7

Monitoring the Quality of PISA

It is essential that users of the PISA data have confidence that the data collection activities have been undertaken to a high standard. The quality assurance that provides this confidence consists of two methods. The first is to carefully develop and document procedures that will result in data of the desired quality, the second is to monitor and record the implementation of the documented procedures. Should it happen that the documented processes are not fully implemented, it is necessary to understand to what extent they were not, and the likely implications for the data.

Quality monitoring is, therefore, the process of systematically observing and recording the extent to which data are collected, retrieved, and stored according to the procedures described in the field operations manuals. Quality monitoring is a continuous process that identifies potential issues and allows forestalment of operational problems. The responsibility for quality control resides with the National Project Managers (NPMs) while quality monitoring is a collaborative process between the NPM and the consortium that assists the NPM.

A comprehensive program of continuous quality monitoring was central to ensuring full, valid implementation of the PISA 2003 procedures and the recording of deviation from those procedures. The main elements of the quality monitoring procedures were:

- *Consortium experts* – To assist NPMs in the planning and implementation of key processes, consortium experts systematically monitored the key processes of school and student sampling, translation and preparation of instruments, coding of responses, field operations, and data preparation.
- *National centre quality monitors* (NCQMs) – To observe the implementation of PISA field operations at the national level, consortium representatives visited NPMs in each country.
- *PISA quality monitors* (PQMs) – Employed by the consortium and located in participating countries, PQMs visited a sample of schools to record the implementation of the documented field operations in the main study. They typically visited 15 schools in each country.
- *NPM quality surveys* – The consortium developed a series of instruments through which NPMs systematically self-reported on the implementation of key processes at the national level.
- *PISA test administration reports* – PISA test administrators completed a report after each PISA test administration, thus providing an overview of the test administration at the national level.

PREPARATION OF QUALITY MONITORING INSTRUMENTS

The purpose of quality monitoring is to observe and record the implementation of the described procedures; therefore, the field operations manuals provided the foundation for all the quality-monitoring procedures. The manuals that formed the basis for the quality-monitoring procedures were the national project manager manual, test administrator manual, school co-ordinator manual, school sampling preparation manual and the PISA data management manual. The quality monitoring instruments developed from these manuals include a range of sampling forms, a translation and verification schedule for instruments, a NCQM interview schedule, PQM instruments, NPM quality surveys, and a PISA test administrator test session report.

Sampling forms

The consortium developed a series of forms for monitoring school and student sampling activities. The NPM and consortium experts negotiated agreement on sampling plans and outcomes (see Chapter 4).

Translation and verification schedule

This is an instrument detailing the quality monitoring activities for the preparation and translation of instruments monitored instrument preparation at the national level (see Chapter 5).

National centre quality monitor interview schedule

A standard schedule was prepared by the consortium to systematically record the outcomes of the NCQM site visit. The interview schedule recorded information on:

- The general organisation of PISA in that country;
- The quality of test administrators;
- The adequacy of security and confidentiality provisions;
- The selection of the school sample;
- The selection of the student sample;
- The quality of the student tracking procedures;
- The quality of translation procedures;
- The quality of assessment booklet assembly procedures;
- The adequacy and quality of the coding procedures; and
- The independence of the PQMs.

PISA quality monitor instruments

A PQM data collection sheet was developed for PQMs to systematically record their observations during school visits. The data collection sheet recorded information on:

- The use of test script;
- The test session timing;
- The security of materials;
- The environment of the test session;
- The implementation of the student tracking procedures;
- The conduct of the students; and
- The views of the school co-ordinator.

A general observation sheet recorded their general impressions of the implementation of PISA at the national level. The general observation sheet recorded information on:

- The security of materials;
- The overall contribution of test administrators;
- The overall contribution of school co-ordinators;
- The attitude and response of students to the cognitive sessions;

- The attitude and response of students to the questionnaire session; and
- Suggestions for improvement.

NPM quality surveys

An NPM field trial review, an NPM main study review, and a data submission questionnaire enabled NPMs to self-report systematically on all key aspects of field operations and data submission. The NPM main study review made provision for NPMs to self-report on their:

- Use of *KeyQuest* for sampling and data entry;
- Translation, adaptation and verification procedures;
- Preparation of instruments;
- Implementation of exclusions standards; and
- Implementation of coding procedures.

The data submission questionnaire focused on matters specifically relating to the data, including the implementation of national and international options.

PISA test administrator test session report

A test session report for the recording of key test session information enabled the systematic monitoring of test administration. The test session report recorded data on:

- The session date and timing;
- The position of the test administrator;
- The conduct of the students; and
- The testing environment.

IMPLEMENTATION OF QUALITY MONITORING PROCEDURES

Milestone database

The consortium used project milestones negotiated individually with each NPM to monitor the progress of each national centre. Main study testing dates, national centre requirements, and consortium reporting imperatives provided the basis for timeline negotiation. Consortium experts used the milestone database to monitor the progress of national centres through key parts of the project and, when problems were identified, to advise on rectifying actions in order to forestall further operational problems.

National centre quality monitors – Site visits

A consortium representative visited most national centres in the two weeks prior to their main study. For some national centres it was not possible to visit before commencement of the main study due to international health and security alerts. In these cases the consortium representative visited at a time when the alerts were lifted. Consortium representatives visited all national centres. The NCQM used the visit to conduct a half-day PQM training session and a face-to-face interview with the NPM or their representative. Potential problems identified by the NCQM at a national centre were forwarded to the relevant consortium expert for appropriate action.

Video-conferencing facilities enabled the training of PQMs prior to the main study, where the site visit occurred after the main study had commenced.

A comprehensive knowledge of PISA operations and an extensive experience in PISA operations were the criteria for NCQM selection. The NCQMs were trained in conducting site visits to ensure their independence. Nationals with a formal association with the consortium did not visit their own national centre.

PISA quality monitors

NPMs nominated PQMs to the consortium. The candidate's formal independence from the national centre, their experience in, or familiarity with, school operations, their experience or familiarity with educational research, and an ability to speak English or French provided the basis for nomination and selection. Candidates nominated for PQM submitted a resume to the consortium. Where the resume did not match the selection criteria, further information or an alternate nomination was sought. In some countries where the PISA national centre was part of the ministry of education, and where there was a legislative requirement that all staff entering school be ministry employees, it was not possible to fulfil the criteria of PQM independence from the national centre. One national centre was not able to nominate candidates with the required criteria and in this case the consortium organised suitably qualified PQMs.

Typically, two PQMs were engaged for each country, with each PQM visiting seven or eight schools. An NCQM trained all PQMs. The NCQM and PQMs collaborated to develop a schedule of school visits, to ensure that a range of schools was covered and to ensure that the schedule of visits was both economically and practically feasible. The consortium paid the PQM expenses and fees.

The majority of school visits were unannounced. However, the need to organise transport and accommodation made it impractical to keep all PQM visits unannounced.

QUALITY MONITORING DATA

The quality-monitoring data collected from the quality-monitoring instrument was centralised in a single database. Data from the NCQMs, the PQMs, and the NPM quality surveys were data entered by the consortium. Consortium experts used consolidated quality-monitoring reports from the resulting central database to make country-by-country judgements on the quality of field operations, translation, school and student sampling, and coding. The consortium experts used the collected quality-monitoring information to cross check against their own records. The final reports by consortium experts were then used for the purpose of data adjudication (see Chapter 15).

An aggregated report on quality monitoring is also included as Appendix 9.

READER'S GUIDE

Country codes

The following country codes are used in this report:

OECD countries

AUS Australia
AUT Austria
BEL Belgium
BEF Belgium (French Community)
BEN Belgium (Flemish Community)
CAN Canada
CAE Canada (English Community)
CAF Canada (French Community)
CZE Czech Republic
DNK Denmark
FIN Finland
FRA France
DEU Germany
GRC Greece
HUN Hungary
ISL Iceland
IRL Ireland
ITA Italy
JPN Japan
KOR Korea
LUX Luxembourg
LXF Luxembourg (French Community)
LXG Luxembourg (German Community)
MEX Mexico
NLD Netherlands
NZL New Zealand
NOR Norway
POL Poland
PRT Portugal
SVK Slovak Republic
ESP Spain
ESB Spain (Basque Community)
ESC Spain (Catalonian Community)
ESS Spain (Castillian Community)
SWE Sweden
CHE Switzerland
CHF Switzerland (French Community)
CHG Switzerland (German Community)
CHI Switzerland (Italian Community)
TUR Turkey
GBR United Kingdom
IRL Ireland
SCO Scotland
USA United States

Partner countries

BRA Brazil
HKG Hong Kong-China
IND Indonesia
LVA Latvia
LVL Latvia (Latvian Community)
LVR Latvia (Russian Community)
LIE Liechtenstein
MAC Macao-China
RUS Russian Federation
YUG Serbia and Montenegro (Serbia)
THA Thailand
TUN Tunisia
URY Uruguay

List of abbreviations

The following abbreviations are used in this report:

ACER	Australian Council for Educational Research
AGFI	Adjusted Goodness-of-Fit Index
BRR	Balanced Repeated Replication
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CITO	National Institute for Educational Measurement, The Netherlands
CIVED	Civic Education Study
DIF	Differential Item Functioning
ESCS	Economic, Social and Cultural Status
ENR	Enrolment of 15-year-olds
ETS	Educational Testing Service
IAEP	International Assessment of Educational Progress
I	Sampling Interval
ICR	Inter-Country Coder Reliability Study
ICT	Information Communication Technology
IEA	International Association for the Evaluation of Educational Achievement
INES	OECD Indicators of Education Systems
IRT	Item Response Theory
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
ISEI	International Socio-Economic Index
MENR	Enrolment for moderately small school
MOS	Measure of size
NCQM	National Centre Quality Monitor
NDP	National Desired Population
NEP	National Enrolled Population
NFI	Normed Fit Index
NIER	National Institute for Educational Research, Japan
NNFI	Non-Normed Fit Index
NPM	National Project Manager
OECD	Organisation for Economic Cooperation and Development
PISA	Programme for International Student Assessment
PPS	Probability Proportional to Size
PGB	PISA Governing Board
PQM	PISA Quality Monitor
PSU	Primary Sampling Units
QAS	Questionnaire Adaptations Spreadsheet
RMSEA	Root Mean Square Error of Approximation
RN	Random Number
SC	School Co-ordinator
SD	Standard Deviation
SEM	Structural Equation Modelling
SMEG	Subject Matter Expert Group
SPT	Study Programme Table
TA	Test Administrator
TAG	Technical Advisory Group
TCS	Target Cluster Size
TIMSS	Third International Mathematics and Science Study
TIMSS-R	Third International Mathematics and Science Study – Repeat
VENR	Enrolment for very small schools
WLE	Weighted Likelihood Estimates

References

Adams, R.J., Wilson, M.R. and **W. Wang** (1997), "The multidimensional random coefficients multinomial logit model", *Applied Psychological Measurement 21,* pp. 1-24.

Aiken, L. R. (1974), "Two scales of attitudes toward mathematics," *Journal for Research in Mathematics Education 5,* National Council of Teachers of Mathematics, Reston, pp. 67-71.

Andersen, Erling B. (1997), "The Rating Scale Model", in van der Linden, W. J. and R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory,* Springer, New York/Berlin/Heidelberg.

Bandura, A. (1986), *Social Foundations of Thought and Action: A Social Cognitive Theory,* Prentice Hall, Englewood Cliffs, N.J.

Baumert, J. and **O. Köller** (1998), "Interest Research in Secondary Level I : An Overview", in L. Hoffmann, A. Krapp, K.A. Renninger & J. Baumert (eds.), *Interest and Learning*, IPN, Kiel.

Beaton, A.E. (1987), *Implementing the New Design: The NAEP 1983-84 Technical Report* (Report No. 15-TR-20), Educational Testing Service, Princeton, N.J.

Bryk, A. S. and **S.W. Raudenbush** (1992), *Hierarchical Linear Models: Applications and Data Analysis Methods,* SAGE Publications, Newbury Park.

Bollen, K.A. and **S.J. Long** (eds.) (1993), *Testing Structural Equation Models*, SAGE publications, Newbury Park.

Branden, N. (1994), *Six Pillars of Self-Esteem*. Bantam, New York.

Brennan, R.L. (1992), *Elements of Generalizability Theory*, American College Testing Program, Iowa City.

Buchmann, C. (2000), *Measuring Family Background in International Studies of Educational Achievement: Conceptual Issues and Methodological Challenges,* paper presented at a symposium convened by the Board on International Comparative Studies in Education of the National Academy of Sciences/National Research Council on 1 November, in Washington, D.C.

Cochran, W.G. (1977), *Sampling Techniques* (3rd edition), Wiley, New York.

Cronbach, L.J., G.C. Gleser, H. Nanda and **N. Rajaratnam** (1972), *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*, Wiley and Sons, New York.

Eccles, J.S. (1994), "Understanding Women's Educational and Occupational choice: Applying the Eccles *et al.* Model of Achievement-Related Choices", *Psychology of Women Quarterly 18*, Society for the Psychology of Women, Washington, D.C., pp. 585-609.

Eccles, J.S. and **A. Wigfield** (1995), "In the mind of the achiever: The structure of adolescents' academic achievement-related beliefs and self-perceptions", *Personality and Social Psychology Bulletin 21*, Sage Publications, Thousand Oaks, pp. 215-225.

Ganzeboom, H.B.G., P.M. de Graaf and **D.J. Treiman** (1992), "A standard international socio-economic index of occupational status", *Social Science Research 21*, Elsevier, pp. 1-56.

Gifi, A. (1990), *Nonlinear Multivariate Analysis*, Wiley, New York.

Greenacre, M.J. (1984), *Theory and Applications of Correspondence Analysis*, Academic Press, London.

Grisay, A. (2003), "Translation procedures in OECD/PISA 2000 international assessment", *Language Testing 20,* Holder Arnold Journals, pp. 225-240.

Gustafsson, J.E and **P.A. Stahl** (2000), *STREAMS User's Guide, Version 2.5 for Windows,* MultivariateWare, Mölndal, Sweden.

Hacket, G. and **N. Betz.** (1989), "An Exploration of the mathematics Efficacy/mathematics Performance Correspondence", *Journal of Research in Mathematics Education 20,* National Council of Teachers of Mathematics, Reston, pp. 261-273.

Harvey-Beavis, A. (2002), "Student and Questionnaire Development" in OECD, *PISA 2000 Technical Report*, OECD, Paris.

Hatcher, L. (1994), *A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling,* SAS Institute Inc., Cary.

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

Jöreskog, K.G. and **Dag Sörbom** (1993), *LISREL 8 User's Reference Guide,* Scientific Software International, Chicago.

Judkins, D.R. (1990), "Fay's Method for Variance Estimation", *Journal of Official Statistics 6,* Statistics Sweden, Stockholm, pp. 223-239.

Kaplan, D. (2000), *Structural Equation Modeling: Foundation and Extensions,* SAGE Publications, Thousand Oaks.

Keyfitz, N. (1951), "Sampling with probabilities proportionate to science: Adjustment for changes in probabilities", *Journal of the American Statistical Association 46,* American Statistical Association, Alexandria, pp. 105-109.

Lepper, M. R. (1988), "Motivational considerations in the study of instruction", *Cognition and Instruction 5,* Lawrence Erlbaum Associates, Mahwah, pp. 289-309.

Ma, X. (1999), "A Meta-Analysis of the Relationship Between Anxiety Toward mathematics and Achievement in mathematics", *Journal for Research in Mathematics Education 30,* National Council of Teachers of Mathematics, Reston, pp. 520-540.

Macaskill, G., R.J. Adams and **M.L. Wu** (1998), "Scaling methodology and procedures for the mathematics and science literacy, advanced mathematics and physics scales", in M. Martin and D.L. Kelly (eds.) *Third International Mathematics and Science Study, Technical Report Volume 3: Implementation and Analysis,* Center for the Study of Testing, Evaluation and Educational Policy, Boston College, Chestnut Hill.

Marsh, H. W. (1990), *Self-Description Questionnaire (SDQ) II: A theoretical and Empirical Basis for the Measurement Of Multiple Dimensions of Adolescent Self-Concept: An Interim Test Manual and a Research Monograph,* The Psychological Corporation, San Antonio.

Marsh, H. W. (1994), "Confirmatory factor analysis models of factorial invariance: A multifaceted approach" *Structural Equation Modeling 1,* Lawrence Erlbaum Associates, Mahwah, pp. 5-34.

Marsh, H. W. (1999), *Evaluation of the Big-Two-Factor Theory of Motivation Orientation: Higher-order Factor Models and Age-related Changes*, paper presented at the 31.62 Symposium, Multiple Dimensions of Academic Self-Concept, Frames of Reference, Transitions, and International Perspectives: Studies From the SELF Research Centre. Sydney: University of Western Sydney.

Masters, G. N. and **B. D. Wright** (1997), "The Partial Credit Model", in W. J. van der Linden and R.K. Hambleton (eds.), *Handbook of Modern Item Response Theory,* Springer, New York/Berlin/Heidelberg.

Meece, J., A. Wigfield and **J. Eccles** (1990), "Predictors of Maths Anxiety and its Influence on Young Adolescents' Course Enrolment and Performance in Mathematics", *Journal of Educational Psychology 82,* American Psychological Association, Washington, D.C., pp. 60-70.

Middleton, J.A. and **P.A. Spanias** (1999), "Findings, Generalizations, and Criticisms of the Research", *Journal for Research in Mathematics Education 30*, National Council of Teachers of Mathematics, Reston, pp. 65-88.

Mislevy, R.J. (1991), "Randomization-based inference about latent variable from complex samples", *Psychometrika 56*, Psychometric Society, Greensboro, pp. 177-196.

Mislevy, R.J. and **K.M. Sheehan** (1987), "Marginal estimation procedures", in A.E. Beaton (ed.), *The NAEP 1983-1984 Technical Report* (Report No. 15-TR-20), Educational Testing Service, Princeton, N.J.

Mislevy, R.J. and **K.M. Sheehan** (1980), "Information matrices in latent-variable models", *Journal of Educational Statistics 14.4,* American Educational Research Association and American Statistical Association, Washington, D.C., and Alexandria, pp. 335-350.

Mislevy, R.J., A.E. Beaton, B. Kaplan and **K.M. Sheehan.** (1992), "Estimating population characteristics form sparse matrix samples of item responses", *Journal of Educational Measurement 29*, National Council on Measurement in Education, Washington, D.C., pp. 133-161.

Multon, K. D., S. D. Brown and **R.W. Lent** (1991), "Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation", *Journal of Counselling Psychology 38,* American Psychological Association, Washington, D.C., pp. 30-38.

Muthén, B. O., S. H. C. du Toit and **D. Spisic** (1997), "Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical outcomes", *Psychometrika*, Psychometric Society, Greensboro.

Muthen, L. and **B. Muthen** (2003), *Mplus User's Guide Version 3.1,* Muthen & Muthen, Los Angeles.

Nishisato, S. (1980), *Analysis of Categorical Data: Dual Scaling and its Applications,* University of Toronto Press, Toronto.

OECD (Organisation for Economic Co-Operation and Development) (1999), *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 (2002), *PISA 2000 Technical Report,* OECD, Paris.

OECD (2003), *Student Engagement at School: A Sense of Belonging and Participation: Results from PISA 2000,* OECD, Paris.

OECD (2004a), *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, OCED, Paris.

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

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

OECD (2005a), *PISA 2003 Data Analysis Manual: SAS® Users,* OECD, Paris.

OECD (2005b), *PISA 2003 Data Analysis Manual: SPSS® Users,* OECD, Paris.

Owens L. and **J. Barnes** (1992), *Learning Preference Scales*, Australian Council for Educational Research, Hawthorn.

Rasch, G. (1960), *Probabilistic models for some intelligence and attainment tests,* Nielsen and Lydiche, Copenhagen.

Rust, K. (1985), "Variance estimation for complex estimators in sample surveys", *Journal of Official Statistics 1,* Statistics Sweden, Stockholm, pp. 381-397.

Rust, K. and **J.N.K. Rao** (1996), "Variance estimation for complex surveys using replication techniques", *Statistical Methods in Medical Research 5,* Holder Arnold Journals, pp. 283-310.

Sändal, C.E., B. Swensson and **J. Wretman** (1992), *Model Assisted Survey Sampling,* Springer-Verlag, New York.

Schaffer, E. C., P. S. Nesselrodt and **S. Stringfield** (1994), "The Contribution of Classroom Observation to School Effectiveness Research" in Reynolds *et. al.* (eds.), *Advances in School Effectiveness Research and Practice,* Pergamon, Oxford/New York/Tokyo.

Schulz, W. (2003), *Validating Questionnaire Constructs in International Studies. Two Examples from PISA 2000,* paper presented at the Annual Meeting of the American Educational Research Association (AERA) in Chicago, 21-25 April.

Schulz, W. (2004), "Mapping Student Scores to Item Responses", in W. Schulz and H. Sibberns (eds.), *IEA Civic Education Study. Technical Report*, IEA, Amsterdam.

Sirotnik, K. (1970), "An analysis of variance framework for matrix sampling", *Educational and Psychological Measurement 30,* SAGE Publications, pp. 891-908.

Slavin, R. E. (1983), "When does cooperative learning increase student achievement?" *Psychological Bulletin 94,* American Psychological Association, Washington, D.C., pp. 429-445.

Statistical Solutions (1992), *BMDP Statistical Software*, Statistical Solutions, Los Angeles.

Teddlie, C. and **D. Reynolds** (2000) (eds.), *The International Handbook of School Effectiveness Research,* Falmer Press, London/New York.

Thorndike, R.L. (1973), *Reading Comprehension Education in Fifteen Countries: An Empirical Study*, Almquist & Wiksell, Stockholm.

Travers, K. J., R.A. Garden and **M. Rosier** (1989), "Introduction to the Study", in D.A. Robitaille and R.A. Garden (eds.), *The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics Curricula,* Pergamon Press, Oxford.

Travers, K. J. and **I. Westbury** (1989), *The IEA Study of Mathematics I: Analysis of Mathematics Curricula,* Pergamon Press, Oxford.

Verhelst, N. (2004), "Generalizability Theory", in Council of Europe, *Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, (Section E), Council of Europe (DGIV/EDU/LANG (2004) 13), Strasbourg.

Warm, T. A. (1989), "Weighted Likelihood Estimation of Ability in Item Response Theory", *Psychometrika* 54, Psychometric Society, Greensboro, pp. 427-45.

Wigfield, A., J. S. Eccles and **D. Rodriguez** (1998), "The development of children's motivation in school contexts", in P. D. Pearson. and A. Iran-Nejad (eds.), *Review of Research in Education 23,* American Educational Research Association, Washington D.C., pp. 73-118.

Wilson, M. (1994), "Comparing Attitude Across Different Cultures: Two Quantitative Approaches to Construct Validity", in M. Wilson (ed.), *Objective Measurement II: Theory into Practice,* Ablex, Norwood, pp. 271-292.

Wolter, K.M. (1985), *Introduction to Variance Estimation*, Springer-Verlag, New York.

Wu, M.L., R.J. Adams and **M.R. Wilson** (1997), *ConQuest: Multi-Aspect Test Software* [computer program], Australian Council for Education Research, Camberwell.

Zimmerman, B.J. and **D.H. Schunk** (eds.) (1989), *Self-Regulated Learning and Academic Achievement. Theory, Research and Practice,* Springer, New York.

Table of Contents