

Technical paper on weighting and sample outcomes approach

Assessments for Minimum Proficiency Levels (AMPL) ab
October 2023

The ACER Centre for Global Education Monitoring supports the monitoring of educational outcomes worldwide, holding the view that the systematic and strategic collection of data on education outcomes, and factors related to those outcomes, is required to inform high quality policy aimed at improving educational progress for all learners.

Acknowledgments

This project is referred to as the AMPL (Assessments for Minimum Proficiency Levels) project. This UNESCO Institute for Statistics (UIS) project is funded by the Global Partnership for Education (GPE).

The Australian Council for Educational Research (ACER) is the technical partner for this project. Support is provided from the Global Education Monitoring (GEM) Centre, an ACER initiative in partnership with the Australian government's Department of Foreign Affairs and Trade. The GEM Centre is also contributing to the UIS Global Item Bank. Technical and implementation support, and contribution to the assessment item pool, is provided by CONFEMEN.

Contents

Preface.....	1
Background.....	2
Introduction.....	2
Section 1: Base Weights.....	3
School Base Weight (WGTFAC1).....	3
Class base Weight (WGTFAC2).....	3
Student base weight (WGTFAC3).....	4
Section 2: Adjustment for Non-Response.....	5
School non-response adjustment (WGTADJ1).....	5
Student non-response adjustment (WGTADJ2).....	5
Trimming of weights.....	6
Final student weight (TOTWGTS).....	6
Section 3: Variance Estimation.....	7
Section 4: Participation Statistics.....	9
Participation Rates.....	10
Unweighted School Participation Rate.....	10
Unweighted Student Participation Rate.....	11
Overall Unweighted Participation Rate.....	11
Weighted participation rates including substitute schools.....	11
Weighted response rates excluding substitute schools.....	12

Preface

This document is near identical to deliverable '4.3 – Survey weights and response rates overview paper.' The aforementioned deliverable proposed the sampling approach related to survey weights and response rate. This deliverable describes what actually occurred. There were no changes to the proposed approach, that is, we undertook exactly what we proposed. Hence, the only changes relate to describing activities in the past rather than future tense.

Background

As part of SDG 4, Indicator 4.1.1 aims to measure the “proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex.” To meet this goal, UIS has coordinated efforts to establish common reading and mathematics scales for all three points of Indicator 4.1.1, building on existing cross-national and national assessments. As a result of these efforts, two important points of consensus have been reached: the definition of the Minimum Proficiency Level (MPL) and the Global Proficiency Framework (GPF). The MPL has been so far identified in each of the major cross-national assessments, based on the proficiency level descriptors in each of them.

The overarching objective of the AMPLab project is to measure and analyze the performance of students at the end of lower and upper primary using an assessment that aligns with GPF Level 2. This will:

- enable the collection of more informative data about where students are performing in terms of the GPF in reading and mathematics,
- produce baseline measures to set targets and compare learning gains/losses
- facilitate reporting on SDG 4.1.1
- aid the tracking of learning progress over time
- complement tools that had been already developed in 2021 in the Monitoring the Impacts on Learning Outcomes (MILO) study.

Introduction

This document describes the methodology and processes for calculating weights and generating replicates for the purposes of variance estimation for AMPLab. In developing these methods, we have drawn from the methodologies of other large scale international education surveys and tailored these to the specific characteristics of the AMPLab study.

Section 1 describes the process used to calculate sample weights. Section 2 describes adjustments for non-response and Section 3 explains the method for estimation of the variance of sample estimates, which involves a replication method. Section 4 explains the calculation of response rates and participation statistics.

Section I: Base Weights

The final student weight is a product of several weight components, associated with each stage in the sample selection process. For AMPLab, three main weights have been estimated: a School, Class and a Student weight.

There are two components for each stage:

- *Base weights* reflect the selection probabilities of sampled units. At each level of sample selection, the base weight is the inverse of the selection probability of a sampled unit.
- *Non-response adjustments* account for the non-participation of sampled units at each stage.

School Base Weight (WGTFAC1)

The School Base Weight reflects the probability of selection of each school within each explicit stratum.

The first stage of sampling involves selecting schools from the population. The school base weight reflects the selection probabilities of this sampling step. When explicit stratification is used, the school samples are selected independently within each explicit stratum h , with $h=1, \dots, H$. If no explicit strata are formed, the entire site is regarded as one explicit stratum.

A systematic sample of schools is selected with the selection probability of school i being proportional to its school size. The measure of school size for school i in stratum h (mos_{hi}) is based on the number of students in the target grade at that school.

The school base weight is defined as the inverse of the school's selection probability. For school i in stratum h , the school base weight, $WGTFAC1_{hi}$, is given by:

$$WGTFAC1_{hi} = \frac{MOS_h}{n_h^s \times mos_{hi}} \quad \text{Eq. 1}$$

Where n_h^s is the number of sampled schools in stratum h ; MOS_h is the total number of students enrolled in the schools of explicit stratum h , and mos_{hi} is the measure of size of the selected school i from stratum h .

A special case in Eq. 1 occurs when schools are relatively large so that $mos_{hi} > (MOS_h/n_h^s)$. These correspond to so-called 'certainty' schools, that will be selected under any circumstance. In this case, the value of $WGTFAC1$ cannot fall below 1, so the value must be set to 1. This means the school represents only itself.

Class base Weight (WGTFAC2)

In each participating school, we assume that an entire class is selected with equal probability among all the classes of the eligible grade. The Class Base Weight is the inverse of that selection probability. In this case, this corresponds to the total number of classes in the school:

$$WGTFAC2_{hi} = nclass_{hi} \quad \text{Eq. 2}$$

Where $WGTFAC2_{hi}$ corresponds to the Class Base Weight for school i in stratum h , and $nclass_{hi}$ is the number of classes of the eligible grade in school i from stratum h .

Student base weight (WGTFAC3)

Finally, it is assumed that all students in the selected class are selected for the survey. Therefore, the students in the class are selected with certainty, with probability equal to one. The student Base Weight is the inverse of that selection probability, which is also equal to one.

For each sampled student k from class j in school i from stratum h , the student base weight, $WGTFAC3_{hijk}$ is defined as

$$WGTFAC3_{hijk} = 1 \quad \text{Eq. 3}$$

In cases where a sub-sample of a class is taken, the student base weight is adjusted to reflect the probability of a student being selected from the selected class.

Section 2: Adjustment for Non-Response

School non-response adjustment (WGTADJ1)

Some sampled schools might not participate in the survey or must be removed from the dataset due to low student participation¹. The school base weights are adjusted to account for the sample loss. Adjustments are calculated within non-response groups defined by the explicit strata. Within each explicit stratum, a school non-response adjustment, $WGTADJ1_{hi}$, is calculated for each participating school i in stratum h as:

$$WGTADJ1_{hi} = \frac{n_h^{s,e}}{n_h^p} \quad \text{Eq. 4}$$

Where $n_h^{s,e}$ is the number of sampled eligible schools and n_h^p is the number of participating schools in the survey in explicit stratum h .

The number $n_h^{s,e}$ in this section is not necessarily equal to n^s in the preceding section, as $n_h^{s,e}$ is restricted to schools deemed eligible to participate in the survey. Some selected schools might turn out to no longer be eligible for participation in the survey. This happens, for example, when a school has recently closed, does not have target students in the target grade, or has only excluded students enrolled. In these cases, the ineligible school is not considered when calculating the non-response adjustment.

Student non-response adjustment (WGTADJ2)

Since only one class will participate in the school, there is no need to adjust for class participation. However, it is possible that within the selected class, some students may not participate. Therefore, a second adjustment factor must be estimated. For all schools, the adjustment for student non-response for each participating student k in class j , $WGTADJ2_{hijk}$ is calculated as follows:

$$WGTADJ2_{hijk} = \frac{std_{hij}^e}{std_{hij}^p} \quad \text{Eq. 5}$$

Where std_{hij}^e is the number of eligible students and std_{hij}^p is the number of participating students from the selected class j in school i in stratum h . In the context of student weight adjustment, students of the target population are regarded as eligible if they had not been excluded due to valid exclusions categories, according to the technical standards, including physical and intellectual disability or insufficient language experience and if they had not left the sampled school.

¹ See Section 4: Participation Statistics for a description of participation standards.

Trimming of weights

A preliminary weight is calculated as the product of the base weights and non-response adjustments at each stage of selection.

$$\begin{aligned} WGT_{Prelim_{hij}} &= WGT_{FAC1_{hi}} * WGT_{FAC2_{hi}} * WGT_{FAC3_{hij}} \\ &* WGT_{ADJ1_{hi}} * WGT_{ADJ2_{hij}} \end{aligned} \quad \text{Eq. 6}$$

An analysis of the distribution of weights within each explicit stratum is undertaken to identify any outlier weights, defined as a weight that is larger than four times the median weight for the stratum². For most surveys, outlier weights occur rarely or not at all. They usually reflect operational issues such as problems with the sampling frame data. Outlier weights are undesirable because they open the possibility of certain sampled units having an excessive influence on survey outcomes. Large variations in weights also have a major impact on the precision of survey estimates. Any outlier weights have their weights trimmed to four times the median weight for the stratum.

Final student weight (TOTWGTS)

The final student weight, $TOTWGTS_{hijk}$, of each student k in class j of school i in stratum h is the product of the four weight and non-response adjustment components with any necessary weight trimming applied:

$$TOTWGTS_{hijk} = WGT_{Prelim_{hijk_trimmed}} \quad \text{Eq. 7}$$

² This definition is based on a convention used in most large scale educational surveys.

Section 3: Variance Estimation

Cluster sampling techniques permit an efficient and economic data collection. However, unlike for Simple Random Samples and some other simpler sampling schemes, ‘closed form’ formulae for calculating the sampling variance of population estimates are not usually available.

Replication techniques provide tools with which to estimate the sampling variance. The paired jackknife replication technique (JK2) is the replication method proposed for AMPLab. This method incorporates two main features of the sample design – the stratification of schools, and the clustering of students within schools.

JK2 requires the pairing of primary sampling units (PSUs) – in this survey, schools – into pseudo strata of two schools. These pseudo-stratum pairs of schools are constructed within explicit strata and are referred to as ‘sampling zones’. To preserve the stratification structure of the original sample, sampling zones are created in the order that the schools were sampled. In the event of an odd number of schools within an explicit stratum, the sampled units within the remaining school are randomly divided into two halves, thereby forming a sampling zone of two ‘quasi-schools’.

By default, the above process produces approximately half the number of sampling zones as eligible participating schools in the survey. In cases involving multiple survey participants - for example countries within an international survey – a decision may be taken to have a uniform number of sampling zones for each survey participant. This allows for efficient data processing.

If, the sampling zone formation for a particular country leads to more pairs than the number of sampling zones to be used, then schools from the excess pairs are simply combined with schools within the already formed zones. For example, if there will be 75 pairs across all survey participants, then the 76th pair from any one participant will be combined with the 1st sampling zone, the 77th pair will be combined with the 2nd sampling zone and so on.

A total of j replicate weight variables (or ‘replicates’), j equal to the number of sampling zones, are constructed using these variables. Each replicate weight is the array that represents the product of the final student weight and the replicate factors as described below, for that student’s school.

Within each sampling zone, j , from stratum h , the first school is randomly allocated to one of two states, with a School Replicate Indicator, u_{hj} , value of 1 or 0. The second school in the sampling zone takes the other value.

For each sampling zone, a replicate weight variable is formed by:

- doubling the weight of the school in the corresponding zone with replicate indicator value 1;

- reducing the weight to zero of the school in the corresponding zone with replicate indicator value 0;
- leaving unchanged the weights for schools in all other sampling zones.

Table 1 illustrates the way sampling zones, school replicate indicators and replicate factors are estimated, for the case of a database with 8 schools grouped in 4 sample zones.

Table 1: Example of the computation of Replicate Weights

ID	Student Weight	School	Sampling Zone	School Replicate Indicator	Replicate Weights			
					Zone 1	Zone 2	Zone 3	Zone 4
1	5.2	A	1	1	10.4	5.2	5.2	5.2
2	5.2	A	1	1	10.4	5.2	5.2	5.2
3	5.2	A	1	1	10.4	5.2	5.2	5.2
4	5.2	A	1	1	10.4	5.2	5.2	5.2
5	9.8	B	1	0	0.0	9.8	9.8	9.8
6	9.8	B	1	0	0.0	9.8	9.8	9.8
7	9.8	B	1	0	0.0	9.8	9.8	9.8
8	9.8	B	1	0	0.0	9.8	9.8	9.8
9	6.6	C	2	0	6.6	0.0	6.6	6.6
10	6.6	C	2	0	6.6	0.0	6.6	6.6
11	6.6	C	2	0	6.6	0.0	6.6	6.6
12	6.6	C	2	0	6.6	0.0	6.6	6.6
13	7.2	D	2	1	7.2	14.4	7.2	7.2
14	7.2	D	2	1	7.2	14.4	7.2	7.2
15	7.2	D	2	1	7.2	14.4	7.2	7.2
16	7.2	D	2	1	7.2	14.4	7.2	7.2
17	4.9	E	3	1	4.9	4.9	9.8	4.9
18	4.9	E	3	1	4.9	4.9	9.8	4.9
19	4.9	E	3	1	4.9	4.9	9.8	4.9
20	4.9	E	3	1	4.9	4.9	9.8	4.9
21	8.2	F	3	0	8.2	8.2	0.0	8.2
22	8.2	F	3	0	8.2	8.2	0.0	8.2
23	8.2	F	3	0	8.2	8.2	0.0	8.2
24	8.2	F	3	0	8.2	8.2	0.0	8.2
25	9.5	G	4	1	9.5	9.5	9.5	19.0
26	9.5	G	4	1	9.5	9.5	9.5	19.0
27	9.5	G	4	0	9.5	9.5	9.5	0.0
28	9.5	G	4	0	9.5	9.5	9.5	0.0

For each replicate therefore, the students from one school are ‘zero-weighted’, i.e., their contribution toward the estimated outcome is removed, its pair (from the same sampling zone) is ‘double-weighted’, and all other students across the other sampling zones contribute according to their final student weight. The formation of replicates can be thought of as a systematic approach to subsampling the final student sample.

The variance in outcomes from this set of subsamples is an estimate of the variance that is associated with the full sample from the population, which is the variance we are attempting to estimate.

$$\widehat{SV}_\mu = \sum_{i=1}^j [m_i - m_s]^2 \quad \text{Eq. 8}$$

Where m_s is the statistic μ estimated for the population through use of the original sampling weights and m_i is the same statistic estimated by using the weights for the i^{th} of j Jackknife replicates (i.e. 1 replicate for each of the j sampling zones). The standard error SE for statistic μ , which reflects the uncertainty of the estimate due to sampling, is computed as:

$$\widehat{SE}_\mu = \sqrt{\widehat{SV}_\mu} \quad \text{Eq. 9}$$

The computation of sampling variance using Jackknife replication can be obtained for any statistic, including means, percentages, standard deviations, correlations, regression coefficients, and mean differences. Standard statistical software does not always include procedures for replication techniques.

Section 4: Participation Statistics

During data collection, some sampled schools and students may, for various reasons, not participate in the survey. In some instances, selected schools may turn out to have no eligible students, or may have recently closed. In the context of a large scale educational survey, it is not unusual to find a small number of such instances. In effect, the survey has uncovered an aspect of the dynamic nature of an educational system, for example a school closure since the time the sampling frame was developed. In other cases, the school has eligible students but refuses or is unable to participate. In these cases, a similar school, systematically assigned at the time of school sampling, may be used as a substitute. If neither the sampled school nor an assigned substitute school participates, the sampled school is classified as a non-respondent.

According to the AMPLab Technical Standards, rates within schools will be measured against certain benchmarks, with consequences when these rates fall to low levels, because of the increased risk of non-response bias in these situations. If fewer than 50% of sampled eligible students respond to the survey, the school is considered as a non-

respondent for the purpose of estimating participation rates³. Furthermore, schools with less than 25% student participation are removed from the database and weights for the remaining schools from the stratum are adjusted to take into account their non-response⁴.

Within participating schools some students may be found to be ineligible, or may be identified as within-school exclusions, for example because of a functional or intellectual disability or insufficient language experience⁵. Other students may not participate due to absence or refusal. All eligible students in the school are classified as non-respondents if they are absent from the survey session or are present but do not manage to complete at least 10% of the of the assessment items⁶.

Participation Rates

AMPLab will report weighted and unweighted participation rates for schools and students, as well as overall participation rates that are a combination of both. To distinguish between participation based solely on originally sampled schools and participation that also relies on substitute schools, school and overall participation rates will be computed separately for originally sampled schools only and for originally sampled together with substitute schools.

Unweighted School Participation Rate

The unweighted school participation rate is the ratio of the number of participating schools to the number of originally sampled schools, excluding any sampled schools found to be ineligible. A school is considered to be a participating school if it has a student participation rate of at least 50 percent. The two unweighted school participation rates are calculated as follows:

R_{unw}^{sc-s} unweighted school participation rate for originally sampled schools only

R_{unw}^{sc-r} unweighted school participation rate, including originally sampled and first and second substitute schools

$$R_{unw}^{sc-s} = \frac{n_s}{n_s + n_{sub1} + n_{sub2} + n_{nr}} \quad \text{Eq. 10}$$

$$R_{unw}^{sc-r} = \frac{n_s + n_{sub1} + n_{sub2}}{n_s + n_{sub1} + n_{sub2} + n_{nr}} \quad \text{Eq. 11}$$

³ Technical Standards Assessment of Minimum Proficiency Level (AMPLab), Standard 1.12, Note 1.3

⁴ Technical Standards Assessment of Minimum Proficiency Level (AMPLab), Standard 1.12, Note 1.2

⁵ Technical Standards Assessment of Minimum Proficiency Level (AMPLab), Standard 1.12, Note 1.4

⁶ Technical Standards Assessment of Minimum Proficiency Level (AMPLab), Standard 1.12, Note 1.1

Where n_s is the number of sampled participating schools; n_{sub1} and n_{sub2} are respectively the first and second substitute schools that participated in the survey; and n_{nr} the number of sampled schools that did not participate and were not able to be replaced by their corresponding substitute schools⁷.

Unweighted Student Participation Rate

The unweighted student participation rate is the ratio of the number of selected students that participated in AMPLab to the total number of eligible, sampled students from the participating schools. The unweighted student participation rate R_{unw}^{st} is computed as follows:

$$R_{unw}^{st} = \frac{\sum_{i,j}^{s+sub1+sub2} S_{rs}^{i,j}}{\sum_{i,j}^{s+sub1+sub2} S_{rs}^{i,j} + \sum_{i,j}^{s+sub1+sub2} S_{nr}^{i,j}} \quad \text{Eq. 12}$$

Where $S_{rs}^{i,j}$ is the number of respondent students from school j in stratum i ; $S_{nr}^{i,j}$ the number of non-respondent eligible students from school j in stratum i .

Overall Unweighted Participation Rate

The overall unweighted participation rate is the product of the unweighted school and student participation rates. Because AMPLab will compute two versions of the unweighted school participation rate, one based on originally sampled schools only and the other including substitutes as well as originally sampled schools, there also are two overall unweighted participation rates:

R_{unw}^{ov-s} unweighted overall participation rate for originally sampled schools only

R_{unw}^{ov-r} unweighted overall participation rate, including originally sampled and first and second substitute schools

$$R_{unw}^{ov-s} = R_{unw}^{sc-s} \times R_{unw}^{st} \quad \text{Eq. 13}$$

$$R_{unw}^{ov-r} = R_{unw}^{sc-r} \times R_{unw}^{st} \quad \text{Eq. 14}$$

Weighted participation rates including substitute schools

For the weighted response rates, sums of weights are used instead of counts of schools and students. School and student base weights (BW) are the weight values before correcting for non-response, so they generate estimates of the population being represented by the responding schools and students. The final weights (FW) at the school and student levels are the base weights corrected for non-response.

⁷ This number includes those schools that participated but the student response rate fell below 50% of eligible students sampled for the survey.

Weighted school response rate including substitutes

School response rates are computed as follows:

$$RR_3^{sc} = \frac{\sum_i^{s+sub1+sub2} (BW_i \times \sum_j^{r_i} (FW_{ij}))}{\sum_i^{s+sub1+sub2} (FW_i \times \sum_j^{r_i} (FW_{ij}))} \quad \text{Eq. 15}$$

Where i indicates a school, $s + sub1 + sub2$ all responding schools, j a student, and r_i the responding students in school i . First, the sum of the student final weights FW_{ij} for the responding students from each school is computed. Second, this sum is multiplied by the school's base weight (numerator) or the school's final weight (denominator). Third, these products are summed over the responding schools (including substitute schools). The ratio of these values is the response rate.

Weighted student response rate including substitutes

The student response rate is computed over all responding schools. Of these schools, the population represented by the responding students is divided by the population being represented by the schools of those responding students. The numerator of the school response rate calculation (Equation 14) is the denominator of the student response rate:

$$RR_3^{st} = \frac{\sum_i^{s+sub1+sub2} (BW_i \times \sum_j^{r_i} (BW_{ij}))}{\sum_i^{s+sub1+sub2} (BW_i \times \sum_j^{r_i} (FW_{ij}))} \quad \text{Eq. 16}$$

Overall weighted student response rate including substitutes

The overall response rate is the product of the school and student response rates:

$$RR_3 = RR_3^{sc} \times RR_3^{st} \quad \text{Eq. 17}$$

Weighted response rates excluding substitute schools

Practically, substitute schools are excluded by setting their school base weight to zero and applying the same computations as above. More formally, the parts of the response rates are computed below.

Weighted school response rate excluding substitutes

$$RR_4^{sc} = \frac{\sum_i^s (BW_i \times \sum_j^{r_i} (FW_{ij}))}{\sum_i^{s+sub1+sub2} (FW_i \times \sum_j^{r_i} (FW_{ij}))} \quad \text{Eq. 18}$$

Weighted student response rate excluding substitutes

$$RR_4^{st} = \frac{\sum_i^s (BW_i \times \sum_j^{r_i} (BW_{ij}))}{\sum_i^s (BW_i \times \sum_j^{r_i} (FW_{ij}))} \quad \text{Eq. 19}$$

Overall weighted student response rate excluding substitutes

$$RR_4 = RR_4^{sc} \times RR_4^{st} \quad \text{Eq. 20}$$