ANNEX A3
Technical notes on analyses in this volume

STANDARD ERRORS, CONFIDENCE INTERVALS AND SIGNIFICANCE TESTS

The statistics in this report represent estimates based on samples of students, rather than values that could be calculated if every student in every country had answered every question. Consequently, it is important to measure the degree of uncertainty of the estimates. In PISA, each estimate has an associated degree of uncertainty, which is expressed through a standard error. The use of confidence intervals provides a way to make inferences about the population parameters (e.g. means and proportions) in a manner that reflects the uncertainty associated with the sample estimates. If numerous different samples were drawn from the same population, according to the same procedures as the original sample, then in 95 out of 100 samples the calculated confidence interval would encompass the true population parameter. For many parameters, sample estimators follow a normal distribution and the 95% confidence interval can be constructed as the estimated parameter, plus or minus 1.96 times the associated standard error.

In many cases, readers are primarily interested in whether a given value in a particular country is different from a second value in the same or another country, e.g. whether students in public schools perform better than students in private schools in the same country. In the tables and figures used in report, differences are labelled as statistically significant when a difference would be observed less than 5% of the time if there were actually no difference in corresponding population values (statistical significance at the 95% level). In other words, the risk of reporting a difference as significant when such difference, in fact, does not exist, is contained at 5%.

Throughout the report, significance tests were undertaken to assess the statistical significance of the comparisons made.

Statistical significance of differences related to type of school and differences between subgroup means

Differences in student performance by type of school or other indices were tested for statistical significance. Positive differences indicate higher scores for students in private schools while negative differences indicate higher scores for students in public schools. Generally, differences marked in bold in the tables in this volume are statistically significant at the 95% confidence level.

Similarly, differences between other groups of students (e.g. students in urban schools and students in rural schools, or socio-economically advantaged and disadvantaged students) were tested for statistical significance. The definitions of the subgroups can, in general, be found in the tables and the text accompanying the analysis. All differences marked in bold in the tables presented in Annex B of this report are statistically significant at the 95% level, unless otherwise indicated.

Statistical significance of differences between subgroup means, after accounting for other variables

For many tables, subgroup comparisons were performed both on the observed difference ("before accounting for other variables") and after accounting for other variables, such as the PISA index of economic, social and cultural status of students. The adjusted differences were estimated using linear regression and tested for significance at the 95% confidence level. Significant differences are marked in bold.

Statistical significance of performance differences between the top and bottom quartiles of PISA indices and scales

Differences in average performance between the top and bottom quarters of the PISA indices and scales were tested for statistical significance. Figures marked in bold indicate that performance between the top and bottom quarters of students on the respective index is statistically significantly different at the 95% confidence level.

Change in the performance per unit of an index

For many tables, the difference in student performance per unit of an index was calculated. Figures in bold indicate that the differences are statistically and significantly different from zero at the 95% confidence level.

ODDS RATIOS

The odds ratio is a measure of the relative likelihood of a particular outcome across two groups. The odds ratio for observing the outcome when an antecedent is present is simply

$$OR = \frac{(p_{11} / p_{12})}{(p_{21} / p_{22})}$$

where p_{11}/p_{12} represents the "odds" of observing the outcome when the antecedent is present, and p_{21}/p_{22} represents the "odds" of observing the outcome when the antecedent is not present.

Logistic regression can be used to estimate the odds ratio: the exponentiated logit coefficient for a binary variable is equivalent to the odds ratio.

Statistical significance of odds ratios

Figures in bold in the data tables presented in Annex B1 of this report indicate that the odds ratio is statistically significantly different from 1 at the 95% confidence level. To construct a 95% confidence interval for the odds ratio, the estimator is assumed to follow a log-normal distribution, rather than a normal distribution.

In some tables, odds ratios after accounting for other variables are also presented. These odds ratios were estimated using logistic regression and tested for significance against the null hypothesis of an odds ratio equal to 1 (i.e. equal likelihoods, after accounting for other variables).

OVERALL RATIOS AND AVERAGE RATIOS

In this report, the comparisons of ratios related to teachers, such as student-teacher ratio or the proportion of fully certified teachers, are made using overall ratios. This means, for instance, that the student-teacher ratio is obtained by dividing the total number of students in the target population by the total number of teachers in the target population. The overall ratios are computed by first computing the numerator and denominator as the (weighted) sum of school-level totals, then dividing the numerator by the denominator. Similar estimations are made for the proportion of teachers with at least a master's degree, the proportion of novice teachers, the proportion of fully certified teachers, participation in teacher training and teacher participation in selected professional development activities. In most cases (i.e. unless all schools are exactly the same size) this overall ratio differs from the average of school-level ratios.

USE OF STUDENT AND SCHOOL WEIGHTS

The target population in PISA is 15-year-old students, but a two-stage sampling procedure was used. After the population was defined, school samples were selected with a probability proportional to the expected number of eligible students in each school. Only in a second sampling stage were students drawn from amongst the eligible students in each selected school.

Although the student samples were drawn from within a sample of schools, the school sample was designed to optimise the resulting sample of students, rather than to give an optimal sample of schools. It is therefore preferable to analyse the school-level variables as attributes of students (e.g. in terms of the share of 15-year-old students affected), rather than as elements in their own right.

Most analyses of student and school characteristics are therefore weighted by student final weights (or their sum, in the case of school characteristics), and use student replicate weights for estimating standard errors.

In PISA 2018, as in PISA 2012 and 2015, multilevel model weights are used at both the student and school levels. The purpose of these weights is to account for differences in the probabilities of students being selected in the sample. Since PISA applies a two-stage sampling procedure, these differences are due to factors at both the school and the student levels. For the multilevel models, student final weights (W_FSTUWT) were used. Within-school weights correspond to student final weights, rescaled to amount to the sample size within each school. Between-school weights correspond to the sum of final student weights (W_FSTUWT) within each school.

STATISTICS BASED ON MULTILEVEL MODELS

Statistics based on multilevel models include variance components (between- and within-school variance), the index of inclusion derived from these components, and regression coefficients where this has been indicated. Multilevel models are specified as two-level regression models (the student and school levels), with normally distributed residuals, and estimated with maximum likelihood estimation. Models were estimated using the Stata (version 15.1) "mixed" module.

The intra-cluster correlation coefficient, or proportion of the variation that lies between schools, is defined and estimated as:

$$100 * \frac{\sigma_B^2}{\sigma_W^2 + \sigma_B^2}$$

where σ_B^2 and σ_W^2 represent the between- and within-variance estimates, respectively.
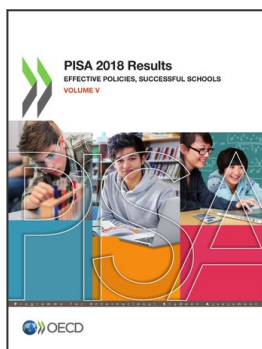
Standard errors in statistics estimated from multilevel models

For statistics based on multilevel models (such as the estimates of variance components and regression coefficients from two-level regression models) the standard errors are not estimated with the usual replication method, which accounts for stratification and sampling rates from finite populations. Instead, standard errors are "model-based": their computation assumes that schools, and students within schools, are sampled at random (with sampling probabilities reflected in school and student weights) from a theoretical, infinite population of schools and students, which complies with the model's parametric assumptions. The standard error for the estimated index of inclusion is calculated by deriving an approximate distribution for it from the (model-based) standard errors for the variance components, using the delta method.

References

OECD (forthcoming), *PISA 2018 Technical Report*, OECD Publishing, Paris.   [1]