IEA Civic Education Study User Guide for the International Database

Heiko Sibberns

IEA Civic Education Study User Guide for the International Database

Heiko Sibberns

The International Association for the
Evaluation of Educational Achievement

CONTENTS

The International Association for the Evaluation of Educational Achievement, known as IEA, is an independent, international consortium of national research institutions and governmental research agencies, with headquarters in Amsterdam. Its primary purpose is to conduct large-scale comparative studies of educational achievement with the aim of gaining more in-depth understanding of the effects of policies and practices within and across systems of education.

1

Content and Format of Database Files

1.1 INTRODUCTION

This chapter describes the content and format of the IEA Civic Education Study Standard Population and Older Population data files. It is organized in three major sections corresponding to the types of files included in the database. Each section describes the contents of the files. The file types are:

Data Files

The IEA Civic Education Study data files reflect the result of an extensive series of data management and quality control steps taken to ensure that the international comparability, quality, accuracy, and general utility of the database provide a strong foundation for secondary analyses. As part of the international data files, all variables derived for reporting in the international reports are included.

Codebook Files

These specifically document the format of all variables in each of the data files.

Program Files

Several program files are provided for use in secondary analyses. The macro programs for computing statistics using the jackknife repeated replication method are discussed in Chapter 2.

In addition to this user guide, the *IEA Civic Education Study Technical Report* (2004) should be available when analyzing the data. The technical report provides information about all technical aspects of the study, with the sampling, weighting, scaling, scoring and reporting procedures presented in particular detail. The technical report can be downloaded from the following site:
http://www.iea-dpc.org/download/ieahq/studies_datasets/CIVED_datasets/CIVED_TechReport.pdf

1.2 DATA FILES

The IEA Civic Education Study international database contains two basic types of data files:
• Background Files
• Student-Teacher Linkage Files.

These files and the variables contained in each are described below. Data files are provided for each country that participated in the IEA Civic Education Study and for which internationally comparable data are available. The two different file types reflect data collected from students, teachers, and schools, as well as student-teacher linkage information.

1.2.1 Data File and Variable Naming Convention

Before describing the different data files, it is necessary to describe the file and variable naming conventions used in the IEA Civic Education Study database. This part of the guide also describes how the variables are organized.

1.2.1.1 Data File Naming Convention

The data files included in this database are named using the following convention.

- The first character indicates the IEA Civic Education Study population:
  - B   indicates the Standard Population
  - C   indicates the Older Population

- The second character indicates the source or level of information in the file:
  - C   indicates a school-level file
  - T   indicates a teacher-level file
  - L   indicates a linkage file
  - S   indicates a student-level file.

- The next three digits identify the country, with the three digit alphanumeric country abbreviation following the ISO coding scheme.

- The seventh and eighth characters indicate the version:
  - F2   stands for final version 2.

1.2.1.1.1 File Extension Definitions

The three-character file extensions used for the files are listed below in alphabetic order:

.SAV   SPSS file

The data files for each country are identified by a three-digit alphabetic country abbreviation designated as <country> in the general file names shown in Table 1.1.

Table 1.1: IEA Civic Education Study data files

File Type	Standard Population	Older Population
Student File	BS<Country>F2	CS<Country>F2
Teacher Background File	BT<Country>F2	NA
School Background File	BC<Country>F2	NA
Student-Teacher Linkage File	BL<Country>F2	NA

The three-digit abbreviations used for each IEA Civic Education Study country and the available files are listed in Table 1.2 along with the numerical code values used in the country identification variable contained in the background data files (see the following section, which discusses identification variables).

Table 1.2: Country identification and availability of data files

Country	ISO Code	Numerical Code	BS	BT	BL	BC	CS
Australia	AUS	036	•	•	•	•	
Belgium (Fr)	BFR	957	•	•	•	•	
Bulgaria	BGR	100	•	•	•	•	
Chile	CHL	152	•	•	•	•	•
Colombia	COL	170	•			•	(•)
Cyprus	CYP	196	•	•	•	•	•
Czech Republic	CZE	200	•	•	•	•	•
Denmark	DNK	208	•	•	•	•	•
England	ENG	826	•	•	•	•	
Estonia	EST	233	•	•	•	•	•
Finland	FIN	246	•	•	•	•	
Germany	DEU	280	•	•	•	•	
Greece	GRC	300	•	•	•	•	
Hong Kong (SAR)	HKG	344	•	•	•	•	(•)
Hungary	HUN	348	•	•	•	•	
Israel	ISR	376				•	•
Italy	ITA	380	•	•	•	•	
Latvia	LVA	428	•	•	•	•	•
Lithuania	LTU	440	•	•	•	•	
Norway	NOR	578	•	•	•	•	•
Poland	POL	616	•	•	•	•	•
Portugal	PRT	620	•	•	•	•	•
Romania	ROM	642	•	•	•	•	
Russian Federation	RUS	643	•	•	•	•	•
Slovak Republic	SVK	201	•	•	•	•	
Slovenia	SVN	890	•	•	•	•	•
Sweden	SWE	752	•	•	•	•	•
Switzerland	CHE	756	•	•	•	•	•
United States	USA	840	•	•	•	•	

All IEA Civic Education Study international data files are provided in SPSS format, enabling users to access the data directly without having to go through intermediate steps. All details of the file structure are provided in codebook files related to each of the data files listed in Table 1.1. The use of these codebooks is described later in this chapter.

1.2.1.2 Variable Naming Convention

International background variables obtained from the student, teacher, and school questionnaires are provided in the corresponding files. In general, the background variables are provided for all countries where the data are considered internationally comparable. The assessment of international comparability for background variables was based on information regarding any national adaptations of the background questionnaire items. This information was provided by the IEA Civic Education Study National Research Coordinators (NRCs). In a few cases, some slightly modified specific country options were retained in the international variables.

For a description of the information obtained from the international student, teacher, and school questionnaire items, see the *IEA Civic Education Study Technical Report* (2004). The technical report also contains copies of the international versions of the questionnaires.

The values for the variables are either categorical options or open-ended numerical values, in accordance with the corresponding questionnaire item formats. The codebook files contain the international variable names, descriptive labels, response code definitions, formats, and field locations corresponding to each questionnaire item.

In addition to the variables contained in the questionnaires, a number of derived variables were computed for use in the international reports. These derived variables, many of which use data from more than one source, are also included in the international database for use in secondary analyses. All derived variables are self-explanatory. The information from the codebook can also be used.

The naming system for the variables permits the determination of the population and source based on 7- or 8-digit codes. These are the variables that capture the responses given to the survey instruments. The following convention is followed in naming these variables:

- The first character indicates the population. "B" is used to indicate the IEA Civic Education Study Standard Population. "C" is used to indicate the IEA Civic Education Study Older Population.
- The second character indicates the type of respondent. The following abbreviations are used:
  C   School Principal
  T   Teacher
  S   Student
- The last characters indicate a question name and are unique to each variable.

1.2.1.2.1 Identifying Background Variables by Questionnaire Numbers

The international variables are listed in the codebooks in order of the corresponding questions in the international version of the questionnaires. For each variable, the corresponding international questionnaire location is given. The questionnaire item numbers associated with each variable are indicated by field locations according to the formats given in Table 1.3.

Table 1.3: Background questionnaire item field location format conventions

Questionnaire	Location
Student Questionnaire, Standard Population	SQ2-***
Student Questionnaire, Older Population	SQ3-***
Teacher Questionnaire	TQ2-***
School Questionnaire	SCQ2-***

1.2.1.3 Identification Variables

In all files, several identification variables are included that provide information used to identify students, teachers, or schools, and to link cases between the different data files. Most identification variables carry the prefix "ID". They are listed below.

1.2.1.3.1 Identification Variables Included in Student, Teacher, and School Files

IDCNTRY

Three-digit country or jurisdiction identification code. This variable should always be used as one of the link variables whenever files are linked within and across countries.

COUNTRY

Three-digit alphanumeric country or jurisdiction identification code.

IDPOP

Identifies the population.

IDSTRAT

Identification number that identifies the stratum to which the corresponding respondent belongs.

IDSCHOOL

Identification number that uniquely identifies each school within each country. The codes for the school are not unique across countries. Schools across countries can be uniquely identified only by the IDCNTRY and IDSCHOOL combination.

1.2.1.3.2 Additional Identification Variables Included in the Student Files

IDSTUD

Identification number that uniquely identifies each student in the country sampled. The variable IDSTUD is a hierarchical identification number. It is formed by combining the variables IDSCHOOL and IDCLASS, followed by a two-digit sequential number within each classroom. Students can be uniquely identified in the database by the combination of IDSTUD and IDCNTRY.

IDCLASS

Identification number that uniquely identifies the sampled class within the school.

IDGRADE

Indicates the actual grade denomination within the country.

1.2.1.3.3 Additional Identification Variables Included in the Teacher Files

IDTEACH

Identification number that uniquely identifies the selected teacher within the school. This hierarchical identification number is formed by combining IDSCHOOL and a two-digit sequential number within each school. This variable is unique to each teacher within each country but is not unique in the teacher file.

IDLINK

This variable uniquely identifies the class for which the teacher answered the questionnaire. The combination of variables IDCNTRY, IDTEACH, and IDLINK uniquely identifies a teacher-class combination in the database.

In the student file, the IDSTUD variable provides a unique identification number to identify each student within each country. Since teachers may teach more than one class, the IDTEACH and IDLINK combinations in the teacher background files provide a unique identification for each teacher teaching a specific class. Teacher background variables are linked to appropriate students using the student-teacher linkage file. The variable IDSCHOOL, contained in all three background files, is a unique identification number for each school within a country that can be used to link school background data to corresponding students or teachers.

1.2.1.3.4 Linking and Tracking Variables

Information about students, teachers, and schools provided on the survey tracking forms is included in linking or tracking variables. These variables have the prefixes "IL" or "IT". Some of the important linking and tracking variables are listed below.

1.2.1.3.5 Linking and Tracking Variables Included in the Student Background Files

ITSEX

Gender of each student.

ITBIRTHM AND ITBIRTHY

Month and year of birth of each student.

ITDATEM AND ITDATEY

Month and year of testing of each student.

ITLANG

Language of testing for each student.

ITPART, ITPART1, ITPART2

Participation status variable indicating whether each student participated in any part (ITPART), in Part 1 and 2 (ITPART1), or in Part 3 (ITPART2).

ILTEACH1 – ILTEACH3

Identification number that identifies the teachers teaching the student under consideration. This hierarchical identification number is formed by combining IDSCHOOL and a two-digit sequential number within each school. This variable is unique to each teacher within each country but is not unique in a student record. Identical ILTEACHx variables in a student record indicate different subjects taught by the same person. Teaching of different subjects is shown by identical ILTEACHx figures accompanied by different ILLINKx numbers.

ILLINK1 – ILLINK3

Within a student record, these variables identify the class or subject a teacher is teaching.

1.2.1.3.6 Linking and Tracking Variables Included in the Teacher Background Files

ITSUBJCT

The subject(s) taught by the teacher according to tracking form information. Since codes vary by country and educational systems, the figures are not comparable across countries.

1.2.1.3.7 Linking and Tracking Variables in the School Background Files

ITPARTC

Participation status variable indicating whether each school participated and returned a questionnaire.

1.2.1.3.8 Scores

Several scores are included in the student file as well as in the student-teacher linkage file. In broad terms, scores pertaining to cognitive and attitude scales have the suffix "mle" to indicate that they are maximum likelihood estimates. For a description of the scaling and scoring processes applied in the IEA Civic Education Study, see the *IEA Civic Education Study Technical Report* (2004).

The following scores are included in the data:

KNOWLMLE

Maximum likelihood estimate of the knowledge scale; internationally scaled to a mean of 100 and a standard deviation of 20.

SKILSMLE

Maximum likelihood estimate of the skills scale; internationally scaled to a mean of 100 and a standard deviation of 20.

TOTCGM LE

Maximum likelihood estimate of the combined cognitive scale; internationally scaled to a mean of 100 and a standard deviation of 20.

CTCONMLE

Maximum likelihood estimate of the attitude scale IMPORTANCE OF CONVENTIONAL CITIZENSHIP; internationally scaled to a mean of 10 and a standard deviation of 2.

CTSOCMLE

Maximum likelihood estimate of the attitude scale IMPORTANCE OF SOCIAL-MOVEMENT-RELATED CITIZENSHIP; internationally scaled to a mean of 10 and a standard deviation of 2.

GOVECMLE

Maximum likelihood estimate of the attitude scale ECONOMY-RELATED GOVERNMENT RESPONSIBILITIES; internationally scaled to a mean of 10 and a standard deviation of 2.

GOVSOMLE

Maximum likelihood estimate of the attitude scale SOCIETY-RELATED GOVERNMENT RESPONSIBILITIES; internationally scaled to a mean of 10 and a standard deviation of 2.

TRUSTMLE

Maximum likelihood estimate of the attitude scale TRUST IN GOVERNMENT-RELATED INSTITUTIONS; internationally scaled to a mean of 10 and a standard deviation of 2.

PATRIMLE

Maximum likelihood estimate of the attitude scale POSITIVE ATTITUDES TOWARD ONE'S NATION; internationally scaled to a mean of 10 and a standard deviation of 2.

WOMRTMLE

Maximum likelihood estimate of the attitude scale ATTITUDES TOWARD WOMEN'S POLITICAL AND ECONOMIC RIGHTS; internationally scaled to a mean of 10 and a standard deviation of 2.

IMMIGMLE

Maximum likelihood estimate of the attitude scale POSITIVE ATTITUDES TOWARD IMMIGRANTS; internationally scaled to a mean of 10 and a standard deviation of 2.

CONFSMLE

Maximum likelihood estimate of the attitude scale CONFIDENCE IN PARTICIPATING AT SCHOOL; internationally scaled to a mean of 10 and a standard deviation of 2.

POLATMLE

Maximum likelihood estimate of the attitude scale POLITICAL ACTIVITIES; internationally scaled to a mean of 10 and a standard deviation of 2.

CCLIMMLE

Maximum likelihood estimate of the attitude scale OPEN CLIMATE FOR CLASSROOM DISCUSSION; internationally scaled to a mean of 10 and a standard deviation of 2.

1.2.1.4 Weights

Weighting variables are included in the student, teacher, and school files. These variables must be used in order to get unbiased analyses results. For a description of the weighting processes applied in the IEA Civic Education Study, see the *IEA Civic Education Study Technical Report* (2004).

1.2.1.4.1 Weight Variables Included in the Student Files

The weight variables reflecting the weight components show the prefix "WGT". Actual weights show the suffix "WGT". Both groups of weight variables are listed below.

WGTFAC1

Weighting factor 1: This variable reflects the selection probability of a school at the school level.

WGTADJ1

Weight adjustment 1: This variable reflects school non-response adjustment within explicit strata at the school level.

WGTFAC2

Weighting factor 2: This variable reflects the selection probability of a class within a participating school.

WGTADJ1

Weight adjustment 2: This variable reflects class non-response adjustment within a participating school.

WGTFAC3

Weighting factor 3: This variable reflects the selection probability of a student at the class level.

WGTADJ3

Weight adjustment 3: This variable reflects student non-response adjustment within a participating class.

TOTWGT

Total weight: This variable represents the weight attached to each student. The sum of TOTWGT across all participating students is an estimate of the population size. This variable must always be used in order to obtain unbiased results.

SENWGT

Senate weight: This variable represents a weight equivalent to TOTWGT. The sum of SENWGT across all participating students within each country is 500. This variable must always be used for country comparisons if countries should contribute equally to the comparisons.

HOUWGT

House weight: This variable represents a weight equivalent to TOTWGT. The sum of HOUWGT across all participating students is equal to the number of participating students.

1.2.1.4.2 Weight Variables Included in the School File

The weight variables show the suffix "WGT".

SCHWGT

School weight: This variable is the inverse of the school's inclusion probability. It indicates the number of schools a particular school represents.

STDWGT

Sum of student weights: This variable indicates the number of students represented by a particular school.

1.2.1.4.3 Weight Variables Included in the Student-Teacher Linkage File

The following important variables required to compute appropriately weighted teacher-level data are included in the linkage files.

NTEACH

Number of teachers for each student.

TOTWGTCH

Adjusted weight for total teachers.

SENWGTCH

Adjusted senate weight for total teachers.

The TOTWGTCH or SENWGTCH variables contain teacher weights adjusted for the total number of teachers for each student. For example, if a student has three teachers, the adjusted weight for each teacher (TOTWGTCH) will be equal to one-third so that each teacher contributes equally to the teacher-based data for students. These variables must always be used in order to obtain unbiased results. However, the user must keep in mind that the IEA Civic Education Study data represent a sample of students, not a sample of teachers.

1.2.1.5 Jackknife Variables

Two variables needed for variance estimation are included in the student files. These variables must be used in order to obtain correct estimates for the sampling error variances. For a description of the processes applied to create the jackknife variables in the IEA Civic Education Study, see the *IEA Civic Education Study Technical Report* (2004).

JKZONE

The sampling zone to which the student's school is assigned.

JKREP

The primary sampling unit to which the student is assigned.

1.2.1.5.1 Identification Variables Included in Student, Teacher, and School Files

IDCNTRY

Three-digit country or jurisdiction identification code. This variable should always be used as one of the link variables whenever files are linked within and across countries.

1.2.2 Student, Teacher, and School Files

There are three different types of IEA Civic Education Study files: student, teacher, and school.

1.2.2.1 Student File

Students who participated in the IEA Civic Education Study were administered a questionnaire containing test items, questions related to home background and school experiences, and items relating to students' concepts, attitudes, and actions. The student file contains students' responses to these questions.

1.2.2.2 Teacher File

The civic education teachers of the students who were sampled for the IEA Civic Education Study were administered a questionnaire containing questions pertaining to their background and their teaching practices in the classes of the sampled students. Each teacher was asked to respond to a questionnaire for each class taught that contained sampled students. The teacher file contains one record for each of the teachers who completed a questionnaire.

In the teacher background data files, each teacher was assigned a unique identification number (IDTEACH) and a teacher link number (IDLINK) specific to the class taught by the teacher and to which the information in the data record corresponds. The IDTEACH and IDLINK combination uniquely identifies a teacher teaching one specific class. So, for example, students linked to teachers identified by the same IDTEACH but different IDLINK would have been taught by the same teacher but in different classes or subjects.

The teacher background data files cannot be merged directly with the student data files, and they do not contain sampling information or achievement scores. It is important to note that the teacher background data files do not constitute a representative sample of teachers in a country, but rather consist of the teachers teaching a representative sample of students. The teacher data should therefore be analyzed only in conjunction with the weights taken from the student-teacher linkage file. The teacher background data files contain a series of other identification and link variables, as well as the responses of the teachers to the background questions.

1.2.2.3 School File

The principals or administrators of the schools in the IEA Civic Education Study sample were administered a school background questionnaire containing questions about school policy and school environment. The school background data file contains the responses given to the questions in this questionnaire. This file also contains a series of identification variables, link variables, and sampling variables. The school data files can be merged with the student data files by using the country and school identification variables.

1.2.3 Student-Teacher Linkage Files

The student-teacher linkage files for the IEA Civic Education Study contain information required to link the student and teacher files and to compute appropriately weighted teacher-level data using the student as the unit of analysis.

The student-teacher linkage files contain one entry per student-teacher linkage combination in the data. In many cases, students are linked to more than one teacher. In these cases, there is one record for each student-teacher link combination. For instance, where three teachers are linked to a student, three entries in the file will correspond to that student.

The linkage files also contain the identification variables required to identify cases and to link the student and teacher files. Some tracking and achievement score variables plus all teacher variables are also included. The codebooks for the student-teacher linkage files contain a complete list of all variables included.

1.2.4 Missing Codes in the International Data Files

All values assigned to variables in the IEA Civic Education Study international data files are numerical. A subset of the numerical values for each of the variable types is reserved for specific codes related to different categories of missing data. The missing categories defined below are assigned to different values depending on the field width of the variable and the variable type.

1.2.4.1 Omitted Response Codes (System Missing, …)

Omitted response codes are used for questions/items that a student, teacher, or school principal should have answered but did not. These are indicated as "missing" in the codebooks. For questionnaire data, no differentiation has been made between "no answer" and "invalid" answers. The latter include instances of checking two or more response options in a categorical question or responses to open-ended questions that were unreadable or could not be interpreted. In a few cases, data received from a country in an invalid or inconsistent way are also recoded as "missing". For cognitive items, an omitted response code is given only in cases in which the item was left blank; a special code is used for invalid answers as described below. The specific omitted response code value given depends on the number of valid codes available for each item.

For identification, tracking, or questionnaire items: An omitted response code value of 9 is used for categorical items with seven or fewer valid response options. For categorical items with more than seven categories, a code value of 99 is used. For open-ended background questionnaire items or other items containing non-categorical values, the omitted code is the next available 9* code greater than the highest response in the defined valid range. Background questionnaire item values outside the valid ranges are recoded as missing.

For cognitive items: An omitted response code value (SYSMIS) is used for multiple-choice cognitive items.

1.2.4.2 "Not Interpretable" Response Codes (7)
[Standard Population only]

For the cognitive items, separate codes distinguish between totally blank responses (omitted/missing) and not interpretable or invalid responses. For multiple-choice items, cases where more than one response option was checked are classified as not interpretable and given a code of 7.

1.2.4.3 Not Administered Codes (8, 98, 998, …)

Special codes are given for items that were "not administered" to distinguish these cases from data missing due to non-response. The specific not administered code value given depends on the number of valid codes available for each item, as described above for the omitted response codes.

There are two general cases when the not administered code is used.

1. Data not collected for a variable for specific individuals. Reasons for this include:

   Booklet not assigned to the student: All variables corresponding to items not given to a student are coded "Not administered".

   Booklet not received/booklet lost: Where a respondent did not receive the instruments assigned to him/her, or the instruments were lost after administration, the items are coded "Not administered".

   Student absent from session: Where a student was not present for a particular testing session, all variables referring to that session are coded "Not administered". However, if a student participated in a session but did not answer any of the items, the items are coded "Omit".

   Item left out or misprinted: Where a particular question or item (or a whole page) was misprinted or not available to the student, teacher, or school, the corresponding variables are coded "Not administered".

2. Item omitted for all cases in a country. All cases are coded "Not administered".

   Cognitive items omitted or mistranslated in student test: Any items identified during the translation verification or item analysis processes that were mistranslated such that the nature of the question was altered were removed for a country.

   Background items omitted: Questions in the student, teacher, or school questionnaires that were considered not applicable in some countries were not included in their questionnaires.

   Questionnaire items mistranslated or not internationally comparable: In some cases, questions in the international version of the questionnaires were mistranslated or modified to fit the national situation. Whenever possible, modified background questionnaire items were recoded to match as closely as possible the international version of the items. This could not be done in all cases, however, and some national data were recoded "Not administered" in order to include only the internationally comparable data.

1.2.4.4 Not Reached Item Codes (6)

These codes are used only for cognitive items. Test items at the end of the testing session that were left blank were considered "Not reached" because the student did not complete the test. These responses are distinguished from the "Missing" responses, in that they are handled differently during the item calibration process; see the *IEA Civic Education Study Technical Report* (2004). They are treated as incorrect responses, however, in computing achievement scores. For the multiple-choice items, a not reached item code value of 6 is used.

1.2.5 National Data Issues Affecting the Use of International Data Files

In some cases, resources were not available to resolve database issues for specific countries in time for either the release of the international reports or the production of the international data files. As a result, some international data have been modified or are not available for some countries. These general database issues are documented here.

Colombia

> The student and school data for the Older Population submitted by Colombia were not deemed internationally comparable and so are not weighted. Although these data are included in the international database, the results cannot be compared to the data from other countries. Teacher data for the Standard Population are not included because no linkage to student data could be established.

Hong Kong (SAR)

> The student and school data for the Older Population submitted by Hong Kong (SAR) were not deemed internationally comparable and are thus not weighted. They are included in the international database. However, the results for Hong Kong (SAR) cannot be compared to the data from other countries.

1.3 CODEBOOK FILES

All information related to the structure of the data files as well as the source, format, descriptive labels, and response option codes for all variables are contained in codebook files. One codebook file is provided for each of the data files listed earlier.

The naming convention of the codebook files is as follows:

- The first character of the files is either "B" or "C". This indicates that the file refers to the IEA Civic Education Study's Standard Population or Older Population.

- The second character indicates the source or level of the information in the file:

    C   indicates a school-level file

    T   indicates a teacher-level file

L   indicates a student-teacher linkage file

S   indicates a student-level file.

1.3.1 File Extension Definitions

The three-character file extensions used for the codebook files are listed below in alphabetic order:

.TXT   Codebook in ASCII text format

.PDF   Printable pdf format.

1.3.2 Using the Codebooks

The variables in the codebooks appear in order by variable name within the section for each codebook type. The major sections of each codebook type are as follows:

Student, Teacher, and School Codebooks

- Identification Variables
- Tracking/Linking Variables
- International Background Variables (in order of questionnaire item location)
- Derived Variables
- Sampling Variables (Student and School Files only)
- Score Variables (Student Files only).

Student-Teacher Linkage File Codebooks

- Identification Variables
- Sampling Variables
- Score Variables
- Teacher Linking/Weighting Variables
- International Teacher Variables (in order of questionnaire item location).

Figure 1.1 provides an example printout of a page from the codebook for the student data. The fields are as follows:

Variable number: The first column (Var. No.) contains a sequential number for each variable in each codebook file.

Question: The second column contains an abbreviated variable identifier providing descriptive information needed to identify the content of the question and/or the source for each type of variable.

Variable name: The third column (Variable Name) contains the variable name associated with each variable included in the international data files. The naming system used for each variable type is described in the previous sections on the contents of data files.

Variable label: The fourth column (Label) contains an extended textual variable label of up to 40 characters associated with each variable, providing more descriptive information about the content of each variable. For multiple-choice cognitive items, the variable label includes the correct response option enclosed in brackets. During data analysis, the variable labels can be printed out to enhance understanding of the results.

Code: The fifth column (Code) contains the codes used for variable responses. For variables where numerical data are supplied in response to open-ended questions, the keyword VALUE is entered in the Code column. For categorical variables, all possible response options are listed. Any missing codes described in Section 2.1.3 are also included for either numerical or categorical variables.

Option: The sixth column (Option) includes a textual description of each type of response option. For variables containing numerical data, the column contains an explanation of the values contained in the variable.

Location/Format: The seventh column presents the location and format of each variable in the raw data files. The location/format indicates the pattern used to write each value of a numerical or categorical variable, with a general structure of

XX-YY / <N or C> ZZ.Z

The numbers preceding the slash (/) indicate the location of the variable and refer to its position in the raw data file (starting (XX) – ending (YY) column positions). The <N or C> after the slash identifies the variable as numerical (N) or categorical (C). The numerical code after the slash (ZZ.Z) indicates the total number of digits (including the decimal point) and the number of decimal places associated with each variable (e.g., 2.0 = two integer digits, 0 decimal places; 6.2 = six total digits: three integer digits, decimal point, and two decimal digits).

1.4 PROGRAM FILES

Two program files are provided for use in analyses of the IEA Civic Education Study data files:

• Jackknife Statistics Program File for subgroups
• Jackknife Statistics Program File for regressions.

The Jackknife Statistics Program for subgroups is used to compute the percentage of students within defined subgroups and the mean value for each subgroup on specified continuous variables, as well as the standard errors associated with these statistics, using the jackknife repeated replication (JRR) method discussed in the *IEA Civic Education Study Technical Report* (2004).

The Jackknife Statistics Program file for regressions is used to compute the intercepts and slopes as well as the standard errors associated with these statistics, using the jackknife repeated replication (JRR) method.

Figure 1.1: Codebook structure

Var. No.	Question	Variable Name	Variable Label	Code	Option	Location/Format
1	PUNCHER	IDPUNCH	*PUNCHER*			1- 3/N 3.0
2	COUNTRY	IDCNTRY	*COUNTRY ID*			4- 6/N 3.0
3		COUNTRY	ALPHA NUMERIC COUNTRY CODE			7- 9/C 3.0
4	POP	IDPOP	*POPULATION ID*			10 /N 1.0
5	STRAT	IDSTRAT	*STRATUM ID*			11- 13/N 3.0
6	SCHOOL	IDSCHOOL	*SCHOOL ID*			14- 21/N 8.0
7	CLASS	IDCLASS	*CLASS ID*			22- 31/N 10.0
8	GRADE	IDGRADE	*GRADE ID*			32- 33/N 2.0
9	STUD	IDSTUD	*STUDENT ID*			34- 45/N 12.0
10	EXCLUD	ITEXCLUD	*INDICATOR FOR EXCLUDED STUDENTS*	1	student is functionally disabled	46 /C 1.0
				2	student is educable mentally retarded	
				3	student unable to speak or read language	
				4	<nationally defined reason>	
				•	missing	
11	BIRTH_MONT	ITBIRTHM	*STUDENT'S DATE OF BIRTHMONTH*			47- 48/N 2.0
12	BIRTH_YEAR	ITBIRTHY	*STUDENT'S DATE OF BIRTHYEAR*			49- 50/N 2.0
13	SEX	ITSEX	*STUDENT'S SEX*	1	girl	51 /C 1.0
				2	boy	
				8	not admin.	
				•	missing	

2

Performing Analyses with the IEA Civic Education Study Data: Some Examples

This chapter presents some basic examples of analyses that can be performed with the IEA Civic Education Study international database, using the sampling weights and scores discussed above. It also provides details on some SPSS syntax codes necessary to conduct such analyses, and the results of these analyses. Although the analyses presented here are simple in nature, they are designed primarily to familiarize the user with the different files and their structure, as well as the relevant variables that need to be included in most analyses.

The analyses presented compute the percentage of students in specified subgroups, the mean achievement for each subgroup, and the corresponding standard errors for the percent and mean statistics. These analyses, based on student data, replicate analyses that are included in the IEA Civic Education Study international report (J. Torney-Purta, R. Lehmann, H. Oswald, & W. Schulz (2001), *Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen*, Delft, IEA). Table 2.1 displays the results which should be reproduced. The user is invited to compare the results from these analyses to the tables in the reports, and is encouraged to practice analyzing the IEA Civic Education Study data by trying to replicate some of the tables that are presented in the international reports. Two different ways to obtain the results are presented. The first method uses the Jackknife Statistics Program File for subgroups; the second uses the Jackknife Statistics Program File for regressions. While both examples lead to the same result, they use different statistical methods.

The examples given here use macros written for SPSS that can be used to perform any of the analyses described below. These general routines can be used for many purposes, provided the user has some basic knowledge of the SPSS macro language.

2.1 PROGRAM FILES

JACKGEN.SPS

The macro program can be used to compute the weighted percentage of students within defined groups, and their mean on a specified continuous variable. The macro also generates replicate weights and computes JRR sampling variance for the percent and mean estimates. Although, in general, the continuous variable of choice will be one of the scores, this continuous variable can be any other continuous variable in the file. A description of how to use the macro program is given later in this chapter.

JACKREG.SPS

This macro program can be used to compute weighted regression coefficients within defined groups. It can also be used to compute descriptive statistics on the variables. The macro can be used with any variable in the analysis.

SAMPLEJACKREG.SPS AND SAMPLEJACKGEN.SPS

These are the programs used in the examples presented later in this chapter. They are included in the IEA Civic Education Study database. Macros JACKREG.SPS or JACKREG.SPS must be included in these programs in order to obtain results.

Table 2.1: Sample table for student-level analysis taken from the IEA Civic Education Study international report *

Country	Mean Female	Mean Male	Difference
Denmark	99 (0,7)	102 (0,7)	3 (1,0)
Switzerland	97 (0,8)	100 (0,9)	2 (1,2)
Chile	88 (0,8)	89 (0,8)	2 (1,1)
Czech Republic	102 (0,8)	104 (1,0)	2 (1,3)
Portugal	96 (0,8)	97 (0,9)	1 (1,2)
Germany	99 (0,6)	101 (0,7)	1 (0,9)
Norway	103 (0,6)	103 (0,7)	1 (0,9)
Russia	99 (1,2)	100 (1,7)	0 (2,1)
Slovak Republic	105 (0,8)	105 (0,9)	0 (1,1)
England	100 (0,8)	100 (1,0)	0 (1,3)
Cyprus	108 (0,7)	108 (0,6)	0 (0,9)
Colombia	87 (1,3)	86 (1,1)	0 (1,7)
Romania	92 (1,0)	91 (0,9)	0 (1,4)
Hungary	102 (1,3)	101 (1,3)	1 (1,9)
Hong Kong (SAR)	107 (1,3)	106 (1,4)	1 (1,9)
Sweden	100 (0,8)	99 (1,1)	1 (1,3)
Estonia	95 (0,6)	93 (0,7)	1 (0,9)
Finland	110 (0,9)	108 (0,8)	2 (1,2)
USA	107 (1,2)	106 (1,3)	2 (1,8)
Greece	109 (0,8)	107 (0,9)	2 (1,2)
Italy	106 (0,9)	104 (1,1)	2 (1,4)
Bulgaria	99 (1,5)	97 (1,2)	2 (2,0)
Lithuania	95 (0,8)	92 (0,8)	2 (1,1)
Australia	103 (0,9)	101 (1,1)	2 (1,4)
Poland	112 (2,2)	109 (1,5)	3 (2,6)
Slovenia	102 (0,6)	99 (0,6)	4 (0,8)
Latvia	93 (0,9)	90 (0,9)	4 (1,3)
Belgium (Fr.)	97 (1,1)	93 (1,3)	5 (1,7)

Note: ***** J. Torney-Purta, R. Lehmann, H. Oswald, & W. Schulz (2001), *Citizenship and Education in Twenty-eight Countries: Civic Knowledge and Engagement at Age Fourteen*, Delft, IEA.

2.2   COMPUTING SAMPLING VARIANCE USING JACK GEN.SPS

This code is provided in the form of a SPSS macro. This macro computes the percentage of students within a subgroup defined by a set of classification variables, the JRR standard error of this percentage, the mean for the group on one of the scales, and the standard error of that mean calculating the correct sampling variance component.

The macro operates as follows:

1.   Computes a set of replicate weights specified by the parameter NJKZ for each record using the procedure described in chapter 10 of the *IEA Civic Education Study Technical Report* (2004).

2.   Aggregates or summarizes the file by computing the sum of the weights for each category, the sum of the weights overall, and the sum of the weighted score.

3. Computes the percentage of people within each group, their mean on the score, and their corresponding standard errors. In SPSS, the resulting working file contains the corresponding statistics. Here, the macro also aggregates or summarizes the file by computing the sum of the weights for each category, the sum of the weights overall, and the sum of a weighted analysis variable.

When using the macro, the user needs to specify a set of classification variables, the name of the score, the number of replicate weights to be generated, the variable that contains the sampling information, such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. In SPSS, the macro uses the current working file, which is the file that was last read.

The user needs to know some basic SPSS macro language in order to use the macro. The macro needs to be first included in the program file where it is going to be used. If the user is operating in batch mode, then the macro needs to be called in every batch. If the user is using SPSS interactively, then the macro needs to be called once at the beginning of the session, and it will then remain active throughout the session. If the session is terminated and restarted at a later time, the macro needs to be called once again. Once the macro is included in a specific session, the word "JACKGEN" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include "{directory_location}jackgen.sps".
```

where {directory_location} points to the specific drive and directory where the macro JACKGEN.SPS can be found. The macro requires that several arguments be submitted when it is invoked. These parameters are:

WGT

The sampling weight to be used in the analysis. This generally is TOTWGT when using the student files, or TOTWGTCH when using the student-teacher linkage file.

JKZ

The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all IEA Civic Education Study files is JKZONE.

JKR

The variable that captures whether the case is to be dropped or to have its weight doubled for the corresponding replicate weight. The name of this variable in all IEA Civic Education Study files is JKREP.

NJKZ

This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75 for the student, school, and teacher background data. The user working with the data for only one country should set the NJKZ argument to the

number of replicates that were used in the country under consideration (see Table 10.1 of the *IEA Civic Education Study Technical Report*, 2004, for the maximum number of replicates by country). If the data from two or more countries are being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter at, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if the user specifies fewer replicate weights than necessary.

CVAR

This lists the variables that are to be used to classify the students in the data file. It can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to the user at the time. It is recommended that the user always includes the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.

DVAR

This is the variable for which means are to be computed. Only one variable has to be listed here. If the user wants to examine, for example, results for different scales, then the macro needs to be invoked separately to generate each table. Although, in most cases, the continuous variable of interest will be an achievement variable, this can actually be any other continuous variable.

The JACKGEN macro in SPSS does not require the data file containing the data of interest to be specified when calling the macro. By default, SPSS uses the current working file. This needs to be read with the GET FILE command prior to invoking the macro.

The simplest and most straightforward way is to invoke the macro using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked as

```
get file = "{directory_location}BSAUSF2.SAV".
jackgen cvar = IDCNTRY BSGGEND /
dvar = TOTCGMLE/
jkz = JKZONE /
jkr = JKREP /
njkz = 75    /
WGT = TOTWGT  / .
```

where {directory_location} points to the specific drive and directory to where the file BSAUSF2.SAV can be found, it will compute the Australian Standard Population mean total cognitive score, and its corresponding standard error, for boys and girls, using the variable TOTWGT as the sampling weight. It will also compute the percentage of boys and girls, by grade, within the country, and the corresponding standard error. The data are read from the system file BSAUSF2.

The file that contains these results is then called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In the example above, the resulting system file would have two variables—IDCNTRY and BSGGEND—and one unique occurrence for each combination of the categories for these variables.

Weight Variable

Contains the estimate in the population that belongs to the group defined by the specific combination of the classification variable categories.

MNX

Contains the weighted mean of the variable DVAR for the group defined by the corresponding combination of classification variable categories.

MNX_SE

Contains the standard error of MNX computed using the JRR method for computing the standard error.

PCT

Contains the weighted percentage of respondents in the group defined by the last classification variable.

PCT_SE

Contains the standard error of PCT computed using the JRR method for computing the standard error.

The file resulting from using this macro can then be printed out using a SPSS procedure of choice. An example call to this macro and a subset of the resulting file are presented in Figure 2.1. In this example, the macro will compute the percentage of boys and girls by country, and their mean achievement.

The listing presented in Figure 2.1 is interpreted in the following way. The first line shows the results for the students with IDCNTRY=36 (Australia), and who had BSGGEND=1 (Girls). It is estimated that there are 121,022 Standard Population girls in Australia, and that their mean total score is 102.9, with a standard error of 0.9. It can also be ascertained from this line of data that it is estimated that 54.5 percent of the Standard Population in Australia are girls, and that the standard error of this percentage is 1.8. The second line shows the same information, but this time for the Standard Population boys (BSGGEND=2). Here, it is estimated that there are 100,880 Standard Population boys in Australia, and that their mean mathematics score is 100.5, with a standard error of 1.1. This line of data also shows that it is estimated that 45.5 percent of the seventh graders in Australia are boys, and that the standard error of this percentage is 1.8.

Figure 2.1: SPSS control code and extract of output file for using the macro JACKGEN.SPS

```
get file = "z:\civic\cd\bs\data\bsausf2.sav"
  / keep=idcntry bsggend  jkrep jkzone totwgt
       totcgmle.

select if not missing(bsggend).
save outfile = student.

include "z:\civic\cd\prg\jackgen.sps".

jackgen
    infile = student       /
    cvar   = idcntry bsggend /
    dvar   = totcgmle        /
    njkz   = 75            /
    jkz    = jkzone        /
    jkr    = jkrep         /
    wgt    = totwgt   .

print formats idcntry bsggend n (f6.0) totwgt (f10.0) mnx mnx_se pct pct_se (F6.2).

report format = list automatic
/ var   = idcntry bsggend n totwgt mnx mnx_se pct pct_se.
```

COUNTRY ID	GIRL OR BOY	N	TOTWGT	MNX	MNX_SE	PCT	PCT_SE
36	1	1790	121022	102.89	.85	54.54	1.75
36	2	1490	100880	100.54	1.11	45.46	1.75

2.3   COMPUTING SAMPLING VARIANCE USING JACKREG.SPS

In this section, an example of the SPSS code that can be used to compute the JRR standard errors for intercepts and slopes in a regression analysis is described. This code is provided in the form of a SPSS macro. This macro computes the intercept and slope within a subgroup defined by a set of classification variables, the JRR standard error of the intercept and slopes, the weighted R^2, the total weighted sum of squares, and the decomposition of the total into predicted sum of squares and residual sum of squares.

The macro operates as follows:

1. Computes a set of replicate weights specified by the parameter NJKR for each record using the procedure described in Chapter 10 of the *IEA Civic Education Study Technical Report* (2004).

2. Aggregates or summarizes the file computing the sum of the weights for each category, the sum of the weights overall, and the sum of the weighted score.

3. Computes the percentage of people within each group, their mean on the score, and their corresponding standard errors. In SPSS, the resulting working file contains the corresponding statistics. The macro also aggregates or summarizes the file computing the sum of the weights for

each category, the sum of the weights overall, and the sum of a weighted analysis variable.

When using the macro, the user needs to specify a set of classification variables, the name of the score, the names of the independent variables, the number of replicate weights to be generated, the variable that contains the sampling information, such as JKZONE and JKREP, and the sampling weight that is to be used for the analysis. In SPSS, the macro uses the current working file, which is the file that was last read.

The user needs to know some basic SPSS macro language in order to use the macro. The macro needs to be first included in the program file where it is going to be used. If the user is operating in batch mode, then the macro needs to be called in every batch. If the user is using SPSS interactively, then the macro needs to be called once at the beginning of the session, and it will then remain active throughout the session. If the session is terminated and restarted at a later time, the macro needs to be called once again. Once the macro is included in a specific session, the word "JACKREG" should not be used within that program because doing so will invoke the macro.

The macro is included in the program file where it will be used by issuing the following command under SPSS:

```
include "{directory_location}jackreg.sps".
```

where {directory_location} points to the specific drive and directory where the macro JACKREG.SPS can be found. The macro requires that several arguments be submitted when it is invoked. These parameters are:

WGT

The sampling weight to be used in the analysis. This is generally TOTWGT when using the student files, or TOTWGTCH when using the student-teacher linkage file.

JKZ

The variable that captures the assignment of the student to a particular sampling zone. The name of this variable in all IEA Civic Education Study files is JKZONE.

JKI

The variable that captures whether the case is to be dropped or to have its weight doubled for the corresponding replicate weight. The name of this variable in all IEA Civic Education Study files is JKREP.

NJKZ

This indicates the number of replicate weights to be generated when computing the JRR error estimates. When conducting analyses using the data from all countries, the value of NJKZ should be set to 75 for the student, school, and teacher background data. The user working with the data for only one country should set the NJKZ argument to as many replicates as there were in the country (see Table 10.1 in the *IEA Civic Education Study Technical Report*, 2004, for the maximum number of

replicates by country). If the data from two or more countries are being used for an analysis, then the larger number of jackknife zones should be used. When in doubt about what number to set the NJKZ parameter at, it should be set to 75. The error variance will always be estimated correctly if more replicate weights than necessary are computed, but will be underestimated if the user specifies fewer replicate weights than necessary.

CVAR

This lists the variables that are to be used to classify the students in the data file. CVAR can be a single variable, or a list of variables. The maximum number of variables will depend mostly on the computer resources available to the user at the time. It is recommended that the user always includes the variable that identifies the country. At least one variable has to be specified, usually IDCNTRY.

DVAR

This indicates the dependent variable in the regression equation. Only one variable has to be listed here. If the user wants to examine, for example, results for different scales, then the macro needs to be invoked separately to generate each table. Although, in most cases, the continuous variable of interest will be an achievement variable, this can actually be any other continuous variable.

XVAR

This lists the independent variables in the regression equation. At least one variable has to be listed here. In most cases one or more continuous variables of interest will be analyzed. However, binary coded variables can also be included here.

The JACKREG macro in SPSS does not require the data file containing the data of interest to be specified when calling the macro. By default, SPSS uses the current working file. This needs to be read with the GET FILE command prior to invoking the macro.

The simplest and most straightforward way is to invoke the macro using the conventional SPSS notation for invoking macros. This involves listing the macro name followed by the corresponding list of arguments for the analysis, each separated by a slash. For example, if the macro is invoked as

```
get file = "{directory_location}BSAUSF2.SAV".
select if not(missing(bsggend)).
compute regsex = bsggend-1.
save outfile = student.
jackreg  infile  = student   /
    cvar   = idcntry /
    xvar   = regsex    /
    dvar   = totcgmle /
    njkz   = 75      /
    jkz    = jkzone   /
    jkr    = jkrep   /
    wgt    = totwgt   .
```

where {directory_location} points to the specific drive and directory where the file BSAUSF2.SAV can be found, it will compute the Australian Standard Population mean total cognitive score, and its corresponding standard error, for boys who were set as a reference group in this example (BSGGEND=0) using the variable TOTWGT as the sampling weight. The corresponding value will be reported as the slope of the regression equation. It will also compute the difference between boys and girls, expressed as an increment compared to the reference group within the country, and its corresponding standard error. To perform these analyses, the necessary dummy coding must be completed prior to invocation of the macro.

The file that contains the results is then called FINAL and is saved to the default directory being used by SPSS. The variables that are contained in this file are:

Classification Variables

Each of the classification variables is kept in the resulting file. In the example above, there would be one variable in the resulting system file, namely, IDCNTRY, and one unique occurrence for each combination of the categories for the classification variables.

Mult_RSQ

Contains the estimate of the multiple R_ obtained in the regression analyses.

SS_Total

Total weighted sum of squares by subgroups as defined in classification variables for the dependent variable under consideration.

SS_Reg

Weighted regression sum of squares by subgroups as defined in classification variables for the dependent variable under consideration.

SS_Res

Weighted residual sum of squares by subgroups as defined in classification variables for the dependent variable under consideration.

B00

Estimate of the intercept in the regression model.

B00.SE

Estimated standard error of the intercept in the regression equation, using the JRR method.

B01

Estimate of the slope in the regression model. B01 represents the slope of the first independent variable in the regression equation. The slopes of subsequent variables are labeled B0x. (x>1)

B01.SE

Estimated standard error of the slope in the regression equation, using the JRR method. B01.SE represents the standard error of the first independent variable slope in the regression equation. Standard errors of subsequent slopes are labeled B0x.SE. (x>1)

The file resulting from using this macro can then be printed out using a SPSS procedure of choice. An example call to this macro and a subset of the resulting file are presented in Figure 2.2. In this example, the macro will compute the mean achievement of boys as intercept. The slope represents the difference arising out of the comparison between the girls and the boys.

Figure 2.2: SPSS control code and extract of output file for using the macro JACKGEN.SPS

```
get file = "z:\civic\cd\bs\data\bsausj1.sav"
  / keep=idcntry bsggend  jkrep jkzone totwgt
       totcgmle.

select if not(missing(bsggend)).
compute regsex = bsggend-1.

save outfile = student.

include " z:\civic\cd\prg\jackreg.sps".

jackreg  infile  = student   /
  cvar   = idcntry  /
  xvar   = regsex    /
  dvar   = totcgmle /
  njkz   = 75       /
  jkz    = jkzone   /
  jkr    = jkrep    /
  wgt    = totwgt   .

print formats idcntry n (F6.0) Mult_RSQ (f5.3)
    SS_Total SS_Reg SS_Res (F10.0) B00 B00.SE B01 B01.SE (f6.2).

list vars = idcntry n Mult_RSQ SS_Total SS_Reg SS_Res B00 B00.SE B01 B01.SE.

List

IDCNTRY   N  MULT_RSQ  SS_TOTAL  SS_REG   SS_RES    B00  B00.SE  B01 B01.SE

    36  3280    .003    90722446  305981  90416465 102.89    .85  -2.36  1.18
```

For example, we can say that, in Australia, data were available for 3,280 cases. The coefficient of determination between gender and performance is 0.003, with boys reaching a performance level of 102.9 and girls achieving 2.4 score points less on average.