the existing average rank-imputation method and the proposed domain reweighting method. Table 3 provides a comparison of the means, standard deviations, and ranges of simulated overall composite scale scores generated by the two methods. The last column of the table gives the Pearson product-moment correlation coefficients between the scores obtained under the two methods.

[7] This is to be expected, since in the calculation of overall composite scale scores the domains of Listening and Speaking are assigned the weights of 0.15, while Reading and Writing carry heavier weights at 0.35.

[8] Appendix A presents a break-down of correlations by grade.

**Table 3: Descriptive statistics for the simulated imputed scores from Ranking and Reweighting methods.**

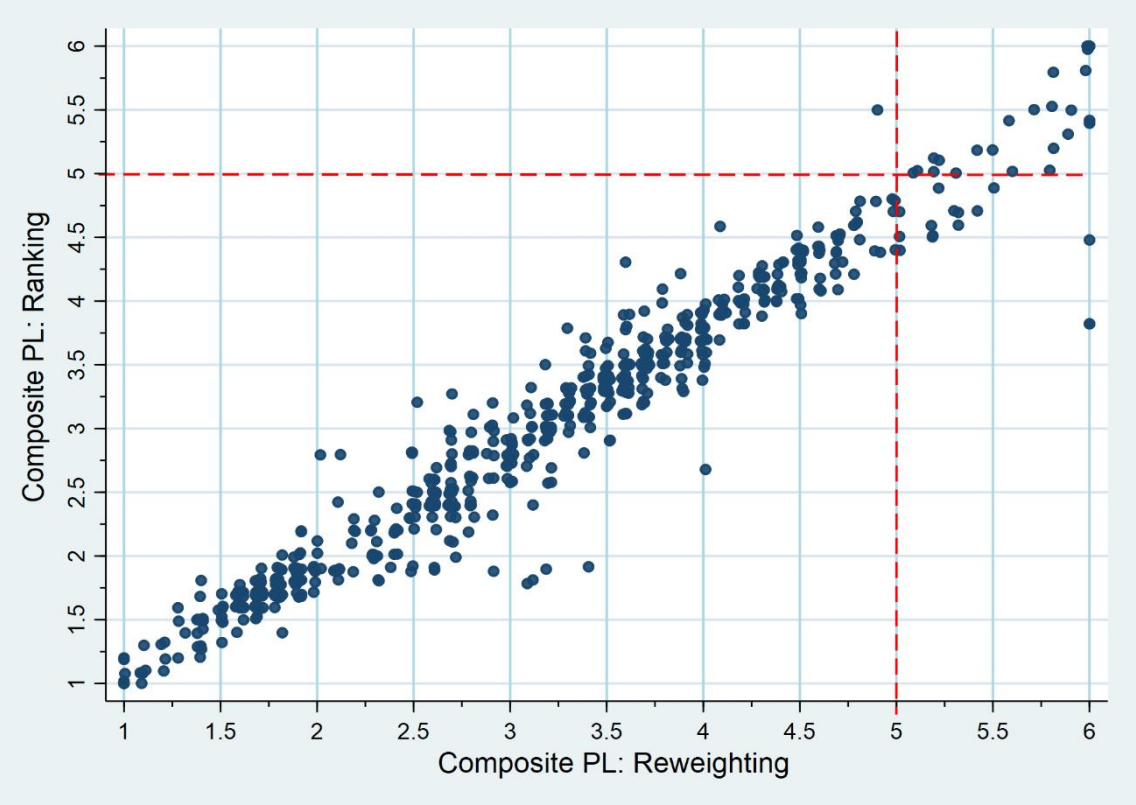
| Variable | Observations | Mean | Std. Dev. | Min | Max | Pearson ρ |
|---|---|---|---|---|---|---|
| **Combined one and two missing domains** | | | | | | N = 665 |
| Ranking Method | 665 | 295 | 57 | 103 | 430 | |
| Reweighting Method | 665 | 302 | 60 | 106 | 437 | **0.95** |
| **One Missing domain only** | | | | | | N = 537 |
| Ranking Method | 537 | 297 | 53 | 103 | 419 | |
| Reweighting Method | 537 | 302 | 55 | 106 | 433 | **0.98** |
| **Two Missing domains only** | | | | | | N = 128 |
| Ranking Method | 128 | 296 | 71 | 103 | 430 | **0.90** |
| Reweighting Method | 128 | 303 | 77 | 110 | 437 | |

Comparing the descriptive statistics of the scores generated by the two methods, we find that the reweighting method produced similar simulated scores, albeit with slightly higher means, standard deviations, and ranges. The Pearson correlation coefficients are very high ($\rho > 0.9$) and support the close alignment of the scores produced by the two methods.

Next, we generate grade-specific imputed overall composite proficiency levels, associated with the simulated imputed overall scale scores. Comparing the proficiency levels associated with the scale scores generated by the two different methods, we find high correlations.[9] Figure 1 provides a scatterplot of generated Imputed overall composite proficiency levels.

[9] Spearman rank correlation = 0.98, Kendall τ-a = 0.88, Kendall τ-β = 0.90.

**Figure 1: Scatterplot of generated Imputed overall composite proficiency levels associated with the Ranking and Reweighting methods. N = 665.**




In Figure 1 above, each of the dots corresponds to the test score generated by either method. The red dashed line indicates the overall composite proficiency level of 5.0, which represents the highest reclassification score in WIDA states. Of interest are the South-East and North-West quadrants separated by these lines. More specifically, the application of the reweighting method could potentially imply reclassification-level proficiency for 12 students who would not be deemed proficient under the ranking method, and for one student in the converse scenario. While a more proportionate distribution of scores in these quadrants would give more credence to the consistency of decisions produced by the two methods, the very small proportion (<2%) of students who may be potentially affected by the use of one method in favor of the other ameliorates these concerns.

In sum, our analysis shows that the two methods produce very similar and highly correlated Imputed composite scale scores and proficiency levels. The simulated scores and PLs are slightly higher for the reweighting method; however, it is important to note that it is based on only 2018-2019 data, and a multitude of factors, such as year to year changes in the demographic of the test-taking population, the relative difficulty of a specific test domain compared to others, as well as federal and state-level policies that determine which students can be waived from taking domains could affect the test scores generated by either method. Since the reweighting method relies on fewer assumptions, is simpler to implement, and produces similar (highly correlated) scores, this approach may be a more efficient way to produce imputed overall composite scale scores for students who are waived from taking one or two domains of ACCESS for ELLs due to a disability.

## APPENDIX A

Pearson Correlations between observed and imputed overall composite scale scores by grade,
ACCESS 2018-2019 administration

| Missing Domain /Grade | K | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | All |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **~Listening** | 0.99 | 0.98 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | **0.99** |
| **~Reading** | 0.95 | 0.94 | 0.96 | 0.96 | 0.97 | 0.97 | 0.96 | 0.97 | 0.98 | 0.97 | 0.97 | 0.97 | 0.97 | **0.97** |
| **~Speaking** | 0.98 | 0.97 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | **0.99** |
| **~Writing** | 0.96 | 0.94 | 0.95 | 0.95 | 0.95 | 0.96 | 0.96 | 0.97 | 0.97 | 0.96 | 0.96 | 0.96 | 0.95 | **0.98** |
| **~Reading and Writing** | 0.82 | 0.83 | 0.85 | 0.87 | 0.88 | 0.89 | 0.86 | 0.89 | 0.90 | 0.90 | 0.90 | 0.90 | 0.88 | **0.85** |
| **~Reading and Listening** | 0.95 | 0.91 | 0.93 | 0.93 | 0.93 | 0.93 | 0.94 | 0.95 | 0.95 | 0.94 | 0.93 | 0.92 | 0.91 | **0.96** |
| **~Reading and Speaking** | 0.95 | 0.92 | 0.94 | 0.94 | 0.95 | 0.95 | 0.95 | 0.95 | 0.96 | 0.95 | 0.94 | 0.94 | 0.93 | **0.97** |
| **~Writing and Listening** | 0.96 | 0.90 | 0.91 | 0.92 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.95 | 0.94 | 0.94 | 0.93 | **0.97** |
| **~Writing and Speaking** | 0.95 | 0.89 | 0.92 | 0.92 | 0.91 | 0.92 | 0.92 | 0.94 | 0.94 | 0.92 | 0.91 | 0.91 | 0.90 | **0.97** |
| **~Listening and Speaking** | 0.95 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 | 0.96 | 0.97 | 0.97 | 0.97 | 0.96 | 0.96 | 0.96 | **0.98** |
| | | | | | | | | | | | | | | |
| **Number of observations** | 232,806 | 239,716 | 240,967 | 235,378 | 218,845 | 178,063 | 128,936 | 116,839 | 108,144 | 121,462 | 101,395 | 90,036 | 64,919 | 2,077,506 |

Wisconsin Center for
Education Research
SCHOOL OF EDUCATION
UNIVERSITY OF WISCONSIN–MADISON

1025 W. Johnson St. | Madison, WI 53706-1706
(866) 276-7735 | help@wida.us | wida.wisc.edu