TIMSS and PIRLS 2011

Resolving Inconsistencies in the TIMSS and PIRLS 2011 Data

Oliver Neuschmidt

e number of inconsistent or implausible responses in the data les varied from country to country, but no country's data were completely free of inconsistent responses. Each problem was recorded in a database along with a description of the problem. Issues that could not be resolved using systematic cleaning rules were reported to the NRC so that original data-collection instruments and tracking forms could be cross-checked. Where the national center could not solve the problems by inspecting the instruments and forms, a set of nal cleaning rules were applied.

As outlined in **Creating and Checking the TIMSS and PIRLS 2011 Databases**, the cleaning process involved the adherence to a number of standardized cleaning rules. e cleaning rules were implemented to ensure that the data from countries and benchmarking entities were processed in a uniform manner.

Listed below are some of the common inconsistencies encountered during the data cleaning process as well as information on how they were resolved:

Filter questions

Filter questions are preliminary questions that direct respondents away from questions that do not apply to them (for example: "Do you have a computer at home? If no, skip to next section"). In every country, a small minority of respondents answered questions in sections that were logically implausible. Inconsistencies between the coding of lter questions and their dependent questions usually were treated automatically by the cleaning programs. If the lter variable contained a valid value and its dependent questions were correctly skipped, dependent variables were coded as "logically not applicable". If a response to a lter question was equivalent to "No", indicating that the dependent questions were not applicable, but the dependent questions contained valid answers then the lter question was recoded to "Yes".

Split Variable Checks

Split variable checks were applied to "Yes/No" lists and percentage list questions where the responses were coded into several variables. For example, question 5 in the TIMSS and PIRLS Grade 4 Student

Questionnaire listed a number of home possessions and asked the student to check all that applied. Student responses were captured in a series of eleven variables, each one was coded as “Yes” if the corresponding possession was checked or it was coded as “No” if left unchecked. Occasionally, students checked the “Yes” boxes but le the “No” boxes unchecked or missing. Since in these cases it could be assumed that the unmarked boxes actually meant “No”, the corresponding variables were imputed accordingly. e individual responses to percentage questions were summed up and, if they fell outside the range of 90 to 110, they were coded to “omitted”.

Implausible Numeric dri