International Data Base (IDB) Analyzer Demonstration Baku, September 17th 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 clustersampling 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 zeroone 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 access 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.

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

Select Study Type	1	2
Select Survey Type	1	<u>-</u>
Select Grade Type	Γ	
Select Participants	Browse For Folder	
	Select Folder Select the folder with the data files from	
Conclidate Participants (0) Code Name	the list below	cipants (0)
Code Trame	(1) Dealtop	
	B C Ny Documenta B B Ny Computer	
	H (Ja 3Va Ploppy (A:) H (See Local Disk (C:)	
	🗄 🛩 DATA (D:)	
	E ← INAGE (E:) E → DAD-RW Drive (*:)	
	E 🕹 DVD Drive (G)	
	OC Cancel	
	Carry Control	
¢	5 K	3
	Exit Nation	

1. Prepare&Merge --Prepare Data

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

Z\Testin	g\IOB Analyzer\Data\TIMS	S\2003			1.44	
Select	Study Type	TIMES			<u>·</u>	
Select	Survey Type	TIMSS 2003			•	
O Select	Grade Type	Grode 8		1	-	
Condicion	e Parlicipante (51) Name	6		Selected	f Participants (0) Name	
		~	· · · · · · · · · · · · · · · · · · ·	Code	Name	1
	ARMENIA AUSTRALIA					
	BAHRAIN		++			
	BASQUE					
	BELGIUM (Flemish) BIOTSWANA		4			
2BGR	BULGARIA		44			
	CANADA ONTARIO					
2COU 2CHL	CANADA QUEBEC					
	CHINESE TAIPEI					
2 CYP	CYPRUS					
	EGYPT					
2ENG 2EST	ENGLAND ESTONIA					
all contract	Survey	×		<		0
10		2		14		100
			Exit Nex	0.0		

2. Prepare&Merge --Prepare Data

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

		sables and Scores (120) ID and Sampling Valiables (14)		And in case of the local division of the loc	Vasiables and Scores (D) ID and Sampling Variables (0)
S BSA	Nerse	Description		Norse	Description
BSG	Scholad Scholad	GENVGRADE LEVEL/LOWEST BRADE GENVGRADE LEVEL/HIGHEST GRADE			
BST	Soboteni (Soboteni	GENATOTAL SCHOOL ENPOLLMENTVALL BRA.		-	
BTM	di bobgeen	GENATOTAL SCHOOL ENROLLMENTALE ONC		-	
	S bobgconu	GENATYPE OF COMMUNITY	P		
BTS	di bobgantd	GENAPERCENT OF STUDENT ABSENT FROM S.			
	d bobgens	GENAPERCENT OF STUDENTS STILL ENROLLE	85		
	a bobgenty	GENAPERCENT OF STUD'S ENROLLED AFTER			
	4 bobgsbed	GENASTUDENTS BACKSPOUNDAECONOMIC DI			
	di bobgsbee	GENASTUDENTS BACKGROUNDAECONDMIC AF			
	🤯 bobgnala	GENA PERCENT OF STUDENTS 1ST LANGUAGE			
	di bobgohis	GENVIN SCHOOL/TEACHERS JOB SATISFACTIO	1000		
	Schootha Schootha	GENVIN SCHOOL/TEADHERS UNDERSTA GOAL GENVIN SCHOOL/TEADHERS DEGREE OF SUC			
	Sobgene	GENVIN SCHOOL/TEACHERS EXPECT STUDS A			
	a bobachas	GENVIN SCHOOL/PARENTS SUPPORT STUDS	44		
	(bobgchai	GENVIN SCHOOL/PARENTS INVOLVED IN SCH.			
Variables By	bobgoha	GENVIN SCHOOLVSTUDS REGARD SCH PROPE			
	@bcbgchtd	GENVIN SCHOOL/STUDS DESIRE TO DO WELL			
eme	di bobgeeps	GENAUMBER OF YEARS PRINCIPAL THIS SCH			
position in File	🎸 bobgapad	GENACTIVITY PERCENTAGES ADMINISTRATI.			
DEBUGHEREN	citis hotocomi	GENNALTINETY PERCENTARESUNGTRULE FALLE			
		and the second se		0	
	Search			Search	
	6	Name C Description			C Hann C Description

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.

⇔ BCG ⇔ BSA ⇔ BSG Ø BST ⇔ BTM ⇔ BTS	Nerre	Description		E BSMMAT01.	Description 1ST TO 5TH PLAUSIBLE VALUE NATHENATICS"
			+ + 4	(bonneso	1ST TO 5TH PLAUSIBLE VALUE SDENCE" TAATIONAL MATH RAKEN SCORE (150.10) TAATIONAL SCIENCE RASCH SCORE (150.10) SCENCE TEACHER LINK" TAATIBENATICS TEACHER LINK" TUIMBER OF MATH TEACHERS FOR STUDENT" TUIMBER OF MATH TEACHERS FOR STUDENT TUIMBER OF SOENCE TEACHERS FOR STUDENT TUIMBER OF SOENCE TEACHERS FOR STUDENT THE OREATION DATE AT OPC HAMBURG"
t Variables By Name Location in File					
6	iearch /	T Nime. C Desaption		Search (* p	iane C Description

4. Prepare&Merge -- Select FileType 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)

AT North					
Select Paraneters Name Description	Select any say file in th	REAL PROPERTY.	+ © d 🖽 -	2 88	
	My Process My Process Decision Decision	rgel, sav rge2, sav rge3, sav			
	My Documents My Computer				_
	Ny Documents	: hurnege zav	× 00	m	_
	My Documents My Corputer	in the second second second second	× Do • Can		
	My Decarvents My Computer	per SPSS Data Files (*, sav)	10000		

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

Select Analysis T C Backgrour IF None			F Daysfeed	110 F Exclude Mooings From Anaboli
Select Parameters				
Nane	Description	C Departer Va	and the second se	
💕 ildontay	"COUNTRY ID"		Norm	Description
ダ idschool	"SCHOOL ID"			
💕 idolasis	"CLASS ID"	4		
ダ idekud	"STUDENT ID"			
🥑 idpop	"POPULATION ID"	C Internation	Same .	
🦸 idgrader	"STANDARDIZED GRADE ID"	C. C. Statementsky, A.	a antinez	1
and the second	"S TUDENT WEIGHT"	E I	Name	Description
a paore	"ZONE CODE"			
Фанр	"REPLICATE CODE"	4		
🦸 intrai03	'IN 2003 ASSESSMENT POPULATION'			
ESMMATOT	1ST TO 5TH PLAUSIBLE VALUE MATHEMATICS"			
ESSSCI01-05	1ST TO 5TH PLAUSIBLE VALUE SCIENCE*	Charitarian V	wate	
Stenned	"NATIONAL MATH PASCH SCORE (150,10)"		Name	Description
Stermo	"NATIONAL SCIENCE PASCH SCORE (150,10)"	P	100000	0.000000
d ideach	"TEACHER ID"			
ダ idirik	"TEACHER LINK NUMBER"	4		
d acouse	"NAME OF COURSE"			
🤞 eciculti	"SCIENCE TEACHER LINK"	A Maide Variability	÷	
Smalaubi	"MATHEMATICS TEACHER LINK"		Nene	Description
Sinnteach	"NUMBER OF MATH TEACHERS FOR STUDENT"	+ 1	1000	
Sinteach	"NUMBER OF SCIENCE TEACHERS FOR STUDENT"			
Sinleach	'NO. OF MATH+SDIENCE TEACHERS CONBINED*	4		
aching!	WEIGHT FOR SCIENCE TEACHER DATA'	×	1	
<		Jackkniling Vari	able	
			Name	Description
Search			179010	
In gr	None C Description			

6. Analyze -- Select analysis type and variables

- xii. Select the type of Analysis. The following analysis procedures are available:
 - Background Means: Calculating means for other variables than plausible values
 - PV Means: Estimating mean achievement scores for groups
 - Background Regression: Computing regressions between dependent and independent variables
 - Achievement Regression: Computing regressions between plausible values and dependent variables
 - For the last one, it is also possible to let the IDB analyzer calculate dummy variables for the independent variables that will be used in the achievement regression. This option needs a lot of computer power. Please assure that you have a fast computer running the program or limit your analysis to a small number of countries in the dataset to be analyzed. 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.
- xiii. The next step is to define all necessary parameters for the analysis (dependent variable/ independent variables/ classification

variables). Weights and Jacknife 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.

set Foreit -+b								
<u> * ~ 또미 : 영역</u> 의								
The matro uses spec	olumn 3. Testi VILLIL Mind of the DEFISE com Martin, This instance a	end Buyli III inte p	Called the name	of a				
ie)								
Report								
п								
ACCOMPTON TO A	TUDESTS*	2	TOTACI	ESPY	RADATINE	207	1-CT_RE	
AWSTRALIA	GEN.	2465	LS.coo	517.0) 847.48	4. 64 4. 65	53.15	2.25	
Baffa Di	613L BOZ	2025 2179	5317 5226	452.01 423.95	2,66 2,27	82.48 45.57	.44 .44	
é				_				

7. Analyze – Output of PV means

In this example the achievement mean for the subgroup gender is reported for two countries (Australia and Bahrein). 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 mathemnatics 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.