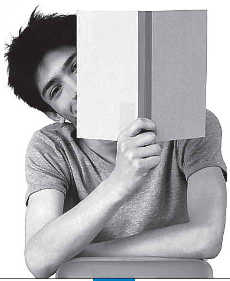


6

Plausible Values

INDIVIDUAL ESTIMATES VERSUS POPULATION ESTIMATES

Education assessments can have two major purposes:

1. To measure the knowledge and skills of particular students. The performance of each student usually will have an impact on his or her future (school career, admission to post-secondary education, and so on). It is therefore particularly important to minimise the measurement error associated with each individual's estimate.
2. To assess the knowledge or skills of a population. The performance of individuals will have no impact on their school career or professional life. In such a case, the goal of reducing error in making inferences about the target population is more important than the goal of reducing errors at the individual level.

National and international education surveys belong to the second category.

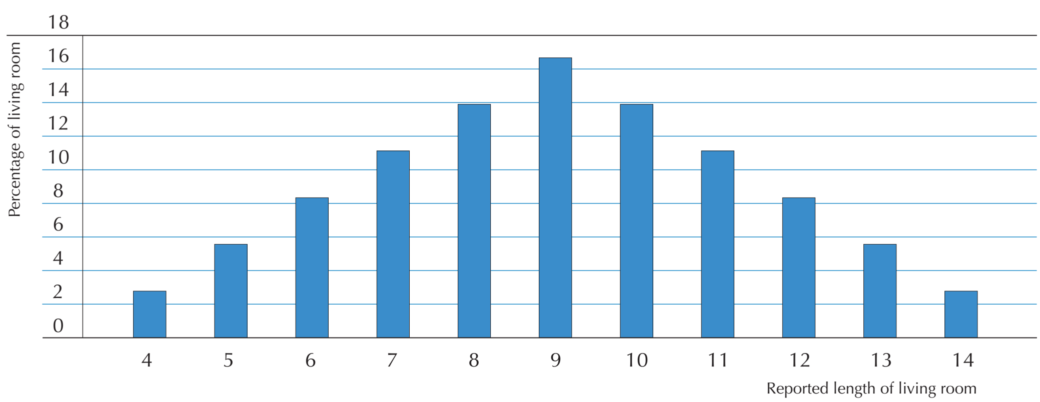
International surveys such as PISA report student performance through plausible values.[1] This chapter will explain the conceptual meaning of plausible values and the advantage of reporting with them. Individual estimators (such as the *weighted likelihood estimate* [WLE] defined in Chapter 5) will be compared with plausible values for the purposes of estimating a range of population statistics.

THE MEANING OF PLAUSIBLE VALUES (PVs)

An example taken from the physical sciences measurement area can help illustrate the complex concept of plausible values. Suppose that a city board decides to levy a new building tax to increase the city's revenue. This new tax will be proportional to the length of the family house living room. Inspectors visit all city houses to measure the length of the living rooms. They are given a measuring tape and are instructed to record the length in term of integers only, *i.e.* 1 metre, 2 metres, 3 metres, 4 metres and so on.

The results of this measure are shown in Figure 6.1. About 2% of the living rooms have a reported length of 4 metres; slightly over 16% of the living rooms have a reported length of 9 metres and so on.

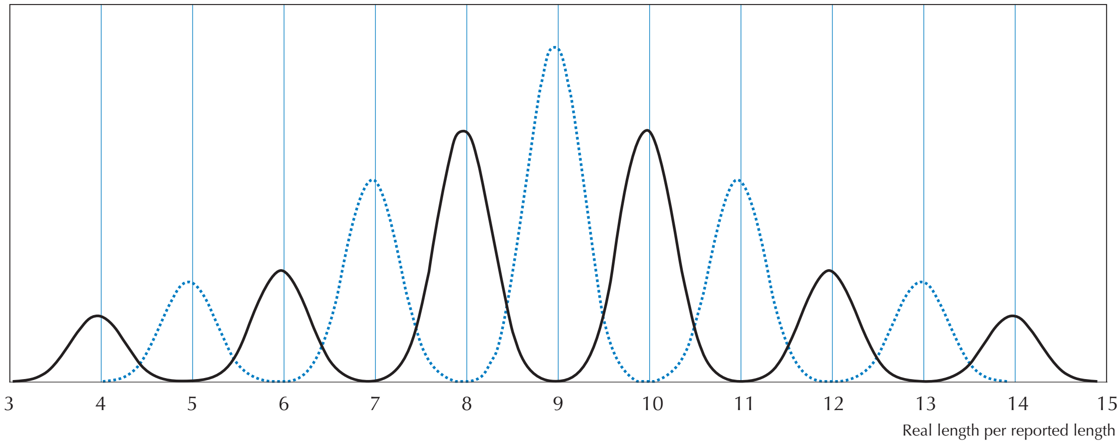
Figure 6.1
Living room length expressed in integers



Of course, the reality is quite different, as length is a continuous variable. With a continuous variable, observations can take any value between the minimum and the maximum. On the other hand, with a discontinuous variable, observations can only take a predefined number of values. Figure 6.2 gives the length distribution of the living rooms per reported length.

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 – © OECD 2009

Figure 6.2
Real length per reported length



All living rooms with a reported length of 5 metres are not exactly 5 metres long. On average, they are 5 metres long, but their length varies around the mean. The difference between reported length and real length is due to the rounding process and measurement error. An inspector might incorrectly report 5 metres for a particular living room, when it really measures 4.15 metres. If the rounding process were the only source of error, then the reported length should be 4 metres. The second source of error, the error in measuring, explains the overlap in the distribution.

In this particular example, the lengths of the living rooms are normally distributed around the mean, which is also the reported length. If the difference between the length and the closest integer is small, then the probability of not reporting this length with the closest integer is very small. For instance, it is unlikely that a length of 4.15 be reported as 5 metres or 3 metres. However, as the distance between the real length and the closest integer increases, the probability of not reporting this length with the closest integer will also increase. For instance, it is likely that a length of 4.95 will be reported as 5 metres, whereas a length of 4.50 will be reported equally as many times as 4 metres as it is 5 metres.

The methodology of plausible values consists of:

- mathematically computing distributions (denoted as posterior distributions) around the reported values and the reported length in the example; and
- assigning to each observation a set of random values drawn from the posterior distributions.

Plausible values can therefore be defined as random values from the posterior distributions. In the example, a living room of 7.154 metres that was reported as 7 metres might be assigned any value from the normal distribution around the reported length of 7. It might be 7.45 as well as 6.55 or 6.95. Therefore, plausible values should not be used for individual estimation.

This fictitious example from the physical sciences can be translated successfully to the social sciences. For example, with a test of six dichotomous items, a continuous variable (*i.e.* mental ability) can be transformed into a discontinuous variable. The discontinuous variable will be the student raw score or the number of correct answers. The only possible scores are: 0, 1, 2, 3, 4, 5 and 6.

Contrary to most measures in the physical sciences, psychological or education measures encompass substantial measurement errors because:

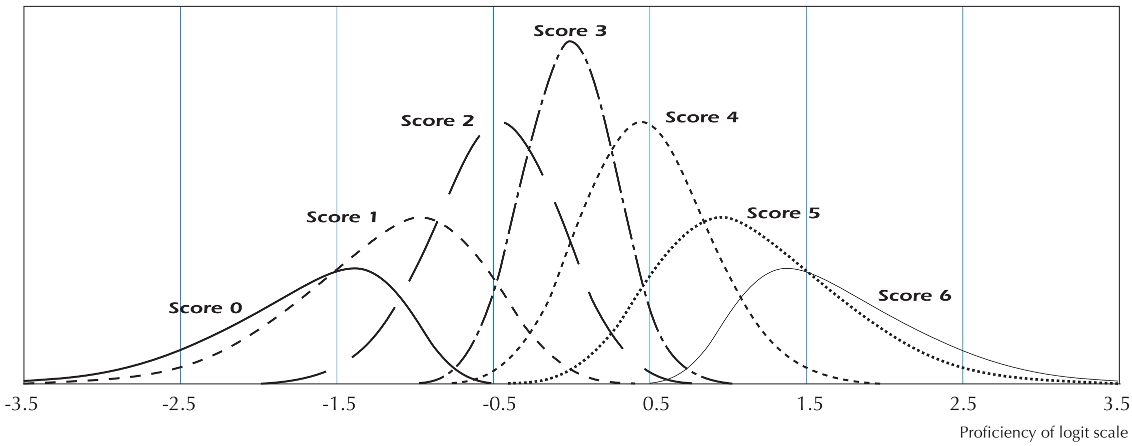
- The concept to be measured is broader.
- They might be affected by the mental and physical dispositions of the students on the day of the assessment.
- The conditions in which students are tested might also affect the results.

This means that there are large overlaps in the posterior distributions, as shown in Figure 6.3.

Further, with the example of the living room, the measurement error of the posterior distributions can be considered as independent of the living room.[2] In education, the measurement error is not always independent of the proficiency level of the students. It may be smaller for average students, and larger for low and high achievers, depending on the test average difficulty.

Further, in this particular example, the posterior distributions for score 0 and score 6 are substantially skewed, as the posterior distributions of the living rooms with a reported length of 4 and 14 metres would be, if all living rooms smaller than 4 metres were reported as 4, and if all living rooms longer than 14 metres were reported as 14. This means that the posterior distributions are not normally distributed, as shown in Figure 6.3.

Figure 6.3
A posterior distribution on a test of six items



Generating plausible values on an education test consists of drawing random numbers from the posterior distributions. This example clearly shows that plausible values should not be used for individual performance. Indeed, a student who scores 0 might get -3, but also -1. A student who scores 6 might get 3, but also 1.

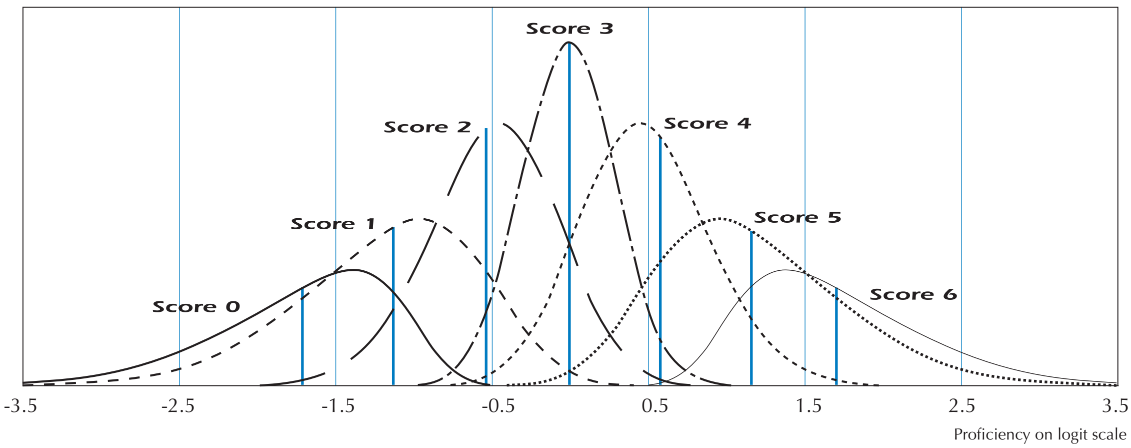
It has been noted that:

> "The simplest way to describe plausible values is to say that plausible values are a representation of the range of abilities that a student might reasonably have. […] Instead of directly estimating a student's ability θ, a probability distribution for a student's θ is estimated. That is, instead of obtaining a point estimate for θ, like a WLE, a range of possible values for a student's θ, with an associated probability for each of these values is estimated. Plausible values are random draws from this (estimated) distribution for a student's θ" (Wu and Adams, 2002).[3]

All this methodology aims at building a continuum from a collection of discontinuous variables (*i.e.* the test score). It is meant to prevent biased inferences occurring as a result of measuring an unobservable underlying ability through a test using a relatively small number of items.

Finally, an individual estimate of student ability can also be derived from the posterior distributions. This derived individual estimate is called the *expected a posteriori* estimator (EAP). Instead of assigning a set of random values from the posterior distributions, the mean of the posterior distributions is assigned. Therefore, the EAP can be considered as the mean of an infinite set of plausible values for a particular student. See Figure 6.4.

Figure 6.4
EAP estimators



As only one value is assigned per posterior distribution, the EAP estimator is also a discontinuous variable.[4] However, EAP estimates and WLEs differ as the former requires a population distribution assumption, which is not the case for the latter. Further, while any raw score for a particular test will always be associated with one and only one WLE, different EAP values can be associated with a particular raw score, depending on the regressors used as conditioning variables.

Researchers not used to working with plausible values might consider this apparent randomisation as a source of imprecision. The comparison of the different types of Rasch ability estimators (WLE, EAP and PV) through the estimation of population statistics will overcome this perception. The PISA database only includes PVs[5] for student performance but uses WLE for contextual indices derived from the student or school questionnaire. Although PISA does not include any EAP in its databases, the comparison will incorporate EAP estimates to show biases that occur when data analysts average the plausible values at the student levels to obtain one score value per student.

COMPARISON OF THE EFFICIENCY OF WLEs, EAP ESTIMATES AND PVs FOR THE ESTIMATION OF SOME POPULATION STATISTICS[6]

A comparison between different student ability estimators could be performed on real data. Such a comparison will outline differences, but it will not identify the best estimators for a particular population statistic. Therefore, a simulation will be used to illustrate and evaluate the differences in various estimators.

The simulation consists of three major steps:

- The generation of a dataset including a continuous variable that represents the student abilities (*i.e.* denoted as the latent variable), some background variables including the gender and an index of social background, denoted HISEI, and a pattern of item responses coded 0 for an incorrect answer and 1 for a correct answer. The results presented hereafter are based on a fictitious test of 15 items.[7]
- The computation of the student ability estimator, in particular the WLEs, EAP estimates and PVs.[8]
- The estimation of some population parameters using the student ability (*i.e.* latent variable) and the different student ability estimators. A comparison will be made for:
  – mean, variance and percentiles,
  – correlation,
  – between- and within-school variance.

The dataset of this simulation contains 5 250 students distributed in 150 schools with 35 students per school. Table 6.1 presents the structure of the simulated dataset before the importation of the Rasch student ability estimators.

Table 6.1
Structure of the simulated data

School ID	Student ID	Sex	HISEI	Item 1	Item 2	…	Item 14	Item 15
001	01	1	32	1	1		0	0
001	02	0	45	1	0		1	0
…	…							
150	5 249	0	62	0	0		1	1
150	5 250	1	50	0	1		1	1

Table 6.2 presents the mean and the variance of the latent variable, the WLEs, the EAP estimates and the five plausible values. The average of the 5 PV mean is also included.

Table 6.2 shows that a good estimate of the population's mean (*i.e.* the latent variable estimate) is obtained regardless of the type of latent variable used (WLEs, EAP estimates or PVs). It can be empirically demonstrated that none of the estimates significantly differs from the expected mean, *i.e.* 0.00 in this particular case (Wu and Adams, 2002). Additionally, it can also be shown that the mean of the WLEs will not be biased if the test is well targeted, *i.e.* if the average of the item difficulties is around 0 on the Rasch scale (Wu and Adams, 2002). That is, on a well-targeted test, students will obtain a raw score of about 50% correct answers. If the test is too easy then the mean of the WLEs will be underestimated (this is called the ceiling effect), while if it is too difficult then the mean of the WLEs will be overestimated (this is called the floor effect).

Table 6.2
Means and variances for the latent variables and the different student ability estimators

	Mean	Variance
Latent variable	0.00	1.00
WLE	0.00	1.40
EAP	0.00	0.75
PV1	0.01	0.99
PV2	0.00	0.99
PV3	0.00	1.01
PV4	0.00	1.01
PV5	-0.01	0.00
Average of the 5 PV statistics	0.00	1.00

These last results explain why the mean of the WLEs provided in the PISA 2000 database differs from the mean of the plausible values, especially for the partner countries. For the reading/reflection and evaluation scale, the means obtained for Canada using WLEs and PVs are 538.4 and 542.5, respectively, which are very close. In contrast, the means obtained for Peru, using WLEs and PVs are 352.2 and 322.7, respectively, a difference of about 0.3 standard deviations. This shows that there is bias when WLEs are used to estimate the mean if the test is not well targeted.

For the population variance, Table 6.2 shows that PVs give estimates closest to the expected value, while WLEs overestimate it and the EAP underestimates it. These results are consistent with other simulation studies.

Table 6.3 presents some percentiles computed on the different ability estimators. For example, because the variance computed using plausible values is not biased, the percentiles based on PVs are also unbiased. However, because the EAP estimates and WLEs variances are biased, the percentiles, and in particular, extreme percentiles will also be biased. These results are consistent with other simulation studies previously cited.

Table 6.4 presents the correlation between the social background index (HISEI), gender and the latent variables and the different estimators of student abilities. The correlation coefficients with the WLEs are both underestimated, while the correlation coefficients with the EAP estimates are overestimated. Only the correlation coefficients with the plausible values are unbiased.[9]

It should be noted that the regression coefficients are all unbiased for the different types of estimators. Nevertheless, as variances are biased for some estimators, residual variances will also be biased. Therefore, the standard error on the regression coefficients will be biased in the case of the WLEs and the EAP estimates.

Table 6.3
Percentiles for the latent variables and the different student ability estimators

	P5	P10	P25	P50	P75	P90	P95
Latent variable	-1.61	-1.26	-0.66	0.01	0.65	1.26	1.59
WLE	-2.15	-1.65	-0.82	-0.1	0.61	1.38	1.81
EAP	-1.48	-1.14	-0.62	-0.02	0.55	1.08	1.37
PV1	-1.68	-1.29	-0.71	-0.03	0.64	1.22	1.59
PV2	-1.67	-1.31	-0.69	-0.03	0.62	1.22	1.58
PV3	-1.67	-1.32	-0.70	-0.02	0.64	1.21	1.56
PV4	-1.69	-1.32	-0.69	-0.03	0.63	1.23	1.55
PV5	-1.65	-1.3	-0.71	-0.02	0.62	1.2	1.55
Average of the 5 PV statistics	-1.67	-1.31	-0.70	-0.03	0.63	1.22	1.57

Table 6.4
Correlation between HISEI, gender and the latent variable, the different student ability estimators

	HISEI	GENDER
Latent variable	0.40	0.16
WLE	0.33	0.13
EAP	0.46	0.17
PV1	0.41	0.15
PV2	0.42	0.15
PV3	0.42	0.13
PV4	0.40	0.15
PV5	0.40	0.14
Average of the 5 PV statistics	0.41	0.14

Table 6.5
Between- and within-school variances

	Between-school variance	Within-school variance
Latent variable	0.33	0.62
WLE	0.34	1.02
EAP	0.35	0.38
PV1	0.35	0.61
PV2	0.36	0.60
PV3	0.36	0.61
PV4	0.35	0.61
PV5	0.35	0.61
Average of the 5 PV statistics	0.35	0.61

Finally, Table 6.5 presents the between- and within-school variances. Between-school variances for the different estimators do not differ from the expected value of 0.33. However, WLEs overestimate the within-school variance, while the EAP estimates underestimate it. These results are consistent with other simulation studies (Monseur and Adams, 2002).

As this example shows, plausible values provide unbiased estimates.

HOW TO PERFORM ANALYSES WITH PLAUSIBLE VALUES

As stated in the previous section, a set of plausible values, usually five, are drawn for each student for each scale or subscale. Population statistics should be estimated using each plausible value separately. The reported population statistic is then the average of each plausible value statistic. For instance, if one is interested in the correlation coefficient between the social index and the reading performance in PISA, then five correlation coefficients should be computed and then averaged.

Plausible values should never be averaged at the student level, *i.e.* by computing in the dataset the mean of the five plausible values at the student level and then computing the statistic of interest once using that average PV value. Doing so would be equivalent to an EAP estimate, with a bias as described in the previous section.

Mathematically, secondary analyses with plausible values can be described as follows. If θ is the population statistic and θ_i is the statistic of interest computed on one plausible value, then:

$$\theta = \frac{1}{M}\sum_{i=1}^{M}\theta_i$$

where M is the number of plausible values.

Plausible values also allow computing the uncertainty in the estimate of θ due to the lack of precision in the measurement test. If a perfect test could be developed, then the measurement error would be equal to zero and the five statistics from the plausible values would be exactly identical. Unfortunately, perfect tests do not exist and never will. This measurement variance, usually denoted imputation variance, is equal to:

$$B_M = \frac{1}{M-1}\sum_{i=1}^{M}(\theta_i - \theta)^2$$

It corresponds to the variance of the five plausible value statistics of interest. The final stage is to combine the sampling variance and the imputation variance as follows:

$$V = U + \left(1 + \frac{1}{M}\right)B_m$$

where U is the sampling variance.

Chapter 7 will demonstrate how to compute sampling variances and imputation variances and how to combine them, using the PISA databases.

CONCLUSION

This chapter was devoted to the meaning of plausible values and the steps required to analyse data with PVs. A comparison between PVs and alternate individual ability estimates was presented to demonstrate the superiority of this methodology for reporting population estimates.

Notes

1. The methodology of plausible values was first implemented in the National Assessment of Educational Progress (NAEP) studies. For more information, see Beaton (1987).

2. The measurement error will be independent of the length of the living rooms if the inspectors are using a measuring instrument that is at least 15 metres long (such as a measuring tape). If they are using a standard metre, then the overall measurement error will be proportional to the length of the living room.

3. The probability distribution for a student's θ can be based on the cognitive data only, *i.e.* the item response pattern, but can also include additional information, such as student gender, social background, and so on. The probability distribution becomes therefore conditioned by this additional information. A mathematical explanation of the model used for the scaling of the PISA 2000 scaling can be found in the *PISA 2000 Technical Report* (OECD, 2002c).

4. If several regressors are used as conditioning variables, then the EAP estimator tends to be a continuous variable.

5. PISA 2000 data files include both WLEs and PVs.

6. PV and EAP estimators can be computed with or without regressors. As the PVs in PISA were generated based on all variables collected through the student questionnaires, this comparison will only include PVs and EAP estimators with the use of regressors.

7. The data generation starts with a factorial analysis on a 3 by 3 squared correlation matrix. The correlation between the latent variable and gender was set at 0.20, the correlation between the latent variable and the social background indicator was set at 0.40 and the correlation between gender and the social background indicator was set at 0.00. Three random variables are drawn from normal distributions and combined according to the factorial regression coefficients to create the three variables of interest, *i.e.* reading, gender and social background. Based on the student score on the latent variable and a predefined set of 20 item difficulties; probabilities of success are computed according to the Rasch Model. These probabilities are then compared to uniform distribution and recoded into 0 and 1. Finally, gender is recoded into a dichotomous variable.

8. The estimators were computed with the Conquest Software® developed by M.L. Wu, R.J. Adams and M.R. Wilson.

9. The results on the EAP and PV correlation coefficients are observed when the probability distributions are generated with conditioning variables. Without the conditioning, the correlation with the plausible values would be underestimated.

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

PISA DATA ANALYSIS MANUAL: SPSS® SECOND EDITION – ISBN 978-92-64-05626-8 –

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

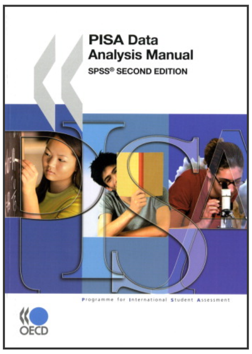
All syntaxes and macros in this manual can be copied from the PISA website (*www.pisa.oecd.org*). These macros were developed for SPSS 17.0. The 19 SPSS® macros presented in Chapter 17 need to be saved under "c:\pisa\macro\", before staring analysis. Each chapter of the manual contains a complete set 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), “Plausible Values”, in *PISA Data Analysis Manual: SPSS, Second Edition*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264056275-7-en