Chapter 8

Creating and Checking the PIRLS International Database

Juliane Barth and Oliver Neuschmidt

8.1   Overview

e PIRLS 2006 International Database is a unique resource for policy makers and analysts, containing student reading achievement and background data

†: $' $#   ? $'   _`{|} $ $ $ # >

TI

8.6.2 Valid Range Check

"Valid range" indicates the range of the values considered to be correct and meaningful for a specific variable. For example, students' gender had two valid values: "1" for girls and "2" for boys. All other values were considered invalid.

There also were questions in the school and teacher background questionnaires where the respondent wrote in a number—the principal was asked to supply the school enrollment, for example. For such variables, valid ranges may vary from country to country, and the broad ranges were set as acceptable to accommodate variations. It was possible for countries to adapt these ranges according to their needs, although countries were advised that a smaller range would decrease the possibility of mispunches. Data cleaning at the IEA DPC did not take smaller national ranges into account. Only if values were found outside the international accepted range were the cases mentioned in the list of inquiries sent to the countries.

8.6.3 Identification Variable (ID) Cleaning

Each record in a data file should have a unique identification number (ID). Duplicate ID numbers imply an error of some kind. If two records shared the same ID and contained exactly the same data, one of the records was deleted and the other remained in the data file. In the rare case that records contained different data apart from the ID numbers, and it was not possible to detect which records contained the "true data", both records were removed from the data. However, the IEA DPC made every effort to keep such losses to a minimum.

The ID cleaning focused on the student background questionnaire, because most of the critical variables were present in this type. Apart from the unique ID, there were variables pertaining to students' participation and exclusion status, as well as dates of birth and dates of testing used to calculate age at the time of testing. The Student Tracking Forms were essential in resolving any anomalies, as was close cooperation with National Research Coordinators.

T

†: $' $#   ? $'   _`{|} $ $ $ # >

†: $' $#   ? $'   _`{|} $ $ $ # >

TI