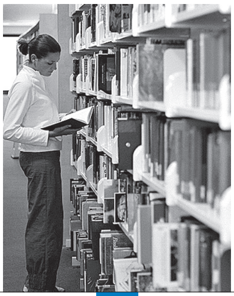


13

Trends

INTRODUCTION

Policy makers and researchers require information on how indicators change over time. An analysis of the impact of reforms on the education system, would be an example, where policy makers would seek to measure changes in the targeted area to gauge the effectiveness of their policies. In the early 1960s, for example, most OECD countries implemented education reforms to facilitate access to tertiary education, mainly through financial help. One indicator of the impact of these reforms would be to calculate the percentage of the population with a tertiary qualification for several years to show how this has evolved. Computing this trend indicator is a straightforward statistical manipulation, since the measure (*i.e.* whether or not an individual has completed tertiary education) is objective and available at the population level, in most cases. Nevertheless, such measures can be slightly biased by, for example, differing levels of immigration over a period of time, student exchange programmes, and so on.

Trends over time on a particular indicator do require careful interpretation. Policy makers also need to take into account changes to the economic context of a country, such as rising unemployment rates. Further, when comparing trend indicators across countries, it is important to consider how comparable the definition of the indicator is from country to country, *e.g.* tertiary education might mean something different in each country.

PISA offers a unique opportunity to extend the computation of trend indicators on educational outcomes by looking at student performance in reading, mathematical and scientific literacy.

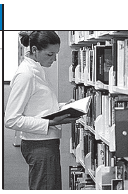
For the trend measures to be reliable, the comparability of the target population, the data collection procedures, and the assessment framework need to be consistent over time. Being able to use the results from PISA as trend indicators is one of its major aims.

Since its inception, PISA has maintained a consistent methodology of data collection. A few small methodological changes have been introduced, however: *(i)* limitation of the target population to 15-year-olds attending grade 7 or above;[1] *(ii)* modification of the student non-response adjustment for counterbalancing gender and grade differential participation rates; *(iii)* revision of the test design between 2000 and 2003.

Other changes were planned when PISA was designed: *(i)* shifts between major/minor domains; and *(ii)* revision/broadening of the assessment framework for the major domain. The changes made in the assessment frameworks have limited the use and the interpretation of the trend indicators in mathematics between 2000 and 2003 and in science between 2003 and 2006.

Figure 13.1 summarises the available trend indicators on student performance for the first three data collections. In reading literacy, the combined scale was constructed in PISA 2000 and later reading assessments were reported on this scale in PISA 2003 and PISA 2006. In PISA 2003 mathematics was the major domain, and the mathematics assessment framework was broadened from two overarching ideas included in PISA 2000 to four overarching ideas.[2] A new combined mathematic scale was constructed in PISA 2003 and two trends scales, provided in a separate database, were computed for the two overarching ideas assessed in both PISA 2000 and PISA 2003. Mathematics reporting scales are directly comparable for PISA 2003 and PISA 2006. For science, a new metric was established in PISA 2006. As mentioned in the *PISA 2006 Technical Report* (OECD, forthcoming), a science link was prepared to permit a comparison of the PISA 2006 science results with the science results in previous data collections. The science link scale provides the results for PISA 2003 and PISA 2006 using only those items that were common to the two PISA studies.

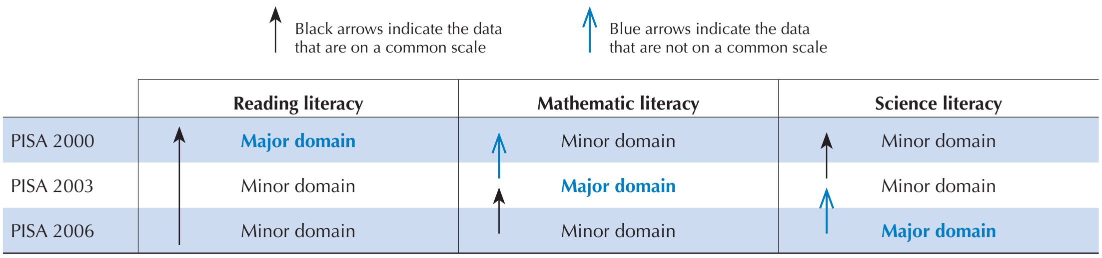
In Figure 13.1, black arrows indicate the data that are on a common scale. For instance, the plausible values for reading literacy, denoted PV1READ to PV5READ in the three international databases, are on a common scale. Trends can therefore be computed directly. Blue arrows indicate the data that are not on a common scale. For instance, the plausible values for science denoted PV1SCIE to PV5SCIE in the PISA 2003



and in the PISA 2006 databases, are not on a common scale. However, the PISA 2000 and the PISA 2003 science plausible values are on a common scale. Trends can therefore be computed in science between PISA 2000 and PISA 2003 without any precautions.

Figure 13.1

Trend indicators in PISA 2000, PISA 2003 and PISA 2006



Black arrows indicate the data that are on a common scale

Blue arrows indicate the data that are not on a common scale

	Reading literacy	Mathematic literacy	Science literacy
PISA 2000	**Major domain**	Minor domain	Minor domain
PISA 2003	Minor domain	**Major domain**	Minor domain
PISA 2006	Minor domain	Minor domain	**Major domain**

Behind these preliminary precautions, the computation of trend indicators in PISA raises two statistical issues:

1. PISA collects data on a sample and therefore any statistic has to be associated with a sampling error. The next section will discuss how to compute such sampling error on a trend indicator.
2. As implicitly represented in Figure 13.1, there are three test-design contexts for trend indicators: *(i)* the *major domain – minor domain* context; *(ii)* the *minor domain – minor domain* context; and *(iii)* the *major domain – minor domain* context. As described previously, with the last context, scales are not on the same metrics and additional data are required for the computation of trends. With the first context, *i.e. major domain – minor domain*, only a subset of items is included to ensure a psychometric link while usually the same anchor items are used in the second context, *i.e. minor domain – minor domain*. As one can easily imagine, selecting other anchor items would have returned slightly different results on the trend performance indicators. It follows that any comparison between two PISA cycles in the student performance will require an addition of another error component, *i.e.* the item sampling error.

THE COMPUTATION OF THE STANDARD ERROR FOR TREND INDICATORS ON VARIABLES OTHER THAN PERFORMANCE

For any country, the PISA samples of two cycles are independent. Therefore, the standard error on any trend indicator not involving achievement variables can be computed as follows:

$\sigma_{(\hat{\theta}_{2003}-\hat{\theta}_{2000})}=\sqrt{\sigma^2_{(\hat{\theta}_{2003})}+\sigma^2_{(\hat{\theta}_{2000})}}$, with θ representing any statistic.

However, the computation of a difference between two PISA cycles and its standard error are relevant only if the two measures are identical. For instance, in the PISA databases, there are several indices derived from the student questionnaires with exactly the same variable names (for instance, HEDRES for Home Educational Resources, BELONG for the student's sense of belonging to the school, and so on). The questions that were used to derive these indices have not changed, but as the scaling was done independently, there is no guarantee that the PISA 2000, PISA 2003 and PISA 2006 metrics are comparable. Further, these indices were standardised at the OECD level to get a mean of 0 and a standard deviation of 1. The standardisation differs between cycles. It is therefore not recommended to compute trend indicators on contextual questionnaire-derived indices.

The Highest International Social and Economic Index (denoted HISEI in the databases) satisfies all the conditions for the computation of trend indicators. Indeed, the questions were not changed and the transformation used on the International Standard Classification of Occupations (ISCO) categories has been implemented without any modification in the three cycles.

Table 13.1 presents, by country, the mean estimate of HISEI and its standard error for PISA 2000 and PISA 2003, as well as the difference between the two estimates, the standard error of this difference and the standardised difference, *i.e.* the difference divided by its standard error.

For Germany (DEU), the means for HISEI in PISA 2000 and PISA 2003 are equal to 48.85 and 49.33 respectively. The difference between these two data collections is therefore equal to:

49.33 – 48.55 = 0.48

The standard errors on these mean estimates are equal to 0.32 and 0.42. The standard error on the difference estimate is equal to:

$$\sigma_{(\hat{\theta}_{2003}-\hat{\theta}_{2000})} = \sqrt{\sigma^2_{(\hat{\theta}_{2000})} + \sigma^2_{(\hat{\theta}_{2003})}} = \sqrt{(0.32)^2 + (0.42)^2} = 0.53$$

The standardised difference, *i.e.* the difference estimate divided by its standard error, is equal to:

$$\frac{0.48}{0.53} = 0.91$$

Table 13.1
Trend indicators between PISA 2000 and PISA 2003 for HISEI, by country

	PISA 2000		PISA 2003		Difference between PISA 2003 and PISA 2000		
	Mean	**S.E.**	**Mean**	**S.E.**	**Dif.**	**S.E.**	**STD difference**
AUS	52.25	(0.50)	52.59	(0.30)	0.34	(0.58)	0.59
AUT	49.72	(0.29)	47.06	(0.52)	-2.66	(0.59)	-4.49
BEL	48.95	(0.39)	50.59	(0.38)	1.65	(0.54)	3.05
CAN	52.83	(0.22)	52.58	(0.27)	-0.25	(0.35)	-0.73
CHE	49.21	(0.53)	49.30	(0.43)	0.09	(0.68)	0.13
CZE	48.31	(0.27)	50.05	(0.34)	1.74	(0.44)	3.98
DEU	48.85	(0.32)	49.33	(0.42)	0.48	(0.53)	0.91
DNK	49.73	(0.43)	49.26	(0.45)	-0.47	(0.63)	-0.75
ESP	44.99	(0.62)	44.29	(0.58)	-0.70	(0.85)	-0.83
FIN	50.00	(0.40)	50.23	(0.36)	0.23	(0.54)	0.42
FRA	48.27	(0.44)	48.66	(0.47)	0.39	(0.64)	0.61
GBR	51.26	(0.35)	49.65	(0.39)	-1.61	(0.52)	-3.07
GRC	47.76	(0.60)	46.94	(0.72)	-0.83	(0.93)	-0.88
HUN	49.53	(0.47)	48.58	(0.33)	-0.95	(0.57)	-1.65
IRL	48.43	(0.48)	48.34	(0.49)	-0.09	(0.69)	-0.13
ISL	52.73	(0.28)	53.72	(0.26)	0.99	(0.38)	2.62
ITA	47.08	(0.31)	46.83	(0.38)	-0.24	(0.49)	-0.50
JPN	50.54	(0.62)	49.98	(0.31)	-0.56	(0.69)	-0.80
KOR	42.80	(0.42)	46.32	(0.36)	3.52	(0.55)	6.36
LUX	44.79	(0.27)	48.17	(0.22)	3.38	(0.35)	9.76
MEX	42.48	(0.68)	40.12	(0.68)	-2.37	(0.96)	-2.46
NLD	50.85	(0.47)	51.26	(0.38)	0.42	(0.61)	0.68
NOR	53.91	(0.38)	54.63	(0.39)	0.72	(0.54)	1.33
NZL	52.20	(0.37)	51.46	(0.36)	-0.74	(0.51)	-1.45
POL	46.03	(0.47)	44.96	(0.34)	-1.07	(0.58)	-1.85
PRT	43.85	(0.60)	43.10	(0.54)	-0.75	(0.81)	-0.92
SWE	50.57	(0.39)	50.64	(0.38)	0.07	(0.55)	0.12
USA	52.40	(0.79)	54.55	(0.37)	2.15	(0.87)	2.47

As the standardised difference is included in the interval [-1.96; 1.96], the difference on the mean estimate for HISEI between PISA 2000 and PISA 2003 is not statistically different from 0 with a type I error of 0.05.

Table 13.1 shows that the difference is statistically different from 0 in nine countries: Austria, Belgium, the Czech Republic, Iceland, Korea, Luxembourg, Mexico, the United Kingdom and the United States.

It would be unrealistic to consider these differences as simply a reflection of social and economic changes in these nine countries. Over a period of three years, some changes can occur, but these could not explain by themselves the size of the observed increases or decreases.

It is also possible that the quality of the samples might explain some of the differences. As the student propensity to participate positively correlate with his/her academic records and as on average low performers come from lower social background variables than high performers, an increase or a decrease in the student participation rates might affect the HISEI mean.

A change in the percentage of missing data for the HISEI variable would be another explanation that can be easily verified. On average, students who do not provide their parents' occupations are lower performers. Therefore, one should expect low socio-economic background characteristics, so that an increase of missing data could be associated with an increase of the HISEI mean, and the inverse.

In summary, changes in the school or student participation rates and in the distribution of missing data might sometimes increase the type I error, *i.e.* rejecting the null hypothesis while it is true. It is therefore recommended to implement some verification before trying to interpret calculated differences as a real change in the population characteristics.

THE COMPUTATION OF THE STANDARD ERROR FOR TREND INDICATORS ON PERFORMANCE VARIABLES

The Technical Reports of the PISA surveys (OECD, 2002; 2005; forthcoming) provide detailed information on the equating methods. These equating methods are usually based on a linear transformation. Such transformations that equate new data with previous data depend upon the change in the difficulty of each of the individual link items, and as a consequence, the sample of link items that have been chosen will influence the choice of the transformation. With an alternative set of link items, the transformation would be slightly different. The consequence is an uncertainty in the transformation due to the sampling of the link items. This uncertainty is referred to as the linking error and this error must be taken into account when making certain comparisons between the results from different PISA data collections.

Similar to the sampling error, the linking error can only be estimated. As the PISA items are clustered in units, mathematical developments for the computation of the linking error estimates are complex. The underlying rationale will therefore be presented on a fictitious example with independent items. Readers interested in the details of the linking error in PISA should consult the *PISA Technical Reports* (OECD, 2002; 2005; forthcoming).

An equating process supposes two data collections and a set of link items. For each link item, we have two item parameter estimates that are, after the application of the linear transformation, on the same metric. Some of these link items might show an increase of the relative difficulty, some might show a decrease, but on average, the difference is equal to 0. This means that some items seem more difficult in one data collection than they were in the other data collection, or the inverse.

Let $\hat{\delta}_i^1$ be the estimated difficulty of link *i* for the first data collection and let $\hat{\delta}_i^2$ be the estimated difficulty of link *i* for the second data collection, where the mean of the two sets difficulty estimates for all of the link items for a domain is set at 0. We now define the value:

$$c_i = \hat{\delta}_i^1 - \hat{\delta}_i^2$$

The value c_i is the amount by which item *i* deviates from the average of all link items in terms of the transformation that is required to align the two scales. If the link items are assumed to be a random sample of all possible link items and each of the items is counted equally, then the link error can be estimated as follows:

$$Link\_error = \sqrt{\frac{1}{L}\sum_{i=1}^{L} c_i^2}$$

where the summation is over the link items for the domain and *L* is the number of link items.

Mathematically, this formula is equal to the one used for computing the sampling error on a population mean estimate.

If the item parameters from one calibration perfectly match the item parameters from the other calibration, then the relative difficulty of the link items would not have changed. All the differences between the relative difficulties would be equal to 0 and therefore, the linking error would be equal to 0.

As the differences in the item parameters increase, the variance of these differences will increase and consequently the linking error will increase. It makes sense for the uncertainty around the trend to be proportional to the changes in the item parameters.

Also, the uncertainty around the trend indicators is inversely proportional to the number of link items. From a theoretical point of view, only one item is needed to measure a trend, but with only one item, the uncertainty will be very large. If the number of link items increases, the uncertainty will decrease.

Table 13.2 presents the linking error estimates by subject domains and by comparison between data collections.

Table 13.2
Linking error estimates

Scales	Compared data collections	Linking errors
Reading combined scale	PISA 2000 – PISA 2003	5.307
	PISA 2000 – PISA 2006	4.976
	PISA 2003 – PISA 2006	4.474
Mathematics combined scale	PISA 2003 – PISA 2006	1.382
Interim science scale	PISA 2000 – PISA 2003	3.112
Science scale	PISA 2003 – PISA 2006	4.963

A common transformation has been estimated from the link items, and this transformation is applied to all participating countries. It follows that any uncertainty that is introduced through the linking is common to all students and all countries. Thus, for example, suppose that the linking error between PISA 2000 and PISA 2003 in reading resulted in an overestimation of student scores by two points on the PISA 2000 scale. It follows that every student's score would be overestimated by two score points. This overestimation will have effects on certain, but not all, summary statistics computed from the PISA 2003 data. For example, consider the following:

- Each country's mean would be overestimated by an amount equal to the link error, in our example this is two score points;
- The mean performance of any subgroup would be overestimated by an amount equal to the linking error, in our example this is two score points;
- The standard deviation of student scores would not be affected because the over-estimation of each student by a common error does not change the standard deviation;
- The difference between the mean scores of two countries in PISA 2003 would not be influenced because the over-estimation of each student by a common error would have distorted each country's mean by the same amount;
- The difference between the mean scores of two groups (*e.g.* males and females) in PISA 2003 would not be influenced, because the overestimation of each student by a common error would have distorted each group's mean by the same amount;
- The difference between the performance of a group of students (*e.g.* a country) between PISA 2000 and PISA 2003 would be influenced because each student's score in PISA 2003 would be influenced by the error; and
- A change in the difference between two groups from PISA 2000 to PISA 2003 would not be influenced. This is because neither of the components of this comparison, which are differences in scores in PISA 2000 and PISA 2003 respectively, is influenced by a common error that is added to all student scores in PISA 2003.

In general terms, the linking error need only be considered when comparisons are being made between results from different data collections, and then usually when group means are being compared.

The most obvious example of a situation where there is a need to use the linking error is in the comparison of the mean performance for a country between two data collections.

In PISA 2000, the mean in reading literacy for Germany is equal to 483.99 with a standard error of 2.47. In PISA 2003, the mean for Germany is equal to 491.36 and the standard error is equal to 3.39. The difference between PISA 2003 and PISA 2000 is therefore equal to 491.36-483.99=7.37. The average performance of German students has therefore increased by 7.37 scores on the PISA 2000 reading scale from PISA 2000 to PISA 2003.

The standard error on this difference, as mentioned previously, is influenced by the linking error. The standard error is therefore equal to:

$$SE = \sqrt{\sigma^2_{(\hat{\mu}_{2000})} + \sigma^2_{(\hat{\mu}_{2003})} + \sigma^2_{(linking\_error)}}$$

$$SE = \sqrt{(2.47)^2 + (3.39)^2 + (5.31)^2} = 6.77$$

As the standardised difference between PISA 2000 and PISA 2003, *i.e.* 7.37/6.77, is included in the interval [-1.96; 1.96], the null hypothesis of no difference is not rejected. In other words, Germany's performance in reading has not changed between PISA 2000 and PISA 2003.

Table 13.3 provides the estimates of the reading performance in Germany by gender in PISA 2000 and PISA 2003, with their respective standard errors, as well as the difference in the mean performance and their respective standard errors.

Table 13.3
Mean performance in reading by gender in Germany

		Performance in reading	S.E.
PISA 2003	Females	512.93	3.91
	Males	470.80	4.23
	Difference	42.13	4.62
PISA 2000	Females	502.20	3.87
	Males	467.55	3.17
	Difference	34.65	5.21

As the comparison for a particular country between PISA 2000 and PISA 2003 is affected by the linking error, the comparison for a particular subgroup between PISA 2000 and PISA 2003 is also affected by the linking error. Therefore, the standard error has to include the linking error.

The trend indicators for males and females in Germany are, respectively, equal to:

$$Trends_{females} = 512.93 - 502.20 = 10.73$$

$$SE_{females} = \sqrt{(3.91)^2 + (3.87)^2 + (5.31)^2} = 7.65$$

$$Trends_{males} = 470.80 - 467.55 = 3.25$$

$$SE_{males} = \sqrt{(4.23)^2 + (3.17)^2 + (5.31)^2} = 7.49$$

Both differences are not statistically different from 0.

On the other hand, the gender difference in PISA 2003 is not affected by the linking error. Indeed, both subgroup estimates will be underestimated or overestimated by the same amount and therefore the computation of the difference will neutralise this difference. Consequently, the trend indicator on the gender difference and its standard error will be equal to:

$$Trends_{Gender\_dif} = 42.13 - 34.65 = 7.43$$

$$SE_{Gender\_dif} = \sqrt{(4.62)^2 + (5.21)^2} = 6.96$$

This means that the change in gender difference in Germany for reading between PISA 2000 and PISA 2003 was not statistically significant, even though it appears from Table 13.3 to have widened considerably.

In the PISA initial reports, student performance is also reported by proficiency levels (see Chapter 9). As the linking error affects the country mean estimates, the percentages of students at each level will also be affected. However, an overestimation or an underestimation of the results of *X* points on one PISA scale will have a different impact on the percentages of students at each proficiency level for each country. If the percentage is small, then the impact will be small. If the percentage is large, then the impact will be larger. It would have been too complex to provide a linking error for each country and for each proficiency level. It was therefore decided not to take into account the linking error for the comparison of percentages of students at each proficiency level between two PISA data collections. This means that the standard errors on the difference between 2000 and 2003 are underestimated.

CONCLUSION

This chapter was devoted to the computation of the standard error on trend indicators. The comparison of any variable other than performance variables is straightforward as the national PISA samples for two cycles are independent. However, such comparisons are only relevant if the measures are comparable from one cycle to another.

The comparison of performance mean estimates is more complex as it might require the inclusion of the linking error in the standard error depending on the statistic. For instance, Table 2.1.d in the PISA 2003 initial report (OECD, 2004) presents the trends in mathematics/space and shape average performance between PISA 2000 and PISA 2003. The trend indicator has integrated the linking error in its standard error. However, Figure 2.6c in the PISA 2003 initial report (OECD, 2004) presents the trends between PISA 2000 and PISA 2003 on the 5th, 10th, 25th, 75th, 90th and 95th percentiles. As mentioned previously, it would require a specific linking error for each percentile and for each country. For that reason, the linking error was not integrated in the standard error of these trends.

Due to the growing interest in trend indicators and their political impacts, it is essential to interpret significant changes with caution. A significant change might simply be due to a difference in the school or student participation rate, in the pattern of missing data or in the composition of the test. For instance, changing the repartition of item types (multiple choice versus open-ended items) might have an impact on the gender difference estimates.

Notes

1. This was introduced from PISA 2003. In PISA 2000, only a very small percentage of 15-year-olds were attending grade 5 or grade 6 (Austria: 0.03%; Canada: 0.03%; Czech Republic: 0.06%; Germany: 0.02%; Hungary: 0.59%; Latvia: 0.27%; Portugal: 1.25%; and Russia: 0.04%). Therefore, except for Portugal, the change in the target population should not significantly affect trend indicators.

2. Four overarching ideas consist of *space and shape*; *change and relationships*; *quantity*; and *uncertainty*. *Space* and *shape* and *change and relationships* were covered in PISA 2000.

References

Beaton, A.E. (1987), *The NAEP 1983-1984 Technical Report*, Educational Testing Service, Princeton.

Beaton, A.E., *et al.* (1996), *Mathematics Achievement in the Middle School Years, IEA's Third International Mathematics and Science Study*, Boston College, Chestnut Hill, MA.

Bloom, B.S. (1979), *Caractéristiques individuelles et apprentissage scolaire*, Éditions Labor, Brussels.

Bressoux, P. (2008), *Modélisation statistique appliquée aux sciences sociales*, De Boek, Brussels.

Bryk, A.S. and **S.W. Raudenbush** (1992), *Hierarchical Linear Models for Social and Behavioural Research: Applications and Data Analysis Methods*, Sage Publications, Newbury Park, CA.

Buchmann, C. (2000), *Family structure, parental perceptions and child labor in Kenya: What factors determine who is enrolled in school?* a*Soc. Forces,* No. 78, pp. 1349-79.

Cochran, W.G. (1977), *Sampling Techniques*, J. Wiley and Sons, Inc., New York.

Dunn, O.J. (1961), "Multilple Comparisons among Menas", *Journal of the American Statistical Association*, Vol. 56, American Statistical Association, Alexandria, pp. 52-64.

Kish, L. (1995), *Survey Sampling*, J. Wiley and Sons, Inc., New York.

Knighton, T. and **P. Bussière** (2006), "Educational Outcomes at Age 19 Associated with Reading Ability at Age 15", Statistics Canada, Ottawa.

Gonzalez, E. and **A. Kennedy** (2003), *PIRLS 2001 User Guide for the International Database*, Boston College, Chestnut Hill, MA.

Ganzeboom, H.B.G., P.M. De Graaf and **D.J. Treiman** (1992), "A Standard International Socio-economic Index of Occupation Status", *Social Science Research* 21(1), Elsevier Ltd, pp 1-56.

Goldstein, H. (1995), *Multilevel Statistical Models,* 2nd Edition, Edward Arnold, London.

Goldstein, H. (1997), "Methods in School Effectiveness Research", *School Effectiveness and School Improvement* 8, Swets and Zeitlinger, Lisse, Netherlands, pp. 369-395.

Hubin, J.P. (ed.) (2007), *Les indicateurs de l'enseignement*, 2nd Edition, Ministère de la Communauté française, Brussels.

Husen, T. (1967), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries,* Almqvist and Wiksells, Uppsala.

International Labour Organisation (ILO) (1990), *International Standard Classification of Occupations: ISCO-88.* Geneva: International Labour Office.

Lafontaine, D. and **C. Monseur** (forthcoming), "Impact of Test Characteristics on Gender Equity Indicators in the Assessment of Reading Comprehension", *European Educational Research Journal*, Special Issue on PISA and Gender.

Lietz, P. (2006), "A Meta-Analysis of Gender Differences in Reading Achievement at the Secondary Level", *Studies in Educational Evaluation* 32, pp. 317-344.

Monseur, C. and **M. Crahay** (forthcoming), "Composition académique et sociale des établissements, efficacité et inégalités scolaires : une comparaison internationale – Analyse secondaire des données PISA 2006", *Revue française de pédagogie*.

OECD (1998), *Education at a Glance – OECD Indicators,* OECD, Paris.

OECD (1999a), *Measuring Student Knowledge and Skills – A New Framework for Assessment,* OECD, Paris.

OECD (1999b), *Classifying Educational Programmes – Manual for ISCED-97 Implementation in OECD Countries,* OECD, Paris.

OECD (2001), *Knowledge and Skills for Life – First Results from PISA 2000,* OECD, Paris.

OECD (2002a), *Programme for International Student Assessment – Manual for the PISA 2000 Database,* OECD, Paris.

OECD (2002b), *Sample Tasks from the PISA 2000 Assessment – Reading, Mathematical and Scientific Literacy,* OECD, Paris.

OECD (2002c), *Programme for International Student Assessment – PISA 2000 Technical Report,* OECD, Paris.

OECD (2002d), *Reading for Change: Performance and Engagement across Countries – Results from PISA 2000,* OECD, Paris.

OECD (2003a), *Literacy Skills for the World of Tomorrow – Further Results from PISA 2000,* OECD, Paris.

OECD (2003b), *The PISA 2003 Assessment Framework – Mathematics, Reading, Science and Problem Solving Knowledge and Skills,* OECD, Paris.

OECD (2004a), *Learning for Tomorrow's World – First Results from PISA 2003,* OECD, Paris.

OECD (2004b), *Problem Solving for Tomorrow's World – First Measures of Cross-Curricular Competencies from PISA 2003,* OECD, Paris.

OECD (2005a), *PISA 2003 Technical Report,* OECD, Paris.

OECD (2005b), *PISA 2003 Data Analysis Manual,* OECD, Paris.

OECD (2006), *Assessing Scientific, Reading and Mathematical Literacy: A Framework for PISA 2006,* OECD, Paris.

OECD (2007), *PISA 2006: Science Competencies for Tomorrow's World*, OECD, Paris.

OECD (2009), *PISA 2006 Technical Report,* OECD, Paris.

Peaker, G.F. (1975), *An Empirical Study of Education in Twenty-One Countries: A Technical report. International Studies in Evaluation VIII,* Wiley, New York and Almqvist and Wiksell, Stockholm.

Rust, K.F. and **J.N.K. Rao** (1996), "Variance Estimation for Complex Surveys Using Replication Techniques", *Statistical Methods in Medical Research,* Vol. 5, Hodder Arnold, London, pp. 283-310.

Rutter, M., *et al.* (2004), "Gender Differences in Reading Difficulties: Findings from Four Epidemiology Studies", *Journal of the American Medical Association* 291, pp. 2007-2012.

Schulz, W. (2006), *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003,* Paper presented at the Annual Meetings of the American Educational Research Association (AERA) in San Francisco, 7-11 April.

Wagemaker, H. (1996), *Are Girls Better Readers. Gender Differences in Reading Literacy in 32 Countries*, IEA, The Hague.

Warm, T.A. (1989), "Weighted Likelihood Estimation of Ability in Item Response Theory", *Psychometrika,* Vol. 54(3), Psychometric Society, Williamsburg, VA., pp. 427-450.

Wright, B.D. and **M.H. Stone** (1979), *Best Test Design: Rasch Measurement,* MESA Press, Chicago.

Table of contents

LIST OF BOXES

LIST OF FIGURES

LIST OF TABLES

User's Guide

Preparation of data files

All data files (in text format) and the SPSS® control files are available on the PISA website (*www.pisa.oecd.org*).

SPSS® users

By running the SPSS® control files, the PISA data files are created in the SPSS® format. Before starting analysis in the following chapters, save the PISA 2000 data files in the folder of "c:\pisa2000\data\", the PISA 2003 data files in "c:\pisa2003\data\", and the PISA 2006 data files in "c:\pisa2006\data\".

SPSS® syntax and macros

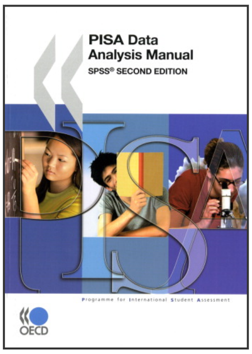
All syntaxes and macros in this manual can be copied from the PISA website (*www.pisa.oecd.org*). These macros were developed for SPSS 17.0. The 19 SPSS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before staring analysis. Each chapter of the manual contains a complete set of syntaxes, which must be done sequentially, for all of them to run correctly, within the chapter.

Rounding of figures

In the tables and formulas, figures were rounded to a convenient number of decimal places, although calculations were always made with the full number of decimal places.

Country abbreviations used in this manual

AUS	Australia	FRA	France	MEX	Mexico
AUT	Austria	GBR	United Kingdom	NLD	Netherlands
BEL	Belgium	GRC	Greece	NOR	Norway
CAN	Canada	HUN	Hungary	NZL	New Zealand
CHE	Switzerland	IRL	Ireland	POL	Poland
CZE	Czech Republic	ISL	Iceland	PRT	Portugal
DEU	Germany	ITA	Italy	SVK	Slovak Republic
DNK	Denmark	JPN	Japan	SWE	Sweden
ESP	Spain	KOR	Korea	TUR	Turkey
FIN	Finland	LUX	Luxembourg	USA	United States



From:

PISA Data Analysis Manual: SPSS, Second Edition

Access the complete publication at:

https://doi.org/10.1787/9789264056275-en

Please cite this chapter as:

OECD (2009), “Trends”, in *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264056275-14-en