

PIRLS 2011 Data Files) and combined them with the IEA DPC, and have since
checked and editing known as 'data cleaning' began. Data cleaning
in the case of checking data for inconsistencies and formatting the data
correctly and standardizing it. The following concerns the data cleaning
processes undertaken by the IEA and the data accuracy of the information collected
in each of the participating countries, the information collected from the
international forms, and the information from the data, each, and
initially could be matched according to the data file. This chapter describes
the data cleaning process undertaken in the case of the PIRLS and TIMSS
2011 data base.

The data cleaning process involved collaboration between the national
centres, the IEA Data Processing and Research Centre (IEA DPC), the TIMSS &
PIRLS International Study Centre and Statistics Canada. IEA Data Processing
and Research Centre was the central hub for the data cleaning process. The IEA
DPC was responsible for checking the data before forming each country, a list of
standardized data cleaning rules to be applied if the accuracy and consistency of the
data and documentation are not satisfactory from the international level.
Data were checked at each country's national centre, and entered into
combined in the IEA DPC. The National Research Centre (NRC)
collaborated with the IEA DPC to ensure the integrity of the data which emerged during
the data cleaning process, and the NRC checked in the information from the
national/benchmarking agencies and data base() created by the IEA DPC.
The TIMSS & PIRLS International Study Centre invited the NRC to help
with the data and almanac containing information on each variable that
the national centre could provide for the data management international electronic.

included a range of data verification checks that facilitated identification and correction of deficiencies before data submission to the IEA DPC.

Uncompleted files for each check of the data are highlighted, the NRC submitted the data as a .csv, .deb, .k and the accompanying documents to the IEA DPC. The documents submitted included National Assessment Form, Student Tracking Form, and Student-Teacher Linkage Form (TIMSS) Student Listing Form (PIRLS). Microsoft submitted all required documents in a single file, which greatly expedited the data checking process. Microsoft's incoming mail data documents were submitted to the handling manager.

Receipt of the Data at the IEA Data Processing and Research Center

Uncompleted files and a list of errors, each submitted to the IEA Data Processing and Research Center for data cleaning and formatting. Upon arrival at the IEA DPC, the data and documents were registered in a special database developed by the center. Each entry in the database included the data file name, a list of errors, and the data cleaning. A record in the database was created for each national data file. All data in the international level were recorded in the database, including information regarding the data match to the international format.

In order to efficiently and accurately complete the data cleaning process, the IEA DPC developed an efficient file-based data cleaning program using the Statistical Analysis Software (SAS) package. All data cleaning programs were highly automated before being applied to the 2011 data.

Documentation and Structure Check

Data cleaning began with an inspection of the data file and a check of the documents submitted by the NRC. Documents submitted included the National Assessment Form, Student Tracking Form, Student-Teacher Linkage Form (TIMSS) Student Listing Form (PIRLS), and Teacher Administration Form. Each of these files was checked to ensure that all documents had been submitted and welcomed. At the same time, the tracking and naming information was added to the WinW3S

It is not clear from the data that the cleaning firm. In the case where the cleaning firm was identified, the IEA DPC applied a general cleaning rule except for the problem. As a result, if it had been identified, IEA DPC applied the best SAS recording criteria and remaining cleaning in the data file.

Identification (ID) Cleaning

Under the current information, the identification (ID) variable in each data file was examined. Each record in the data file had a unique identification variable. If a record had the same ID number and contained the same data, the IEA DPC deleted the record in the database (excluding the record). If a record had the same ID number and data, and it was not possible for the IEA DPC and NRC to identify which record contained the data, both records were removed from the database. The deleted records were sequential, and in this case, all were deleted.

Although the ID cleaning included all files, the ID cleaning effort did not include the background and reference files, which contained most of the critical ID variable. As a result, the unique ID number, it was necessary to check the variable in the file being identified and deleted.

Linkage

Resolving Inconsistencies in Background Questionnaire Data

After each file matched the international standard decided in the checklist, and the identification and linkage cleaning was complete, a separate file for data cleaning was created. The file was then used to clean the background questionnaire data. The cleaning program identified, and in many cases automatically corrected, inconsistencies in the data. When inconsistencies could not be reconciled automatically, the user was alerted and a message box was displayed. The user was then asked to make an informed decision. Among the inconsistencies encountered in the data were: (a) missing data in the background questionnaire, (b) missing data in the main file and the background questionnaire, (c) missing data in the main file and the background questionnaire, and (d) missing data in the background questionnaire. (See [Resolving Inconsistencies in the TIMSS and PIRLS 2011 Data](#)).

The user was then asked to make an informed decision. Among the inconsistencies encountered in the data were: (a) missing data in the background questionnaire, (b) missing data in the main file and the background questionnaire, (c) missing data in the main file and the background questionnaire, and (d) missing data in the background questionnaire. (See [Resolving Inconsistencies in the TIMSS and PIRLS 2011 Data](#)).

The summary statistics were calculated by the TIMSS & PIRLS International Study Centre and included weighted means for all variables available for each country. For categorical variables, the percentage of respondents choosing each response in the study was calculated. For numerical variables, the distribution measures were calculated, including the minimum, maximum, mean, standard deviation, median, mode, and percentile. For bivariate variables, the percentage of missing data were calculated. Additionally, the International Study Centre provided item analysis and reliability statistics including the number of valid cases, percentage, percentage correct, Rasch item difficulty, Cronbach's alpha, and Fisher's exact test for independence of the data at the international and national level in terms of reliability, nonequivalence, etc.

Final Product – the TIMSS and PIRLS 2011 International Databases

The data cleaning process implemented at the IEA DPC ensured that the TIMSS and PIRLS 2011 international databases contained high-quality data (see [2011 TIMSS International Database](#) and [2011 PIRLS International Database](#)).

More specifically, the process ensured that:

- Information coded in each variable was internationally comparable

- National data were entered accurately in all variables

- All entries in the database could be accessed, linked, and accessed at the level

- Sampling weights and student achievement scores were available for international comparison