

**International Data Base (IDB) Analyzer Demonstration**  
**Baku, September 17<sup>th</sup> 2007**  
**Presenter: Oliver Neuschmidt**

**1 Introduction**

The International Database (IDB) Analyzer© - is a plug-in for the Statistical Package for the Social Sciences (SPSS) developed by the IEA Data Processing and Research Center that can be used to combine files and analyze data from large scale assessments. The analysis performed by the SPSS code generated by the IDB Analyzer© takes into account information from the sampling design in the computation of sampling variance. In addition, the SPSS code is also able to handle multiple plausible values and calculates variance of estimates due to imputation. The IDB Analyzer© enables the user to conduct statistical hypothesis testing between groups in the population without having to write any programming code.

The IDB Analyzer consists of two modules. The first module is used to create analysis data sets by combining data from different countries, merging school, teacher, and student level files and selecting subsets of analysis variables. The second module provides different procedures for analysis like the computation of estimates of means and percentages of any variable of interest overall for a country, and for specific subgroups within a population.

**2 System Requirements**

The IDB Analyzer© will work on IBM compatible computers using Microsoft Windows 2000, XP or any later version.

As the software is directly accessing the SPSS engine, SPSS for Windows version 11 or higher needs to be pre-installed.

Additionally, the .NET Framework 1.1 along with service pack 1 is required and will be installed along with the software.

For the installation process local administrator rights are requirement.

**3 Databases supported**

In the current version of the IDB Analyzer, international databases of all published cycles of TIMSS and PIRLS in SPSS format are supported.

**4 Analyzing the IEA Survey data by using the IDB-Analyzer**

The PIRLS & TIMSS sampling design applied a stratified multistage cluster-sampling technique to the problem of selecting efficient and accurate samples of students while working with schools and classes. However, these procedures complicate the task of computing standard errors to quantify sampling variability,

as the computational formulas derived from assumptions of simple random sampling generally underestimate the true variability in the data.

Additionally, each student's achievement on the entire assessment is measured depending on the student's responses to the items that they took and the student's background characteristics. Because there is some error inherent in this imputation process, generally five estimates, or "plausible values," for each student on each of the scales is drawn. Each student, therefore, has five estimates of his or her achievement on each of the scales and subscale measured.

Therefore it is very important to be able to quantify the uncertainty associated with these statistics.

In PIRLS & TIMSS, the jackknife procedure is used to provide a robust estimate of the standard error of each statistic presented in the International Reports. The jackknife standard errors include both an error component due to sampling variation and an error component due to variation among the five plausible values generated for each student. The standard errors may be used to create confidence intervals for statistics computed from the TIMSS data.

All statistical procedures applied to analyses of school, teacher, and student variables by the IDB Analyzer apply the jackknife algorithm described above.

## **5 Analyzing the IEA Survey data by using the IDB-Analyzer**

The IDB Analyzer consists of two modules which are executed as independent applications. Both modules can be started from the system menu or from within the SPSS menu.

### **a) The Merge Module**

Data from IEA studies are generally distributed separately by country. This first module also facilitates the combination of data from different countries into one dataset that subsequently can be used for analyses either with the IDB Analyzer© itself, or with another software chosen by the user. Joining files from different countries will permit conducting cross- country analyses.

The International Databases from IEA studies generally consists of separated data files for each participating country. The merge module for the IDB Analyzer provides a graphical user interface allowing the user to join different countries' data files into one dataset for analysis for conducting cross-country analyses.

Even more important, the module enables merging of school, teacher, and student level information while simultaneously allowing the user to select a subset or all of the variables in the selected files as needed for further analysis.

According to IEA's study design, the student level will be the principal level of analyses and any additional information obtained from teacher or school level will be represented as an attribute of the student. Analyses therefore can only be

reported in reference to the student level and the IDB Analyzer© generates SPSS code that ultimately creates a student level file with the teacher and/or school data merged onto it. Any SPSS code created by the IDB Analyzer can be saved from the SPSS syntax windows to be re-used when rerunning the merging steps. Data created by the Merge module now can be processed either with the Analysis Module of the software or by any other analysis software accepting SPSS input files.

## **b) The Analyses Module**

The second module of the IDB Analyzer is used to analyze the data, whether or not these have been pre-processed with the first module. The second module contains a number of procedures for analysis.

Currently the following procedures are available:

### **Background means:**

This procedure allows to compute the Jackknife standard error for means of variables other than plausible values and percentages.

Example: calculating the mean age for eight grade students in each country

### **PV means:**

Estimating the mean achievement scores (based on plausible values) and percentages for any set of classification variables overall for a country, and for specific subgroups within a population.

Example: calculating mean score by gender for each participating country

### **Background Regression:**

Computing multiple linear regressions between the specified dependent and independent variables within subgroups.

Example: predicting the number of books at home by gender and country

### **Achievement Regression:**

Computing multi linear regressions between a set of plausible values (dependent variables) and independent variables within subgroups

Example: testing for significance of the effect of student-, teacher or school-level background variables on student achievement.

The option 'Dummy recoding' allows the automatical creation of so-called zero-one coded dummy variables out of any categorical independent variable prior to computing the regression coefficient.

All procedures, regardless of whether they focus on means, percentages or regression coefficients are computed making use of appropriate sampling weights while standard errors are computed using the jackknife repeated replicate method (JRR) as described in the TIMSS technical report (Martin et al., 2004). Methods 2) and 4) replicate analyses with all 5 plausible values available in the international databases to assess the impact of the imputation error.

## 6 Availability of the software

The IEA IDB Analyzer© currently is available free of charge at:  
[http://www.iea.nl/iea\\_studies\\_datasets.html](http://www.iea.nl/iea_studies_datasets.html).

## 7 Appendix: Short Tutorial of the IDB Analyzer

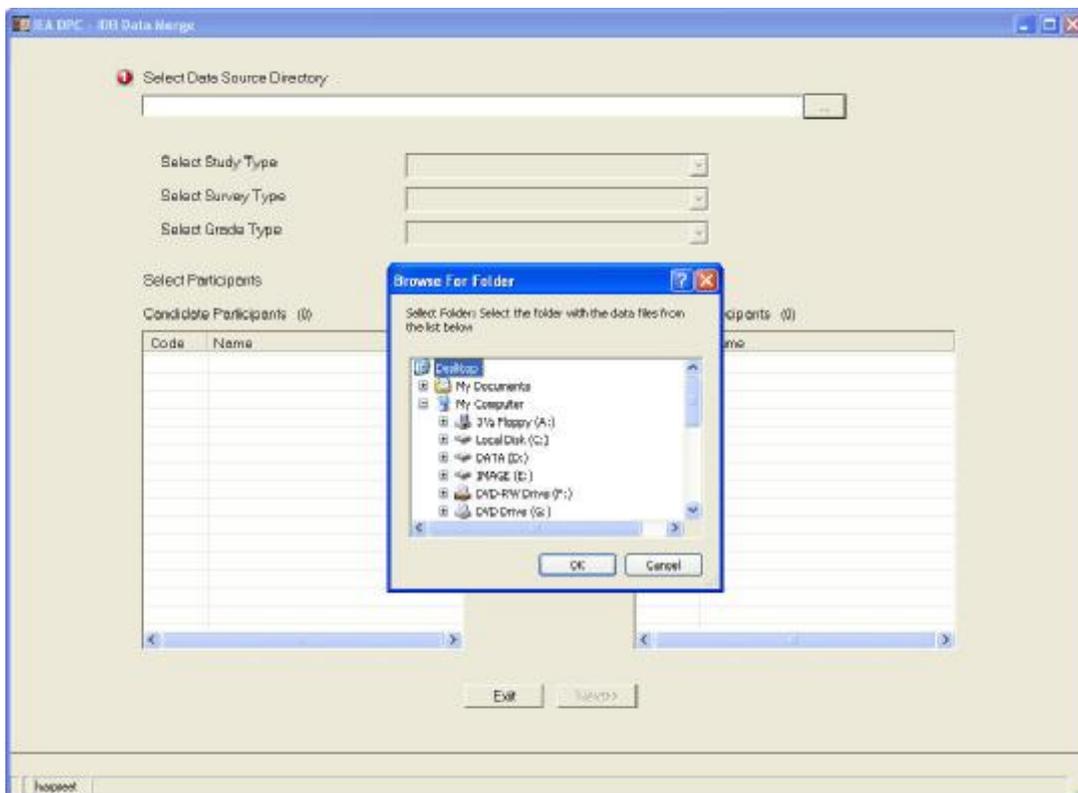
### The Merge Module

#### a) Starting the application

- i. Start from Windows: Select *Start Menu à Programs à IEA à IDBAnalyzer à Prepare Merge Data*
- ii. Start from SPSS menu: Select *IEA- IDBAnalyzer à Merge IDB Data*

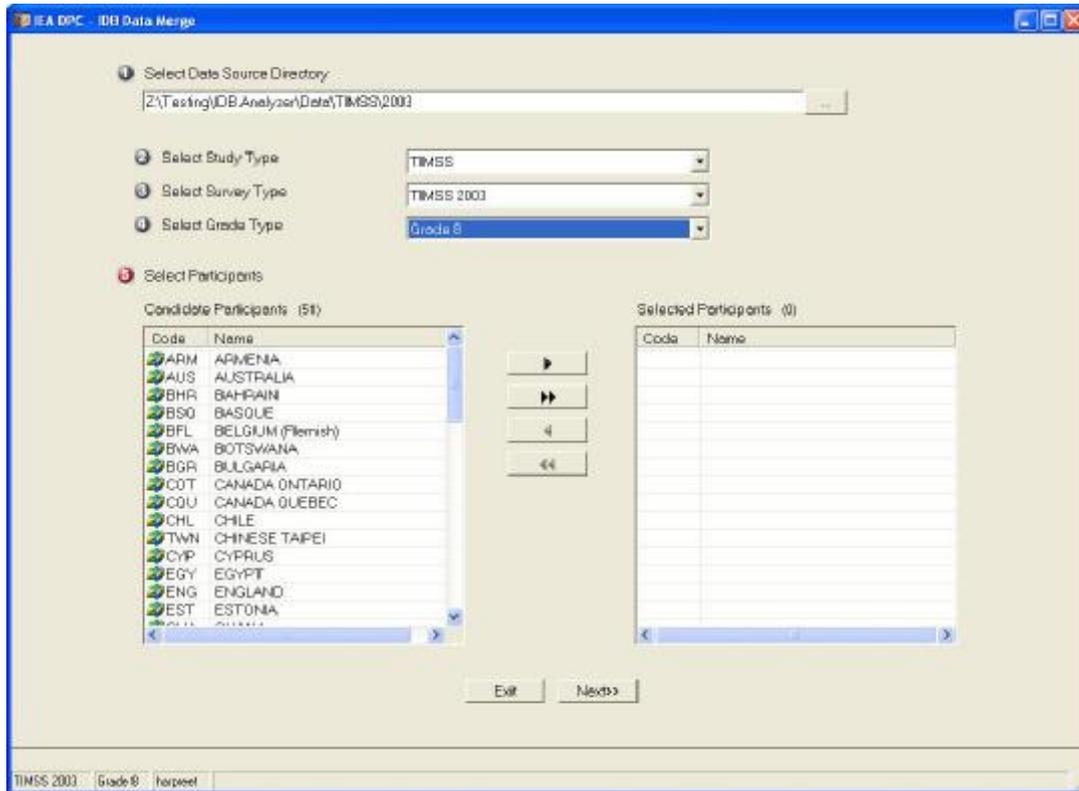
#### b) Running the application

- iii. Select your data source directory by clicking the button next to the upper edit field



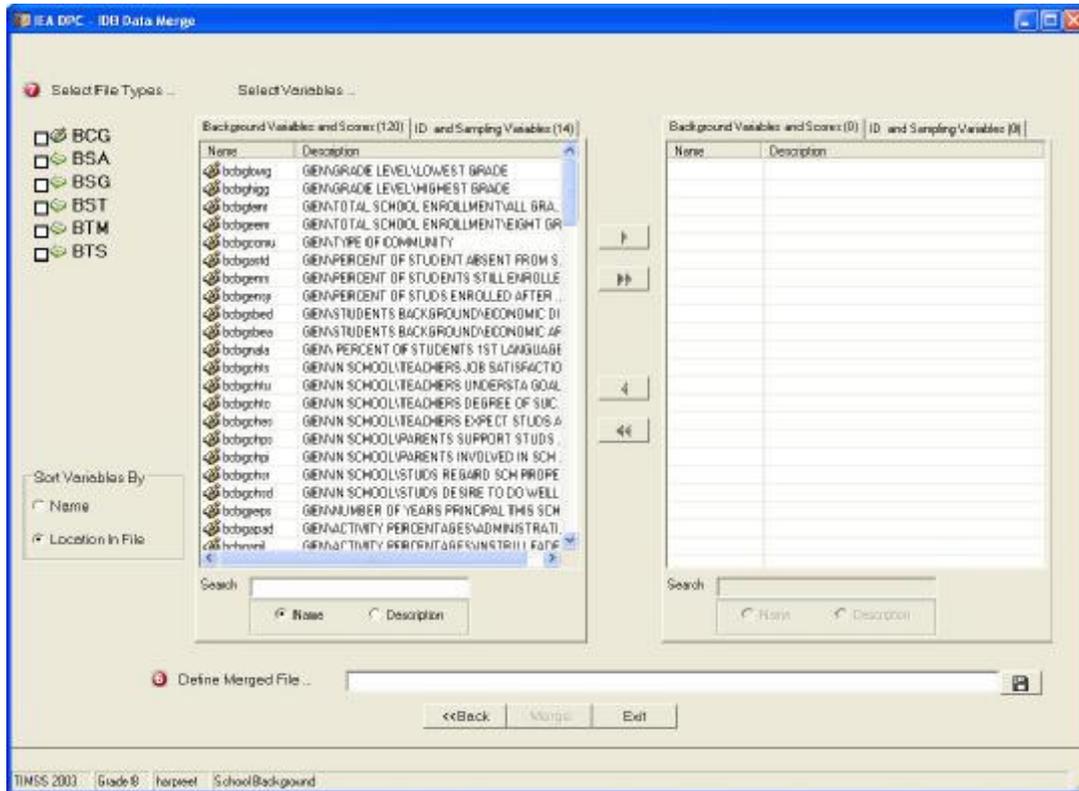
### 1. Prepare&Merge --Prepare Data

- iv. Now the study type, cycle, grade, and participating countries can be selected and de-selected from the window.



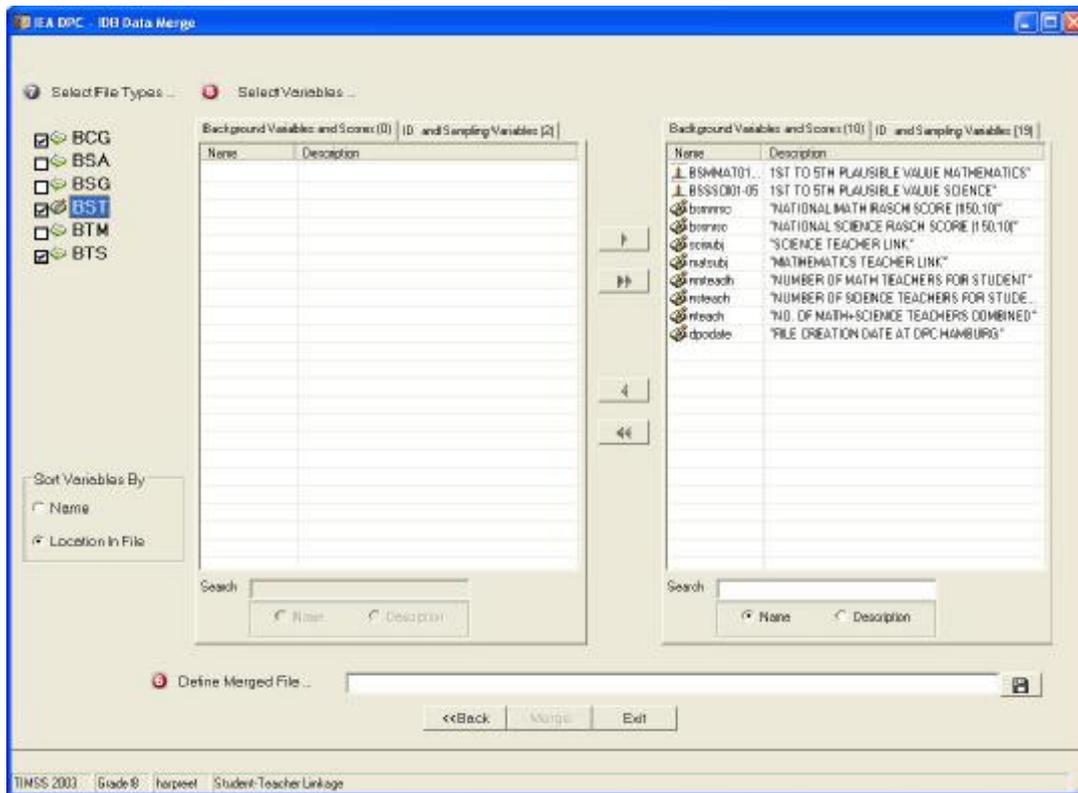
## 2. Prepare&Merge --Prepare Data

- v. Click “Next” to invoke a second screen, where file types and variables can be selected.



### 3. Prepare&Merge -- Select FileType and Variables -1

- vi. After selecting the desired file types in the upper left part of the screen (7), analyses variables to be kept in the final file may be selected.

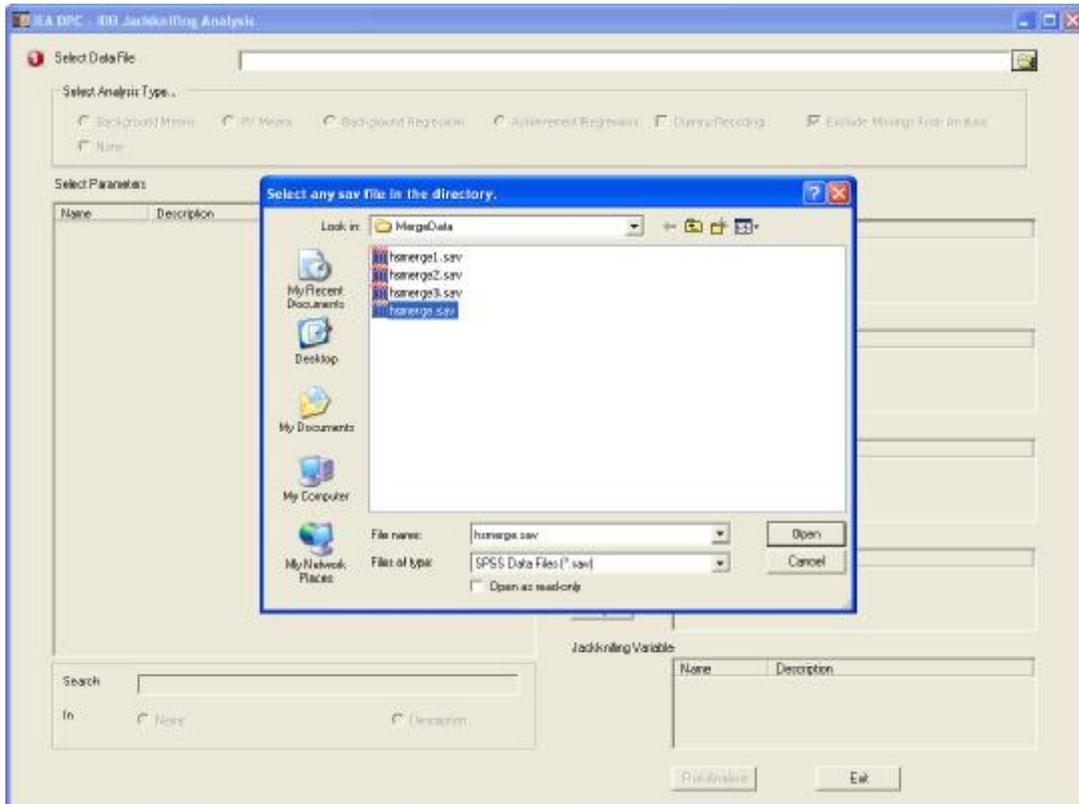


#### 4. Prepare&Merge -- Select File Type and Variables – 2

- vii. After file and variable selection process has been finalized, the file name specification for the merged file should be made. A Dialog box will appear when clicking the button on the right side next to the Define Merged File... edit field.
- viii. Press the *Merge*-button in order to start the merge process

#### Analysis Module

- a) Starting the application
  - ix. Start from Windows: Select *Start Menu à Programs à IEA à IDBAnalyzer à Analyze Data*
  - x. Start from SPSS windows: Select *IEA- IDBAnalyzer à Analyze IDB Data*
- b) Running the application
  - xi. After starting the application, define the data file to be analyzed by clicking the button next to the upper edit field (1)



5. Analyze -- Start-up and select Data source file



variables). Weights and Jackknife variables are pre-selected the software.

- xiv. As soon as all settings are made, click „Run Analysis“, and the software will trigger the SPSS engine to run the necessary analysis routines. All output created is original SPSS output and will be displayed in the corresponding SPSS syntax- and output windows.

### c) Output

Depending on the analysis requested, different SPSS outputs are available. Below you will find an example for the 'PV Means' analysis method.

The screenshot shows the SPSS Output window for a 'PV Means' analysis. The report displays the following data:

COUNTRY ID*	TAKE UP STUDENTS*	N	TOTWGT	MNPV	MNPV_SE	TOT	TOT_SE
AUSTRALIA	G13L	2465	15,600	517,03	4,64	51,15	2,25
	BOF	2198	128807	347,48	4,88	48,87	2,28
BAHRAIN	G13L	1065	5317	452,81	2,66	51,68	,44
	BOF	2174	5126	421,95	2,27	45,67	,44

### 7. Analyze – Output of PV means

In this example the achievement mean for the subgroup gender is reported for two countries (Australia and Bahrain). The following information is listed:

N = Number of cases in the groups

TOTWGT = Total student weight – contains estimates in the population that belong to the group defined by each specific combination of the classification variable (ITSEX)

MNPV = here: Means of the mathematics plausible values (BSMMAT01-05)

MNPV\_SE = Standard error of MNPV

PCT = Weighted percentage of students in the groups for the classification variable (ITSEX)

PCT\_SE = Standard Error of PCT

## 8 Conclusion

The advantage of using the IDB analyzer is that it takes into account the complex sampling design of the TIMSS study resulting in correct estimates of standard errors. In contrast, algorithms implemented in SPSS and many other frequently used statistics programs assume a simple random sample which will lead to misleading results that, in turn, may lead to inappropriate conclusions.

## 9 References

- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grade*. Chestnut Hill, MA: Boston College.
- Martin, M.O. (Ed.) (2005). *TIMSS 2003 User Guide for the International Database*. Chestnut Hill, MA: Boston College.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS Technical Report*. Chestnut Hill, MA: Boston College.