

# TIMSS and PIRLS 2011

## Quality Control in the Data Cleaning Process

Oliver Neuschmidt

In order to ensure high-quality data, the IEA DPC implemented an extensive set of interrelated data checking and cleaning procedures using the Statistical Analysis Software (SAS) software. The quality control steps outlined below were taken to ensure that all procedures were conducted in the correct sequence, that no special requirement was overlooked, and that the cleaning process was implemented in an absolutely standardized way, regardless of the person in charge.

- ◆ Before being used with real data, all data cleaning programs were thoroughly tested using simulated data sets containing a wide range of possible problems and inconsistencies.
- ◆ Incoming data and documentation from national centers were registered in a specific material receipt database along with their dates of arrival and any specific issues relevant for the data cleaning.
- ◆ All national adaptations to the international instruments were recorded in a “National Adaptation Database”, along with any corresponding recoding instructions needed in order to recode data back to the international format.
- ◆ The cleaning process was organized following a standardized data processing workflow that was strictly and consistently applied to all national datasets, so that deviations from the predefined cleaning sequence were not possible.
- ◆ Once the data cleaning had been completed for a country, an additional quality control step repeated the cleaning process from the beginning to detect any problems that might have been inadvertently introduced during the process.
- ◆ IEA DPC staff worked closely with the national centers, and provided each country with processed data files and accompanying documentation and statistics at various stages of data processing, enabling the country to conduct a thorough review and correction of any inconsistencies.

- ◆ As an additional data verification step, each version of the data prepared for send-out either to the national centers or to the International Study Center, was carefully compared with the preceding data version.
- ◆ All national adaptations that countries recorded in their documentation were verified against the structure of the national data files. All deviations from the international data structure were recorded in the National Adaptations Database. Whenever possible, national deviations were recoded to follow the international data structure. However, if international comparability could not be assured, the corresponding data had to be removed from the international database.
- ◆ At the end of the cleaning process, a series of checks to control and identify irregularities in the data pattern were performed and data was corrected where necessary. This included checks about matching response patterns across records, verification of punching errors of missing values, the plausibility of testing dates, etc.