Annex A

TECHNICAL NOTES ON ANALYSES IN THIS REPORT

Standard errors and significance tests

The statistics in this report (e.g. means and patterns of association between variables) represent estimates that are based on samples, rather than values that could be calculated if every student, teacher and principal in every country had answered every question. Consequently, it is important to measure the degree of uncertainty of the estimates when inferences are made about countries' populations of students, teachers and schools. In PISA, each estimate has an associated degree of uncertainty, which is expressed through a 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 the average class size in the modal grade for 15-year-olds increased between 2006 and 2015 or whether it is larger in the most advantaged schools than in the most disadvantaged schools. In the tables and charts used in this report, differences are labelled as statistically significant when a difference of that magnitude or larger would be observed less than 5% of the time, if there were actually no difference in corresponding population values. Similarly, the risk of reporting a correlation as significant if there is, in fact, no correlation between two measures, is contained at 5%. Throughout the report, significance tests were undertaken to assess the statistical significance of the comparisons made.

With the exception of statistics based on multilevel models (see below), the reported standard errors are computed with a balanced repeated replication (BRR) methodology. In some cases – e.g. where a proportion is estimated to be equal to 0% or 100%, or where the entire population of schools with a particular characteristic was included (census), rather than a sample, this method results in an estimated sampling uncertainty of 0. The corresponding estimate is replaced by a "c" in the tables, to indicate that there are too few observations to provide reliable estimates of the statistical uncertainty. Indeed, the lack of variation within the sample does not constitute definite evidence that there is no variation in the population at large; and even if there is no sampling uncertainty when a census is conducted, there may be other sources of uncertainty – such as misreporting, item non-response, etc. – that are not quantified in standard errors.

Overall ratios and average ratios

When comparing the student-teacher ratio or the proportion of fully certified teachers across countries or across types of schools, the quantity of interest is, in this report, the overall ratio that can be obtained by dividing the total number of students (or of fully certified teachers) in the target population by the total number of teachers in the target population. In most cases (i.e. unless all schools are exactly of the same size) this overall ratio differs from the average of school-level ratios.

This report estimates the overall student-teacher ratio, proportion of fully certified teachers, and proportion of teachers with a major in science from school samples, by first computing the numerator and denominator as the (weighted) sum of school-level totals, then dividing the numerator by the denominator.

Definition of advantaged and disadvantaged schools

Chapter 3 defines advantaged and disadvantaged schools in terms of the socio-economic profile and of the academic profile of schools. For both definitions, all schools in each PISA-participating education system are ranked by a composite score and then divided into four groups with approximately an equal number of students (quarters).

The socio-economic profile of a school is measured by the average PISA index of economic, social and cultural status (ESCS) of its 15-year-old students. Schools in the bottom quarter of average ESCS are referred to as "socio-economically disadvantaged schools"; and schools in the top quarter of average ESCS are referred to as "socio-economically advantaged schools".

The academic profile of a school is measured by the expected mean performance, given its students' socio-demographic characteristics (ESCS, immigrant background, language spoken at home, gender and age), computed with a regression model. This model ensures that each characteristic is weighted according to its country-specific importance in determining student disadvantage.

Pearson and Spearman correlation coefficients

Correlation coefficients measure the strength and direction of the statistical association between two variables. Correlation coefficients vary between -1 and 1; values around 0 indicate a weak association, while the extreme values indicate the strongest possible negative or positive association.

The Pearson correlation coefficient (indicated by the letter r) measures the strength and direction of the linear relationship between two variables. In this report, Pearson correlation coefficients are used to quantify relationships between country-level statistics. With only two variables (*x* and *y*), the r-square measure (indicated by R^2) of the linear regression of *y* on *x* (or, equivalently, of *x* on *y*) is the square of the Pearson correlation coefficient between the two variables.

The Spearman correlation is a measure of how well the relationship between two variables can be described by a monotonic (but not necessarily linear) function. Spearman correlation coefficients are computed in Chapter 3 at the school level. Because the Spearman correlation coefficient is based on the rank of each school within the sample rather than on its value, it is not affected by weights. Standard errors based on BRR weights could therefore not be computed.

Use of student, school and teacher weights

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

As an exception, estimates of "overall ratios" in which the denominator corresponds to the population of teachers (student-teacher ratios; proportions of fully certified teachers and proportion of science teachers with a university degree and a major in science) use school weights, which correspond to the inverse of the prior probability of selection for each selected school. Replicate school weights were generated for these analyses in analogy with the student replicate weights in the database, by applying the replicate factors observed for student weights within the school (one value among 0.2929, 0.5, 0.6464,1, 1.3536, 1.5 or 1.7071) to the base school weights (OECD, 2017, pp. 123-124[1])

Analyses based on teacher responses to the teacher questionnaires are weighted by student weights. In particular, in order to compute averages and shares based on teacher responses, final teacher weights were generated so that the sum of teacher weights within each school is equal to the sum of student weights within the same school. The same procedure was used to generate replicate teacher weights in analogy with the student replicate weights in the database. All science teachers within a school have the same weight, as do all non-science teachers within a school. Data for science and non-science teachers are analysed separately, as these define two distinct and non-overlapping populations for sampling.

For the computation of means, this is equivalent to aggregating teacher responses to the school level through simple, unweighted means, and then applying student weights to these school-level aggregates.

Statistics based on multilevel models

Statistics based on multilevel models include variance components (between- and within-school variance), the index of intra-class correlation derived from these components, and regression coefficients where this has been indicated.

Two-level models in Chapter 2

Multilevel models in Chapter 2 are specified as two-level regression models (the student and school levels, or the teacher and school levels), with normally distributed residuals, and estimated with maximum likelihood estimation. Where the dependent variable is science, reading, or mathematics performance, the estimation uses ten plausible values for each student's performance on the performance scale. Models were estimated using the Stata ® (version 15.1) "mixed" module.

In two-level models, weights are used at both the student/teacher and school levels. The purpose of these weights is to account for differences in the probabilities of students being selected in the sample.

For multilevel models where the dependent variable is at the student level, final student weights (W_FSTUWT) were used. Students' within-school weights correspond to student final weights, rescaled to amount to the sample size within each school. School weights correspond to the sum of final student weights within each school. This definition of school weights is the same used in the PISA 2015 and PISA 2012 Initial Reports.

For multilevel models where the dependent variable is at the teacher level, all teachers included in the model were assigned a weight equal to 1; school weights correspond to the sum of final student weights within each school.

The index of intra-class correlation is defined and estimated as:

$$100 * \frac{\sigma_w^2}{\sigma_w^2 + \sigma_b^2}$$

where σ_w^2 and σ_b^2, respectively, represent the within- and between-variance estimates.

Estimates based on multilevel models, and the between-school variance estimate in particular, depend on how schools are defined and organised within countries and by the units that were chosen for sampling purposes. For example, in some countries, some of the schools in the PISA sample were defined as administrative units (even if they spanned several geographically separate institutions, as in Italy); in others they were defined as those parts of larger educational institutions that serve 15-year-olds; in still others they were defined as physical school buildings; and in others they were defined from a management perspective (e.g. entities having a principal). The *PISA 2015 Technical Report* (OECD, 2017, pp. 86-88[1]) provides an overview of how schools are defined.

Because of the manner in which students and teachers were sampled, the within-school variation includes variation between classes as well as between students.

Three-level models in Chapter 4

Chapter 4 estimates a three-level regression model where the dependent variable is binary (expecting a career as a teacher). In this model, students (level 1) are nested within schools (level 2) and within countries (level 3). Because the dependent variable is binary, the regression model is specified as a

hierarchical generalised linear model (HGLM) with level 1 residuals following a Bernoulli distribution (Raudenbush and Bryk, 2002[2]). Normalised student weights were used in the estimation: these normalised weights ensure that each country contributes equally to the analysis (OECD, 2009, p. 219[3]). Models were estimated using HLM 7.

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 multilevel 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.

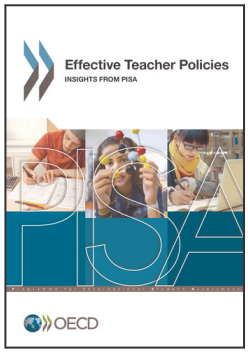
The standard error for the estimated index of intra-class correlation is calculated by deriving an approximate distribution for it from the (model-based) standard errors for the variance components, using the delta-method.

References

OECD (2009), *PISA Data Analysis Manual: SPSS, Second Edition*, PISA, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264056275-en. [3]

OECD (2017), *PISA 2015 Technical Report*, PISA, OECD Publishing, Paris, www.oecd.org/pisa/data/2015-technical-report/ (accessed on 31 July 2017). [1]

Raudenbush, S. and **A. Bryk** (2002), *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publications, Thousand Oaks. [2]



From:
Effective Teacher Policies
Insights from PISA

Access the complete publication at:
https://doi.org/10.1787/9789264301603-en

Please cite this chapter as:

OECD (2018), "Technical notes on analyses in this report", in *Effective Teacher Policies: Insights from PISA*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264301603-8-en

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at info@copyright.com or the Centre français d'exploitation du droit de copie (CFC) at contact@cfcopies.com.

OECD