Annex A

TECHNICAL ANNEX

SAMPLE DESIGN AND WEIGHTING PROCEDURES FOR THE PISA RE-ASSESSMENT

The PISA-24 sample is a representative sample of Canadian youth who were 15 years old in 2000. To ensure that the sample remains representative, the original PISA-15 sample weights applied to the data have to be modified to take into account attrition within YITS and the fact that only a sub-sample of YITS participants took part in the PISA-24 test.

The starting point for the creation of weights for PISA-24 is the final weight from YITS Cycle 5. To derive a final weight for the PISA-24, the following adjustments are applied to initial weights of the individual records on the PISA-24:

- Adjustment for sub-sampling of the Cycle 5 YITS sample

To select the PISA-24 sample, the YITS cycle 5 responding sample was stratified into 12 strata and a random sample was selected within each stratum. To adjust the initial weights to account for this sampling, the initial weight of each sampled unit in stratum h is multiplied by a factor equal to the number of Cycle 5 units in stratum h (N_h) divided by the number of units selected for the Reading Skills Re-assessment sample in stratum h (n_h).

- Adjustment for non-response

To adjust the weights for non-response to the PISA-24, a logistic regression is used to estimate the expected probability of response for each sample unit. Modelling is done within region as data allows (Atlantic, Québec, Ontario, Prairies, and British Columbia). To form response groups within which weight adjustments are to be made, the sample file is sorted by the estimated probability of response within each region. It is then divided into deciles, giving ten response adjustment groups for each region. Within each response adjustment group, the non-response adjustment factor is computed as the ratio of the sum of the weights for all units selected in the PISA-24 sample to the sum of the weights for all responding sample units.

- Adjustment for calibration to Cycle 5 gender totals

To bring estimates for the PISA-24 in line with YITS Cycle 5 estimates, a final calibration adjustment is made. Non-response weights are adjusted such that the survey weights sum to the same totals, by gender, as the Cycle 5 weights for all units in the Cycle 5 sample.

The final PISA-24 weight is the product of the initial weight multiplied by (1) the adjustment for sub-sampling of the YITS Cycle 5 sample, (2) the non-response adjustment, and (3) the calibration adjustment.

Although the participants in the PISA-24 tended to come from more advantageous socio-economic contexts than the original PISA-15 sample, population weights were calculated by Statistics Canada to adjust the representation of the current sample to the population represented by the PISA/YITS 2000 sample. The current study uses data from all cycles of YITS, including the original PISA-15 data.

DESCRIPTION OF INDICES EXAMINED IN THIS REPORT

Sense of mastery

The concept of mastery – the belief that one has control over one's destiny – was measured using the student questionnaire in YITS that was distributed with PISA-15 (Statistics Canada, 2005). The inspiration for this scale is the concept of external versus internal locus of control. A respondent's sense of mastery is defined as "the extent to which one regards one's life chances as being under one's own control in contrast to being fatalistically ruled" (Pearlin and Schooler, 1978). Individuals were asked about the extent to which they agreed with the following statements: "I often feel helpless in dealing with the problems of life"; "I have little control over the things that happen to me"; "There is little I can do to change many of the important things in life"; "There is really no way I can solve some of the problems I have"; "Sometimes I feel I'm being pushed around in life"; "I can do just about anything I really set my mind to"; and "What happens to me in the future mostly depends on me". The sense-of-mastery scale was then developed by combining their answers.

Teacher-student relations

Students' reports on their level of agreement with the following statements: Students get along well with most teachers; Most teachers are interested in students' well-being; Most of their teachers really listen to what they have to say; If they need extra help, they will receive it from their teachers; and Most of their teachers treat them fairly. Based on PISA-15 question ST30Q1-5 in the student questionnaire.

School achievement pressure

Students' reports on the frequency with which: the teacher wants students to work hard; the teacher tells students that they can do better; the teacher does not like it when students deliver careless work; and the teachers says that students have to learn a lot. Based on PISA-15 questions ST26Q2-4 and ST26Q15 in the student questionnaire.

Family educational support

Students' reports on the frequency that their parents and siblings work with them on their schoolwork. Based on PISA-15 question ST20Q01-3 in the student questionnaire.

Parental cultural communication

Students' reports on the frequency with which their parents/guardians discussed political or social issues with them; discussed books, films or television programmes; and listened to classical music. Based on PISA-15 question ST19Qo1-03 in the student questionnaire.

Total number of schooling hours per year

Instructional time for 15-year-old students in the school and derived hours of school per year. Based on PISA-15 question SC06Q01-3 in the school questionnaire.

Teacher morale

The extent to which school principals agreed with the following statements: The morale of the teachers in this school is high; Teachers work with enthusiasm; Teachers take pride in this school; and Teachers value academic achievement. Based on PISA-15 question SC20Q01-4 in the school questionnaire.

Shortage of teachers

The principals' views on how much learning by 15-year-old students was hindered by a shortage or inadequacy of teachers, in general, and in the specific courses of language, mathematics, and science. Based on PISA-15 question SC21Q01-4 in the school questionnaire.

School autonomy

School principals' reports on who had the main responsibility for school administration, staffing, compensation, financing, admission, teaching material and curriculum. A PISA index of school autonomy is derived from the number of categories that principals classified as not being a school responsibility. Based on PISA-15 question Based on PISA-15 questions SC22Q01-12 in the school questionnaire.

Teacher participation in decision making

An index of teacher autonomy was derived from the number of categories that principals classified as being mainly the responsibility of teachers. Based on PISA-15 questions SC22Q01-12 in the school questionnaire.

Student behaviour

Principals' reports on the extent to which learning by 15-year-olds in their school was hindered by: student absenteeism; disruption of classes by students; students skipping classes; students lacking respect for teachers; the use of alcohol or illegal drugs; and students intimidating or bullying other students. Based on PISA-15 questions SC19Q02,06,09,10,13,15 in the school questionnaire.

Teacher behaviour

Principals' reports on the extent to which learning by 15-year-olds was hindered by: low expectation of teachers; poor student-teacher relations; teachers not meeting individual students' needs; teacher absenteeism; staff resisting change; teachers being too strict with students; and students not being encouraged to achieve their full potential. Based on PISA-15 questions SC19Q01,03,07,08,11,14,16 in the school questionnaire.

Material resources

School principals' reports on the extent to which learning by 15-year-olds in their school was hindered by: lack of instructional material; not enough computers for instruction; lack of instructional materials in the library; lack of multi-media resources for instruction; inadequate science laboratory equipment; and inadequate facilities for the fine arts. Based on PISA-15 questions SCQ04-09 in the school questionnaire.

School size

School principals' report of the number of girls/boys enrolled in the school, the total enrolment, and the percentage of girls. Based on PISA-15 questions SC02Q01-02 in the school questionnaire.

Number of teachers

School principals' report of the number of full- and part-time teachers. Based on PISA-15 questions SC14Q01 in the school questionnaire.

Proportion of language teachers

Proportion of full- and part-time language teachers. Based on PISA-15 questions SC14Q07-08 in the school questionnaire.

MEASUREMENT ERROR AND REGRESSION TO THE MEAN

The results of PISA-24 suggest that larger gains in development of reading proficiency are made by individuals with the lower initial proficiency. The importance of this finding should not be overlooked, as it suggests that the disadvantages faced by many youth can be overcome with time. However, the problem of regression towards the mean clouds the analyst's ability to properly observe this relationship. This section looks more closely at the topic of skills convergence and the methodological issues with estimating skills growth using PISA-24. It then provides an option for dealing with regression towards the mean, which is used throughout the report.

The primary limitation to drawing inferences about changes in individuals is the imprecision with which proficiency is measured at the individual level. Although substantial documentation is available to describe in a precise manner what is meant by the construct of "reading proficiency" in the PISA assessment (OECD, 2001), because the construct must be measured by sampling behaviour that is merely indicative of its presence, there will always be imprecision in how it is assessed. Accordingly, an individual's score on the PISA-15 assessment and the PISA-24 assessment in 2009 should be though of as an imprecise signal of actual reading proficiency at ages 15 and 24 and of skills growth during this time.

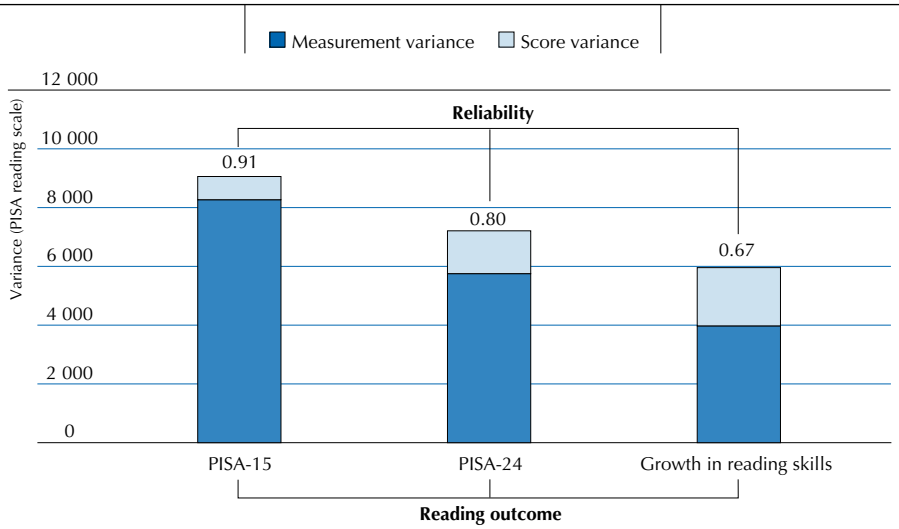
In general, comparability of measurements over time or across individuals can be achieved by increasing the number of observations (i.e. test items) and the similarity in how the observations are collected between the individuals being compared. Several factors with the different test administrations in this study determine the imprecision of the resulting scores: the rotated booklet design in PISA-15 increases the differences between individuals, and the relatively small number of test items in PISA-24 increases the relative impact of random mistakes to influence proficiency estimates. Both of these factors increase the measurement error.

Since the measurements at both time periods are imprecise, the difference between those estimates (i.e. skills growth) is even more imprecise. For example, random measurement error may have resulted in an overestimate for a student at age 15 and an underestimate at age 24, artificially deflating the perceived change in proficiency or skills growth. The reverse, an underestimation at age 15 and an overestimation at age 24, is also possible. This results in a range of possible score differences that is greater than the range of either of the individual scores.

The average measurement variance gives some indication of the relative accuracy of each of the three outcomes used in this study: reading proficiency at age 15, proficiency at age 24, and the change in proficiency between the two periods. One way to consider the measurement error is to show it as a proportion of the total variation in proficiency between individuals, a statistic known as the reliability. Estimates for the amount of change that has occurred, since it has the largest measurement error and the smallest total variation, has the smallest reliability among the three proficiency estimates. The numerical reliabilities for each outcome are displayed in each bar in Figure A.1. The value of 0.67 for change indicates that one third of the individual variations in estimated change in proficiency are due to chance.

■ Figure A.1 ■

Measurement variance, total variance, and scale reliability for reading outcomes based on PISA-15



Source: Cartwright (2012).

StatLink http://dx.doi.org/10.1787/888932576871

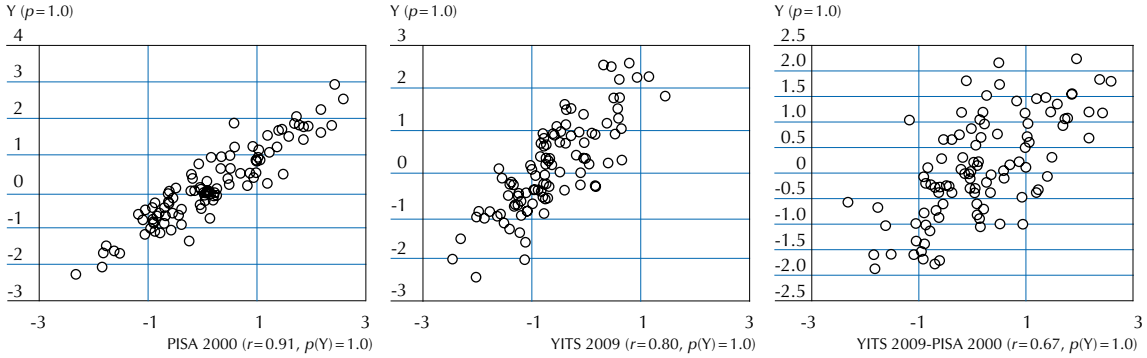
Box A.1 provides details on the analytic implications of measurement error.

Box A.1 The consequences of higher measurement error

The practical consequence of the lower level of reliability is that observed relationships appear weaker than they truly are. The three charts in Figure A.2 illustrate simulated data corresponding to variables that are theoretically perfectly correlated, but where one variable is measured with error. Each chart corresponds to one of the reading outcomes described in this study. From left to right, as reliability decreases, the observable relationship between the two variables decreases, from a clear linear pattern when r=0.91 to a vague cloud of data points when r=0.67. Since lower reliability decreases the maximum observable correlation between two variables, any relationship observed in the presence of measurement error can be considered a 'lower bound' on the true magnitude of the relationship. The consequences for the current study are that many relationships are found by estimating relationships between other variables and these three reading outcomes. The numerical quantification of these relationships is mathematically certain to be underestimated, so traditional metrics of gauging substantive importance may not apply.[1]

■ Figure A.2 ■

Illustration of attenuation of correlation due to measurement error using simulated data based on perfectly correlated variables and reliabilities of reading outcomes



Source: Cartwright (2012).

StatLink http://dx.doi.org/10.1787/888932576871

1. Although there are known corrections for the attenuating effect of measurement error, they are only relevant when all measurement error is known and when the statistic is a product-moment correlation; for the analyses in this study, either the adjustments would be constant to all comparable analyses, no adjustment is possible, or the analyses would be missing reliability estimates for other variables. For simplicity, all estimates have been left unadjusted.

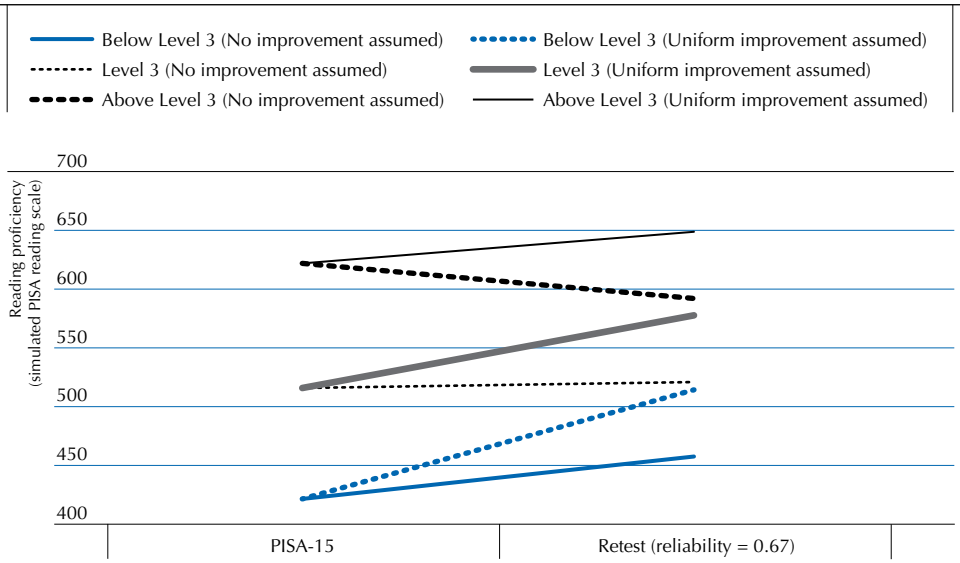
WHAT ARE THE CONSEQUENCES OF MEASUREMENT ERROR FOR A REPEAT ASSESSMENT?

Because measurement error is random, if an individual's proficiency is overestimated the first time it is measured, it will most likely be underestimated the second time. This implies that an individual who is far below or far above the average at the first point in time will tend to be closer to the average at the second point in time, a phenomenon known as "regression towards the mean". If there is a systematic change to the distribution between the two points in time, such as a global increase in scores, regression will be towards the new mean. The consequence of regression towards the mean is that measurement error in our estimate of proficiency growth is negatively correlated to PISA scores at age 15.

Because regression to the mean is a random phenomenon, it is not possible to determine how much any individual's estimated change in proficiency is a product of random variation versus true change. However, some expectations can be made at a group level. Using the initial results from the PISA-15 Canadian sample, regression to the mean for each proficiency level group is illustrated in Figure A.3 under two alternate assumptions: 1) no systematic change happens to the distribution; and 2) the distribution changes uniformly to match the mean and standard deviation of the PISA-24 sample. Under both assumptions, the group averages converge at the second test. Note that under the assumption of systematic uniform improvement, because all three groups are converging on a higher average, the improvement of the lowest group is artificially inflated, and the improvement of the highest group is deflated (in the absence of measurement error, all three groups would have the same growth trajectory). The reliability used to generate these random data is 0.67, the same as that for change in reading proficiency in this study.

■ Figure A.3 ■

Expected regression to the mean for PISA proficiency levels for Canadian PISA-15 participants based on retest reliability of 0.67

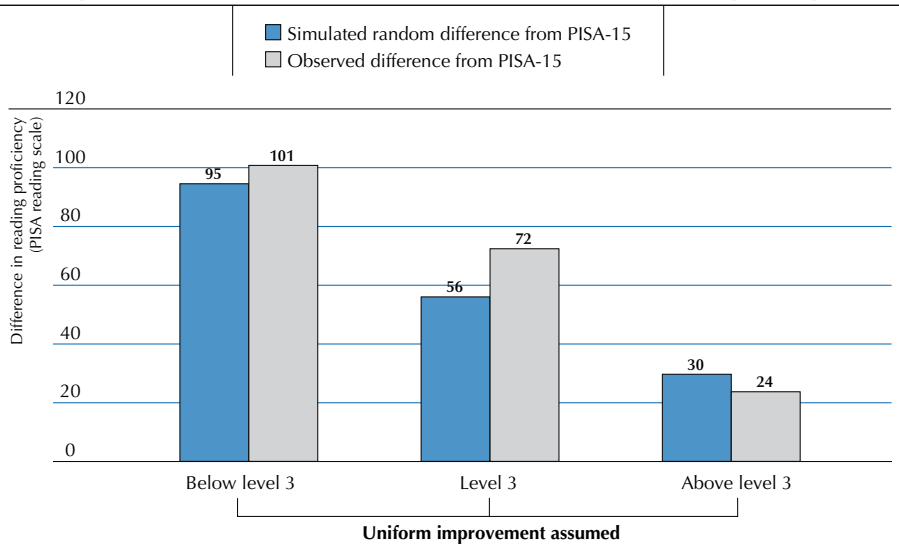


Source: Cartwright (2012).
StatLink http://dx.doi.org/10.1787/888932576871

The observed changes between PISA and PISA-24 are compared with the random simulated data from this example in Figure A.4. The simulated data are based on the specification that there is no relationship between initial status and the degree of change. Any observed relationships between initial status and degree of change are purely artefacts of regression to the mean. However, the magnitude of the group differences in the random scenario is almost identical to that in the observed data. When uniform improvement is assumed and the PISA-24 mean and standard deviation is applied, the average simulated change for each group appears to be of approximately the same magnitude as the average change observed in the actual data. Unequivocally, skills growth is occurring, but the data do not support the inference that skills acquisition is substantially greater or lesser for young people with higher or lower initial reading skill as measured by PISA-15.

■ Figure A.4 ■

Expected regression to the mean for PISA proficiency levels under uniform change assumptions compared to observed differences for Canadian PISA-15 participants



Source: Cartwright (2012).
StatLink http://dx.doi.org/10.1787/888932576871

REGRESSION TO THE MEAN AND SKILLS CONVERGENCE

When interpreting these results, one should recall that randomness does not necessarily mean "noise" – rather it represents complexity that is not described in a statistical model. The pattern of regression in the results occurs because there is real and meaningful variation in the items used in assessing reading. There are different forms of reading, and since growth in reading skills is not the same on each of these forms, the measure of growth reflects this heterogeneity. However, without a greater number of items specific to each of these sub-components of reading, it is not possible to use more detailed sub-divisions of skills and still remain comparable to PISA-15.

So, what possibilities are left? Other than to provide caveats for interpretation and analysis of the PISA-24 results, how does awareness of this phenomenon inform the current analysis? The level of skills at age 15 remains one of the most important precursors to educational and career opportunities. One of the fundamental principles of PISA-15 is that the knowledge and skills at age 15 are key determinants of other choices made during a lifetime. Thus, it is critical that any analysis of downstream outcomes, particularly in early adulthood, considers the role of skills at age 15.

To resolve the conflict between the needs of analysis and the restrictions of the data, it is important to note that the phenomenon of regression to the mean is a random event. As a random event, it should decrease as the effect of measurement error decreases. The results of a single observation typically have more measurement error than the average results of several observations. In the educational context, school grades represent the aggregation of tens or hundreds of formal and informal observations of student proficiency. In the PISA data, several other measures of initial reading proficiency are available, describing the performance of students in their language class at school. While lacking the specificity and international comparability of the PISA test results, these reported grades provide triangulation for the effects of initial status on skills growth.

The relationships of the qualitative reports of grades in school language class to growth in reading skills are shown in Figure 5.2 and Figure 5.3. The percentage of students in each category is given in Table 5.1 in Annex B. Qualitative descriptions of marks relative to a passing standard, as well as grouped percentage scores, show consistent relationships with similar interpretations as the PISA proficiency-level grouping. In all cases, lower initial status is associated with greater improvements in reading proficiency.

For subsequent analyses requiring controls for skills at age 15, this study used school marks (low, medium and high, by percentages). Although this classification may not be as internationally comparable as PISA proficiency groupings are, it has the dual benefits of higher stability at the individual level and errors that are completely independent from measurement errors in the PISA scores. As shown in Figure 5.3, it has the weakest association with changes in reading skills, suggesting the smallest influence of regression to the mean. The similarity between group sizes also facilitates meaningful group comparisons.

SUMMARY AND CONCLUSIONS

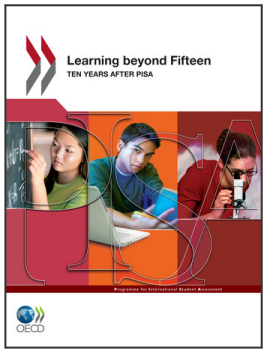
The problem of measurement error, common to all assessments, is compounded in the case of a re-assessment – because estimates of proficiency growth include the error from both assessments, and because the phenomenon of regression to the mean overestimates growth for low-proficiency students and underestimates growth for high-proficiency students. There are two implications of regression to the mean for the questions posed in this report. The first is that it challenges whether or not the skills convergence observed among young people between the ages of 15 and 24 is a real occurrence. However, using students' language grades as a measure of initial reading proficiency shows that skills convergence is in fact happening; students with less proficiency have increased their skills level faster than students who are more highly proficient, no matter which measure of initial proficiency is used.

Second, initial proficiency is strongly correlated to both proficiency growth and to many of the behaviours and choices whose impact on skills growth analysed. For instance, one would like to know if certain behaviours that were associated with high PISA scores at age 15 are also associated with improvements in proficiency. However, without controlling for initial status it will not be clear if the behaviour in question is the cause of the observed improvement in proficiency or if initial proficiency is responsible. Therefore, in determining where students' initial skills level is important, analysis will be conducted for three groups of students defined by their reading proficiency, as measured by school grades in language classes.

DATA AVAILABILITY

For approved research proposals, access to the data file is possible through one of Statistics Canada's Research Data Centre. For more information on this process please contact *educationstats@statcan.gc.ca.*

For more information on the YITS cycle 5.5: Reading Skills Reassessment, please consult: *http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SurvId=4435&SurvVer=2&InstaId=17010&InstaVer=6&SDDS=4435&lang=en&db=imdb&adm=8&dis=2.*



From:
Learning beyond Fifteen
Ten Years after PISA

Access the complete publication at:
https://doi.org/10.1787/9789264172104-en

Please cite this chapter as:

OECD (2012), "Technical Annex", in *Learning beyond Fifteen: Ten Years after PISA*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264172104-11-en