PISA 2000

Manual for the French Subsample Database

Switzerland has been one of the participating countries in the OECD Programme for International Student Assessment (PISA) 2000. PISA 2000 compared the skills of 15-year-old students in reading, mathematic, and scientific literacy. In addition, Switzerland tested a sample of students of grade 9 regardless of their age. In this document this sample is referred as to the *national sample* whereas the sample of the 15-year old students is called *international sample*. Most of these informations can be used for the French Subsample Database too. The French Subsample Database is the subsample of the french speaking cantons which was not integrated in the national sample because of differences regarding the methods of choosing the participants.

The purpose of this document is to provide all of the necessary information to analyse the data in accordance with the methodologies used to collect and process the data.[1] It does not provide detailed information regarding these methods. In addition, a list of related publications can be found in the "References" section at the end of this document.

The following sources can provide additional information about PISA:

At the national level:

- The PISA Web page (www.pisa.admin.ch): it provides descriptions about the programme, contact information and allows users to download different reports.

- The publication **Prepared for Life? Basic Competencies of Young People – A Synthesis of the National PISA 2000 Report** (Moser 2001) summarises the main findings of the PISA study for Switzerland. This publication is available in English, French, German, and Italian.

- The publication **Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen-Nationaler Bericht der Erhebung PISA 2000** (Zahner et al. 2002) presents evidence on student performance in reading, mathematical and scientific literacy. Moreover, it contains detailed analyses on knowledge and use of ICT, self-regulated learning as a cross-curriculum-competence, social origin and equal opportunities, cultural heterogeneity in school and the influence of gender on student performance. This publication is available in French and German.

[1] The document is an adaptation of the "Manual for the PISA 2000 database" published by the OECD. This manual can be downloaded from http://www.oecd.org/pdf/M00036000/M00036694.pdf . We gratefully acknowledge the permission of the OECD to base this description on that manual.

At the international level:

- The **PISA Web page** (www.pisa.oecd.org): i) it provides descriptions about the programme, contact information, participating countries and results of PISA 2000, ii) it allows users to download the complete micro-level database, all questionnaires, publications, national reports and the media cover of PISA 2000, and iii) it provides an opportunity for users to generate their own tables or request specific ones.

- The publication **Knowledge and Skills for Life - First Results from PISA 2000** (OECD, 2001) includes the first results from PISA 2000. It presents evidence on student performance in reading, mathematical and scientific literacy, reveals factors that influence the development of these skills at home and at school, and examines what the implications are for policy development.

- The publication **Sample Tasks from the PISA 2000 Assessment – Reading, Mathematical and Scientific Literacy** (OECD, 2002a) describes the instruments underlying the PISA assessment. It introduces the PISA approach to assessing reading, mathematical and scientific literacy with its three dimensions of processes, content and context. Further it presents tasks from the PISA 2000 assessment together with how these tasks were scored and how they relate to the conceptual framework underlying PISA.

- The publication **PISA 2000 Technical Report** (OECD, 2002b) presents the methodology and procedures used in PISA.

Which instruments were included in PISA 2000?

In PISA 2000, a rotated test design was used to assess student performance in reading, mathematical and scientific literacy (for the complete conceptual frameworks see OECD, 1999b and OECD, 2000). This type of test design ensures a wide coverage of content while at the same time keeping the testing burden on individual students low. Nine test booklets were distributed at random to students. These booklets included questions assessing reading literacy, mathematical literacy and scientific literacy, but not all booklets assessed the same domains. Students were randomly assigned a testing booklet within each of the sampled schools.

- Booklets 8 and 9 contained reading, mathematics and science questions;
- Booklets 1, 3 and 5 contained reading and mathematics questions;
- Booklets 2, 4 and 6 contained reading and science questions; and,
- Booklet 7 contained only reading questions.

Questionnaires

Student questionnaires

A student questionnaire was designed to collect information about the student's family, home environment, reading habits, school and everyday activities. This information was later analysed both independently and in relation to performance.
Additional variables were collected for two research projects that are part of the National Research Programme "Formation and Employment" (NRP 43). These variables are included in the database of the national sample in addition to the international variables.

Additionally, the programme included two additional optional questionnaires for students. The first one was a cross curriculum competencies questionnaire asking about students' strategies of self-regulated learning, motivational preferences and self-concept. The second one was a computer familiarity questionnaire, including questions about students' use of computers, the availability of computers, and students' self-assessment of their computer skills.

School questionnaire

The principals or head administrator of the participating schools responded to a school questionnaire covering issues such as the demographics of the school, school staffing, the school environment, human and material educational resources in the school, selection and transfer policies, and educational and decision-making practices in the school.

Structure of the testing session

The student testing session consisted of:

- two 60-minute sessions assessing reading, mathematical and scientific literacy;
- 20-30 minutes for the student questionnaire;
- 10 minutes for the international option of cross curriculum competencies questionnaire; and
- 5 minutes for the international option of computer familiarity questionnaire.

The school principal or head administrator answered a 20-30 minute school questionnaire.

What is available from the PISA 2000 Swiss database?

Questionnaires

The following questionnaires are available: the student questionnaire (including questions relating to cross curriculum competencies and computer familiarity) and the school questionnaire in French.

Codebooks

In the French Subsample database a standard codebook is NOT avilable.

But there is a list avialble with the variables which were transformed.
And there is also a file called codebook aviable with the variable names, labels and value names and labels for the national sample (students), which is corresponding with most of the variables of the French Subsample Data.

Data files in SPSS® format

The student file of the French Subsample
The database is available in SPSS® format. As it is, it includes one row for each student with his or her performance scores in reading, mathematic and scientific literacy, the student weights, the replicates for computing the sampling variance, the responses to all questions and various indices. All variables are described in detail below.

For each student who participated in the assessment, the following information is available:

- Identification variables for the school and student.
- The student responses on the three questionnaires, i.e., the student questionnaire and the two international options: computer familiarity questionnaire and cross curriculum competencies questionnaire.
- The students' indices derived from the original questions in the questionnaires.
- The student weights for reading, mathematics and science.
- The students' performance scores in reading.
- The students' performance scores in mathematics for each student who was assessed with one of the booklets that contain mathematics material.
- The students' performance scores in science for each student who was assessed with one of the booklets that contain science material.
- The 132 reading, 132 mathematics and 132 science Jacknife's replicates for the computation of the sampling variance estimates..

The assessment items data file of the French Subsample
For each item included in the test, this file shows the students' responses expressed in a one-digit format. The items from mathematics and science used double-digit coding during marking.[2]
For details, see OECD (2002c).

The school questionnaire data file of the French Subsample
For each school that participated in the assessment, the following information is available:

- The identification variables for the school.

- The school responses on the school questionnaire.

- The school indices derived from the original questions in the school questionnaire.

Which records are included in the database of the French Subsample?

Records included in the database:

Student level
- All PISA students who attended one of the two test (assessment) sessions.

- PISA students who only attended the questionnaire session are included if they provided a response to the father's occupation questions or the mother's occupation questions on the student questionnaire (questions 8 to 11).

School level
All participating schools – that is, any school where at least 25 per cent of the sampled eligible students were assessed – have a record in the school level database, regardless of whether the school returned the school questionnaire.

[2] The responses from open-ended items could give valuable information about students' ideas and thinking, which could be fed back into curriculum planning. For this reason the marking guides for these items in mathematics and science were designed to include a two-digit marking so that the frequency of various types of correct and incorrect response could be recorded. The first digit was the actual score. The second digit was used to categorise the different kinds of response on the basis of the strategies used by the student to answer the item. The international database includes only the first digit.

Records excluded from the database:

Student level

- Sampled students who were reported as not eligible, students who were no longer at school, students who were excluded for physical, mental or linguistic reasons, and students who were absent on the testing day.

- Students who refused to participate in the assessment sessions.

- Students from schools where less than 25 percent of the sampled and eligible students participated.

School level

- Schools where fewer than 25 per cent of the sampled eligible students participated in the testing sessions.

How are missing data represented?

For most of the variables, the missing data is not defined.

For some variables, the missing data is represented in accordance of the national sample. In these cases, the coding of the data distinguishes between four different types of missing data:

- Item level non-response: 9 for a one-digit variable, 99 for a two-digit variable, 999 for a three-digit variable, and so on. Missing codes are shown in the codebooks. This missing code is used if the student or school principal was expected to answer a question, but no response was actually provided.

- Multiple or invalid responses: 8 for a one-digit variable, 98 for a two-digit variable, 998 for a three-digit variable, and so on. This code is used for multiple choice items in both test booklets and questionnaires where an invalid response was provided. This code is not used for open-ended questions.

- Not applicable: 7 for a one-digit variable, 97 for a two-digit variable, 997 for a three-digit variable, and so on for the student questionnaire data file and for the school data file. Code “n” is used for a one-digit variable in the test booklet data file. This code is used when it was not possible for the student to answer the question. For instance, this code is used if a question was misprinted or if a question was deleted from the questionnaire by a national centre. The not-applicable codes and code “n” are also used in the test booklet file for questions that were not included in the test booklet that the student received.

- Not reached items: all consecutive missing values starting from the end of each test session were replaced by the non-reached code, “r”, except for the first value of the missing series, which is coded as missing.

How are the students identified?

The student identification consists of three variables, which together form a unique identifier for each student:

- The stratum identification variable labelled STIDSTRT.

- The school identification variable labelled STIDSCH.

- The student identification variable labelled STIDSTD.

For the sampling design and the allocation of the schools to the different strata, see Renaud (2002).

The school identification consists of two variables, which together form a unique identifier for each school:

- The stratum identification variable labelled STIDSTRT.

- The school identification variable labelled STIDSCH.

The unique identifier that is constructed of these two variables is labelled SCHOOLID.

The weights and replicates

The weights

Students included in the final FS Grade9 Sample File are not equally representative of the full student population of the french part of Switzerland. Sampling weights must be applied to compensate for differences in the selection probabilities of students. To account for the sample design during the analyses, so as not to produce biased results, survey weights must be incorporated into the analysis. The calculation of these weights is described in the document "Pondération_SR", included in the folder "Dokumentation Consortium romand". The cases of the canton Jura were not weighted, that's why this canton is not represented in the table in the document "Pondération_SR".

The replicates

Jacknife's replicates are included in the data file because they are needed to compute unbiased-standard error estimates associated with any population parameter estimates. They are the chosen option of the software program WesVar®.
In the national data file of the national sample, the option of Fay's replicates had been chosen.

The standard error (of sampling) provides an estimate of the degree to which a statistic (such as a mean score) may be expected to vary about the true (but unknown) population mean. A 95% confidence interval for a mean (consisting of a region from 1.96 standard errors below the mean to 1.96 standard errors above the mean) may be constructed in such a way that, if the sampling procedure were repeated a large number of times, and the sample statistic re-computed each time, the confidence interval would be expected to contain the population estimate 95% of the time. Jacknife's replicates take into account the complex, two-stage, stratified sample design. If this is not done, one underestimates the standard error, thereby running the risk of obtaining statistical significance when in fact there is none. There are several methods of doing this, WesVar®, SAS®, SPSS® or others.

Information about the students in the French Subsample database

The responses to the questionnaires

The files contain the original variables collected through the student context questionnaires, i.e., the compulsory student questionnaire and the two international options: the cross curriculum competencies questionnaire (CC) and the computer familiarity questionnaire (IT).

The names that are used to represent these variables in the database are directly related to the context questionnaires. Each variable name consists of seven characters.

ST??Q??

The sixth and seventh characters refer to the item number of the question. For instance, ST01Q01 is the day of birth, ST01Q02 is the month of birth and ST01Q03 is the year of birth.

The third and fourth characters refer to the question number as it appears in the questionnaire. For instance, ST01 refers to the first question in the student questionnaire relating to the date of birth.

The first two characters refer to the instrument:

ST for the student questionnaire,
IT for the computer familiarity questionnaire,
CC for the cross curriculum competencies questionnaire.

National adaptations of internationally used variables contain the character N instead of Q (e.g. ST01N04). Moreover, there are some additionally collected variables on the national level that are not being used on the international level. All variables starting from ST42Qxx are of this type.

The student questionnaire indices

Several of PISA's measures reflect indices that summarise responses from students or school representatives (typically principals) to a series of related questions. The questions were selected from larger constructs on the basis of theoretical considerations and previous research. Structural equation modelling was used to confirm the theoretically expected behaviour of the indices and to validate their comparability across countries. For this purpose, a model was estimated separately for each country and, collectively, for all OECD countries.

This section explains the indices derived from the student and school context questionnaires that are used in this report. For a description of other PISA indices and details on the methods see the PISA 2000 Technical Report.

Unless otherwise indicated, where an index involves multiple questions and student responses, the index was scaled using a weighted maximum likelihood estimate, using a one-parameter item response model (referred to as a WARM estimator; see Warm, 1985) with three stages:

- The question parameters were estimated from equal-sized sub-samples of students from each OECD country.

- The estimates were computed for all students and all schools by anchoring the question parameters obtained in the preceding step.

- The indices were then standardised so that the mean of the index value for the OECD student population was zero and the standard deviation was one (countries being given equal weight in the standardisation process).

It is important to note that negative values in an index do not necessarily imply that students responded negatively to the underlying questions. A negative value merely indicates that a group of students (on all students, collectively, in a single country) or principals responded less positively than all students or principals did on average across OECD countries. Likewise, a positive value on an index indicates at a group of students or principals responded more favourably, or more positively, than students or principals did, on average, in OECD countries.

Terms enclosed in brackets < > in the following descriptions were replaced in the national versions of the student and school questionnaires by the appropriate national equivalent. For example the term <classes in the language of assessment> was translated into “German classes”, “French classes” or “Italian classes” depending on whether students received the German, French or Italian version of the assessment instruments.

For the reliabilities of the indices, see the *PISA 2000 Technical Report*.

Indices derived through a direct combination of the answers from the student questionnaire

1) The following indices were included in the student questionnaire file:

- ***Time in minutes spent each week at school in reading (RMINS), mathematics (MMINS) and science (SMINS) courses***. The three variables are simply the product of the following corresponding items:

    - How many <class periods> the students spent in courses in each of the three domains during the last full week (ST27Q01 for <test language> courses, ST27Q03 for <mathematics> courses, and ST27Q05 for <science> courses); and

    - The number of instructional minutes in the average single <class period> from the school questionnaire (SC06Q03).

- ***Age (AGE)***. The age of the student expressed in months computed from the students’ date of birth (ST01).

- ***Family structure (FAMSTRUC)***. Students were asked to report who usually lived at home with them (items of ST04). The responses were then grouped into four categories:

    - single-parent family coded as 1 (students who reported living with one of the following: mother, father, female guardian or male guardian);

    - nuclear family coded as 2 (students who reported living with a mother and a father);

    - mixed family coded as 3 (students who reported living with a mother and a male guardian, a father and a female guardian, or two guardians); and

- other response combinations coded as 4.

- ***Number of siblings (NSIB)***. Students were asked to indicate how many brothers and sisters they had older than themselves, younger than themselves, or of the same age. For the analyses in Chapter 8 (OECD, 2001), the numbers in each category were added together. This variable is based on the three items of question ST05.

- ***Birth order (BRTHORD)***. Also based on ST05, this variable received a value of 0 if the student was the only child, 1 if the student was the youngest child, 2 if the student was a middle child, and 3 if the student was the oldest child.

- ***Father's occupation (BFMJ), mother's occupation (BMMJ) and student's expected occupation at the age of 30 (BTHR)***. Students were asked to report their mothers' and fathers' occupations, and to state whether each parent was: in full-time paid work; part-time paid work; not working but looking for a paid job; or "other". The students' open-ended responses to questions ST08Q0l, ST09Q01, ST10Q01, ST11Q01 and ST40Q01 were then coded in accordance with the International Standard Classification of Occupations (ISCO 1988), with these variables receiving the actual ISCO code and later recoded according to the PISA International Socio-Economic Index of Occupational Status (ISEI) explained below.

- ***PISA International Socio-Economic Index of Occupational Status (ISEI)***. Additionally, these variables were transformed to create the PISA International Socio-Economic Index of Occupational Status, derived from students' responses on parental occupation. The index captures the attributes of occupations that convert parents' education into income. The index was derived by the optimal scaling of occupation groups to maximise the indirect effect of education on income through occupation and to minimise the direct effect of education on income, net of occupation (both effects being net of age). For more information on the methodology, see Ganzeboom *et al.* (1992). The ISEI variable is equal to the father's occupation or to the mother's occupation if the father's ISEI is missing. A second variable is also included ***(HISEI),*** based on either the father's or mother's occupations, whichever is the higher. Values on the index range from 16 to 90; low values represent low socio-economic status and high values represent high socio-economic status.

- ***Parental education (FISCED for fathers and MISCED for mothers)***. Students were asked to classify the highest level of education of their mother and father on the basis of national qualifications, which were then coded in accordance with the International Standard Classification of Education (ISCED 1997) in order to obtain internationally comparable categories of educational attainment. These were collected in two questions about each parent (questions ST12Q01 and ST14Q01 for the mother and questions ST13Q01 and ST15Q01 for the father). The father's educational level (FISCED) and the mother's educational level (MISCED) have the following categories, which are defined in accordance with the International Student Classification of Education (ISCED) (OECD, 1999a):

1. Did not go to school;

2. Completed <ISCED Level 1 only (primary education)>;

   The following terms were used in Switzerland:
   French: L'école primaire (4 - 6 ans de scolarité)
   German: Primarschule besucht (4 bis 6 Schuljahre)

Italian: Scuola elementare (4-6 anni di scuola)

3. Completed <ISCED Level 2 only (lower secondary level)>;

   The following terms were used in Switzerland:
   French: La scolarité obligatoire (7-9 ans de scolarité)
   German: obligatorische Schule abgeschlossen (7 bis 9 Schuljahre)
   Italian: Scuola dell'obbligo (7-9 anni di scuola)

4. Completed <ISCED Level 3B or 3C only (upper secondary level, aimed in most countries at providing direct entry into the labour market)>;

   The following terms were used in Switzerland:
   French: Un apprentissage, une école professionnelle, une école de culture générale de degré diplôme
   German: Berufslehre oder Berufsschule abgeschlossen (dazu gehören z.B. auch Handelsschule oder Diplommittelschule)
   Italian: Scuola cantonale di amministrazione, scuola di cultura generale, scuola di diploma, apprendistato o scuola professionale a tempo pieno

5. Completed <ISCED Level 3A (upper secondary, aimed in most countries at gaining entry into tertiary education)>; and

   The following terms were used in Switzerland:
   French: Une école du niveau de la Maturité (gymnase, collège, école normale)
   German: Gymnasium, Kantonsschule oder Primarlehrer/innenseminar abgeschlossen
   Italian: Scuola di maturità (liceo, scuola cantonale di commercio, scuola magistrale, maturità professionale)

6. Completed <ISCED Level 5A, 5B or 6 (tertiary education)>.

   The following terms were used in Switzerland:
   French: Université, Ecole polytechnique, Haute école spécialisée, Haute école pédagogique, formation professionnelle supérieure ou maîtrise fédérale
   German: Universität, Hochschule, Höhere Fach- und Berufsausbildung
   Italian: Scuola universitaria professionale, università, politecnico, alta scuola pedagogica, formazione professionale superiore o maestria federale

Note: Years of schooling was used in Chapter 8 of the First Results from PISA 2000 (OECD, 2001) as a conversion of the highest level of educational attainment of the parents.

2) Weighted likelihood estimates

Indices from the student questionnaire

Fifteen indices from the student questionnaire were derived using the weighted estimate method (Warm, 1985). These indices are:

- ***Index of cultural communication with parents (CULTCOM***). This index was derived from students' reports on the frequency with which their parents (or guardians) engaged with them in the following activities: discussing political or social issues; discussing books, films or television programmes; and listening to classical music. It was based on questions ST19Q01, ST19Q02 and ST19Q03.

- ***Index of social communication with parents (SOCCOM)***. This index was derived from students' reports on the frequency with which their parents (or guardians) engaged with them in the following activities: discussing how well they are doing at school; eating <the main meal> with them around a table; and spending time simply talking with them. It was based on questions ST19Q04, ST19Q05 and ST19Q06.

- ***Index of family educational support (FAMEDSUP)***. This index was derived from the students' reports on the frequency with which the following people work with them on their schoolwork: their mother, their father, their brothers and sisters. It was derived from questions ST20Q01, ST20Q02 and ST20Q03.

- ***Index of family wealth (WEALTH)***. This index was derived from students' reports on: i) the availability, in their home, of a dishwasher, a room of their own, educational software, and a link to the Internet; and ii) the number of cellular phones, television sets, computers, motor cars and bathrooms at home. It was based on questions ST21Q01, ST21Q02, ST21Q03, ST21Q04, ST22Q01, ST22Q02, ST22Q04, ST22Q06 and ST22Q07.

- ***Index of home educational resources (HEDRES)***. This index was derived from students' reports on: i) the availability, in their home, of a dictionary, a quiet place to study, a desk for study, and textbooks; and ii) the number of calculators at home. It was based on questions ST21Q05, ST21Q06, ST21Q07, ST21Q08 and ST22Q03.

- ***Index of activities related to "classical" culture (CULTACTV)***. This index was derived from students' reports on how often they had participated in the following activities during the preceding year: visited a museum or art gallery; attended an opera, ballet or classical symphony concert; and watched live theatre. It was derived from questions ST18Q02, ST18Q04 and ST18Q05.

- ***Index of possessions related to "classical" culture in the family home (CULTPOSS***). This index was derived from students' reports on the availability of the following items in their home: classical literature (examples were given); and books of poetry and works of art (examples were given). It was based on questions ST21Q09, ST21Q10 and ST21Q11.

- ***Index of time spent on homework (HMWKTIME)***. This index was derived from students' reports on the amount of time they devote to homework per week in the <language of assessment>, mathematics, and science. It was based on questions ST33Q01, ST33Q02 and ST33Q03.

- ***Index of teacher support (TEACHSUP)***. This index was derived from students' reports on the frequency with which: the teacher shows an interest in every student's learning; the teacher gives students an opportunity to express opinions; the teacher helps students with their work; the teacher continues teaching until the students understand; the teacher does a lot to help students; and, the teacher helps students with their learning. It was derived from questions ST26Q05, ST26Q06, ST26Q07, ST26Q08, ST26Q09 and ST26Q10.

- ***Index of disciplinary climate (DISCLIMA)***. This index derived from students' reports on the frequency with which, in their <language of assessment class>: the teacher has to wait a long time for students to <quieten down>; students cannot work well; students don't listen to what the teacher says; students don't start working for a long time after the lesson begins; there is noise and disorder; and, at the start of class, more than five minutes are spent doing nothing. It was based on questions ST216Q01, ST26Q12, ST26Q13, ST26Q14, ST26Q16 and ST26Q17. This index was inverted during reporting so that low values indicate a poor disciplinary climate (OECD, 2001).

- ***Index of teacher-student relations (STUDREL)***. This index was derived from students' reports on their level of agreement with the following statements: students get along well with most teachers; most teachers are interested in students' well-being; most of their teachers really listen to what they have to say; if they need extra help, they will receive it from their teachers; and, most of their teachers treat them fairly. It was based on questions ST30Q01 to ST30Q05.

- ***Index of achievement press (ACHPRESS)***. This index was derived from students' reports on the frequency with which, in their <language of assessment class>: the teacher wants students to work hard; the teacher tells students that they can do better; the teacher does not like it when students deliver <careless> work; and, students have to learn a lot. It was based on questions ST26Q02, ST26Q03, ST26Q04 and ST26Q15.

- ***Index of student's sense of belonging in the school (BELONG)***. This index was derived from students' reports on their level of agreement with the following statements concerning their school: I feel like an outsider (or left out of things); I make friends easily; I feel like I belong; I feel awkward and out of place; other students seem to like me; and, I feel lonely. It was based on questions ST31Q01 to ST31Q06.

- ***Index of engagement in reading (JOYREAD)***. This index was derived from students' level of agreement with the following statements : I read only if I have to; reading is one of my favourite hobbies; I like talking about books with other people; I find it hard to finish books; I feel happy if I receive a book as a present; for me, reading is a waste of time; I enjoy going to a bookstore or a library; I read only to get information that I need; and, I cannot sit still and read for more than a few minutes. It was based on questions ST35Q01 to ST35Q09.

- ***Index of reading diversity (DIVREAD)***. This index was derived from the frequency with which students read the following materials because they wanted to: magazines, comic books, fiction (examples were given), non-fiction books, emails and Web pages, and newspapers. It was based on questions ST36Q01 to ST36Q06. For this index, categories 1 and 2 were recoded as 0 and categories 3, 4, 5 were recoded as 1.

These indices, based on weighted estimates (Warm, 1985), were standardised to have a mean of 0 and a standard deviation of 1 at the international level using the same procedures that were applied to the performance variables. Suggestions for ways of analysing these indices are given in the sub-section on "Analysis of the questionnaire data".

Indices from the computer familiarity questionnaire

Three indices from the student computer familiarity questionnaire were derived using the weighted estimate method (Warm, 1985). These indices are:

- ***Index of comfort with and perceived ability to use computers (COMAB)***. This index was derived from students' responses to the following questions: how comfortable are you with using a computer?; how comfortable are you with using a computer to write a paper?; how comfortable are you with taking a test on a computer?; and, if you compare yourself with other 15-year-olds, how would you rate your ability to use a computer? It was based on questions IT02Q01, IT02Q02, IT02Q03, and IT03Q01. The items were inverted. For information on the conceptual underpinning of the index see Eignor *et al.* (1998).

- ***Index of computer usage (COMUSE)***. This index was derived from students' responses to the frequency to which they use the computer for the following purposes: to help them learn school material; for programming; for word processing (examples of software packages were given); spreadsheets (examples of software packages were given); drawing, painting or graphics; and, educational software. It was based on questions IT05Q03, IT05Q04, IT06Q02, IT06Q03, IT06Q04, and IT06Q05. The items were inverted.

- ***Index of interest in computers (COMATT)***. This index was derived from students' responses to the following statements: it is very important to me to work with a computer; to play or work with a computer is really fun; I use a computer because I am very interested in this; and, I forget the time, when I am working with the computer. It is based on questions IT07Q01, IT08Q01, IT09Q01, and IT10Q01. The items were inverted. For information on the conceptual underpinning of the index see Eignor *et al.* (1998).

These indices, based on weighted estimates (Warm, 1985), were standardised to have a mean of 0 and a standard deviation of 1 at the international level using the same procedures as were applied to the performance variables. Only OECD countries (except Netherlands) that participated in the optional computer familiarity component (IT questionnaire) were included in this transformation.

The indices from the cross-curriculum competencies questionnaire

Fourteen indices from the student cross-curriculum competencies questionnaire (also known as CC questionnaire) were derived using the weighted estimate method (Warm, 1985). These indices are:

- ***Index of control strategies (CSTRAT)***. This index was derived from the frequency with which students used the following strategies when studying: I start by figuring out what exactly I need to learn; I force myself to check to see if I remember what I have learned; I try to figure out, as I read, which concepts I still haven't really understood; I make sure that I remember the most important things; and, when I study and I don't understand something, I look for additional information to clarify the point. It was based on questions CC0lQ03, CC01Q13, CC01Q19, CC01Q23 and CC01Q27. For information on the conceptual underpinning of the index see Baumert *et al.* (1994).

- ***Index of effort and perseverance (EFFPER)***. This index was derived from the frequency with which students used the following strategies when studying: I work as hard as possible; I keep working even if the material is difficult; I try to do my best to acquire the knowledge and skills taught; and, I put forth my best effort. It was based on questions CC01Q07, CC0lQ12, CC01Q20 and CC01Q28.

- ***Index of memorisation strategies (MEMOR)***. This index was derived from the frequency with which students used the following strategies when studying: I try to memorise everything that might be covered; I memorise as much as possible; I memorise all new material so that I can recite it; and, I practise by saying the material to myself over and over. It was based on questions CC01Q01, CC01Q05, CC01Q10 and CC01Q15. For information on the conceptual underpinning of the index see Baumert *et al.* (1994) and Pintrich *et al.* (1993).
- ***Index of perceived self-efficacy (SELFEF)***. This index was derived from the frequency with which students used the following strategies when studying: I am certain I can understand the most difficult material presented in readings; I am confident I can do an excellent job on assignments and tests; and, I am certain I can master the skills being taught. It was based on questions CC01Q02, CC01Q18 and CC01Q26.

- ***Index of control expectation (CEXP)***. This index was derived from the frequency with which students used the following strategies when studying: when I sit myself down to learn something really hard, I can learn it; if I decide not to get any bad grades, I can really do it; if I decide not to get any problems wrong, I can really do it; and, if I want to learn something well, I can. It was based on questions CC01Q04, CC01Q11, CC01Q16 and CC01Q24.

- ***Index of elaboration strategies (ELAB)***. This index was derived from the frequency with which students used the following strategies when studying: I try to relate new material to things I have learned in other subjects; I figure out how the information might be useful in the real world; I try to understand the material better by relating it to things I already know; and, I figure out how the material fits in with what I have learned. It was based on questions CC01Q09, CC01Q17, CC01Q21 and CC01Q25. For information on the conceptual underpinning of the index see Baumert *et al.* (1994).

- ***Index of instrumental motivation (INSMOT)***. This index was derived from the frequency with which students study for the following reasons: to increase my job opportunities; to ensure that my future will be financially secure; and, to get a good job. It was based on questions CC01Q06, CC01Q14 and CC01Q22.

- ***Index of interest in mathematics (INTMAT)***. This index was derived from students' level of agreement with the following statements: when I do mathematics, I sometimes get totally absorbed; mathematics is important to me personally; and, because doing mathematics is fun, I wouldn't want to give it up. It was based on questions CC02Q01, CC02Q10 and CC02Q21. For information on the conceptual underpinning of the index see Baumert *et al.* (1997).

- ***Index of self-concept in mathematics (MATCON)***. This index was derived from students' level of agreement with the following statements: I get good marks in mathematics; mathematics is one of my best subjects; and, I have always done well in mathematics. It was based on questions CC02Q12, CC02Q15 and CC02Q18. For information on the conceptual underpinning of the index see Marsh *et al.* (1992).

- ***Index of interest in reading (INTREA)***. This index was derived from students' level of agreement with the following statements: because reading is fun, I wouldn't want to give it up; I read in my spare time; and, when I read, I sometimes get totally absorbed. It was based on questions CC02Q06, CC02Q13 and CC02Q17. For information on the conceptual underpinning of the index see Baumert *et al.* (1997).

- ***Index of self-concept academics (SCACAD)***. This index was derived from students' level of agreement with the following statements: I learn things quickly in most school subjects; I am good at most school subjects; and I do well in tests in most school subjects. It was based on questions CC02Q03, CC02Q07 and CC02Q20.

- ***Index of self-concept in reading (SCVERB)***. This index was derived from students' level of agreement with the following statements: I'm hopeless in <language of assessment classes>; I learn things quickly in the <language of assessment classes>; and, I get good marks in the <language of assessment>. It is based on questions CC02Q05, CC02Q09 and CC02Q23. For information on the conceptual underpinning of the index see Marsh *et al.* (1992).

- ***Index of competitive learning (COMLRN)***. This index was derived from students' level of agreement with the following statements: I like to try to be better than other students; trying to be better than others makes me work well; I would like to be the best at something; and, I learn things faster if I'm trying to do better than the others. It is based on questions CC02Q04, CC02Q11, CC02Q16 and CC02Q24. For information on the conceptual underpinning of the index see Owens and Barnes (1992).

- ***Index of co-operative learning (COPLRN)***. This index was derived from students' level of agreement with the following statements: I like to work with other students; I learn the most when I work with other students; I like to help other people do well in a group; and, it is helpful to put together everyone's ideas when working on a project. It is based on questions CC02Q02, CC02Q08, CC02Q19 and CC02Q22. For information on the conceptual underpinning of the index see Owens and Barnes (1992).

These indices, based on weighted estimates (Warm, 1985), were standardised to have a mean of 0 and a standard deviation of 1 at the international level using the same procedures as were applied to the performance variables. Only OECD countries (except Netherlands) that participated in the international cross-curriculum competencies option (CC questionnaire) were inclu ded in this transformation.

The weights

For information on the construction of the weight variables see the description in the document "Pondération_SR" in the file "Dokumentation Consortium romand".

The student performance scores

For each domain, i.e., reading, mathematics and science, and for each subscale in reading, two kinds of estimate are provided**: a weighted likelihood estimate** (WLE) and a **set of plausible values**.

It is recommended that the set of plausible values be used when analysing and reporting statistics at the population level. Using WLEs for population estimates will yield biased estimates.

The weighted likelihood estimates

The Swiss database provides six weighted likelihood estimates, respectively labelled:

- variable WLEREAD to represent the reading ability estimate, which is provided for all students who answered at least one reading question;

- variable WLEREAD1 to represent the reading ability estimate for the retrieving subscale, which is provided for all students who answered at least one reading retrieving question;

- variable WLEREAD2 to represent the reading ability estimate for the interpreting subscale, which is provided for all students who answered at least one reading interpreting question;

- variable WLEREAD3 to represent the reading ability estimate for the reflecting and evaluating subscale, which is provided for all students who answered at least one reading reflecting and evaluating question;

- variable WLEMATH to represent the mathematics ability estimate, which is provided only for students who took booklets 1, 3, 5, 8 or 9 or and answered at least one mathematics question; and

- variable WLESCIE to represent the science ability estimate, which is provided only for students who took booklets 2, 4, 6, 8 or 9 and answered at least one science question.

The plausible values

The plausible values represent a set of random values for each selected student at random from an estimated ability distribution of students with similar item response patterns and backgrounds. They are intended to provide good estimates of parameters of student populations (for example, country mean scores), rather than estimates of individual student proficiency, which are better estimated using the weighted likelihood estimates.

The database provides five plausible values for each domain and each reading subscale, respectively labelled:

- PV1read to PV5read for reading ability,

- PV1read1 to PV5read1 for reading ability, *retrieving information subscale*,

- PV1read2 to PV5read2 for reading ability, *interpreting subscale*,

- PV1read3 to PV5read3 for reading ability, *reflecting and evaluating subscale*,

- PV1math to PV5math for mathematics ability,

- PV1scie to PV5scie for science ability.

Each student included in the Swiss database has performance plausible values for the reading domain as well as for the reading subscales. Only students who were assessed with booklets 1, 3, 5, 8 or 9 will have plausible values in mathematics, and only students who were assessed with booklets 2, 4, 6, 8 or 9 will have plausible values in science.

Transformation of the ability estimates

The weighted likelihood estimates and the plausible values were transformed to a scale with a mean of 500 and a standard deviation of 100 by using the data for the participating OECD countries only (except the Netherlands[3]). These linear transformations used weighted data, with an additional adjustment factor so that each country contributed equally. The standardisation parameters were derived from the average of the mean and standard deviation computed from each of the five plausible values. This means that although the mean and standard deviation of individual plausible values will not be exactly 500 and 100, respectively, the average of the five means and the five standard deviations for each scale will be 500 and 100, respectively.

The transformation that was used to give reading a mean of 500 and a standard deviation of 100 was also applied to the three reading subscales. This means that the mean and the standard deviation for the reading subscales will differ from 500 and 100, respectively.

How to analyse data using the plausible values

It is important to recognise that plausible values are not test scores and should not be treated as such. As noted above, plausible values are random numbers that are drawn from the distribution of scores that could be reasonably assigned to each individual. As such, the plausible values contain random error variance components (that is, variation between individual plausible values assigned to each student) and are not optimal as scores for individuals. However, the important characteristic of plausible values is that as a set, they are better suited for describing the performance of the population than a set of scores that are optimal at the individual student level (for example, the weighted likelihood estimates).

Plausible values can be thought of as intermediate values that provide consistent estimates of population parameters. Such estimates can be obtained using statistical software such as WesVar®, SPSS® and SAS®.

During data exploration, there is no need to work with the five plausible values; one can use a single plausible value. On average, one plausible value will provide unbiased estimates of population parameters. However for the final estimates, it is recommended that all five plausible values be used, otherwise the standard error estimated from one plausible value will only contain the sampling variance component while it should also contain the measurement error component. This means that the analysis should be undertaken five times, once with each of the five relevant plausible values. The results of these five analyses need to be combined so that the associated standard error incorporates measurement error associated with the variance between the five plausible values. The method for combining them is described below in two sections: one for users of the WesVar® software, and one for users of the SAS® and SPSS® software systems. An example of computing correlation using plausible values is included later in the document.

1. Using WesVar®

The WesVar® software can incorporate the five plausible values and produce the correct standard errors in the calculation of means of groups, using the ‘PV’ function. The degrees of freedom that WesVar® uses for these analyses are not the actual degrees of freedom but rather the number of replicate weights, 132 in the case of the PISA French Subsample database. This is considered an accurate approximation to the actual degrees of freedom for the vast majority of analyses.

[3] Response rate too low to ensure comparability (Annex A3, OECD(2001)).

For other types of estimate, such as quartiles or medians, the analysis in question must be carried out five times and the five estimates combined as described in the section on SAS® and SPSS® that follows.

2. Using SAS® and SPSS®

As computer packages such as SAS® and SPSS® do not provide standard (measurement) errors associated with estimates, it is necessary to compute such standard errors using the following procedure. (Note that WesVar® only provides correct standard errors associated with means, so all other types of analysis done in WesVar® should also be undertaken using the procedures below.)

1. Separate estimates need to be computed for each plausible value. This will result in five estimated parameters (one associated with each plausible value). Each set (P1 to P5) should then be averaged to provide a mean parameter estimate (MP). Standard errors (SE1 to SE5) also need to be generated for each parameter estimate (P1 to P5).

2. The measurement error and sampling variances for the mean parameter estimate (MP) should then be computed. The measurement error variance should be computed using the following formula:

$$[(MP-P1)^2 + (MP-P2)^2 + (MP-P3)^2 + (MP-P4)^2 + (MP-P5)^2]/4$$

The sampling variance should be computed using the following formula:

$$[(SE1^2 + SE2^2 + SE3^2 + SE4^2 + SE5^2)]/5$$

The total variance should then be computed by summing the measurement error and the sampling variances. In doing so, a weight of 1.2 (1 + 1 /M, where M is the number of plausible values) should be applied to the measurement error variance. The square root of the total variance provides an estimate of the standard error of the parameter estimate. Note that outputs from SAS®, SPSS® and WesVar® can be pasted into spreadsheet packages such as Excel®, which can then be used to semi-automate this procedure, if many such analyses are to be undertaken. An example of how an Excel® spreadsheet can be set up is given in Figure 1, below.[4]

Figure 1: Formulae for computation of standard errors of plausible values in Excel

Plausible Value	Parameter Estimate	Standard Error*
1	[a1]	[b1]
2	[a2]	[b2]
3	[a3]	[b3]
4	[a4]	[b4]
5	[a5]	[b5]
Sampling variance	=(b1^2+ b2^2+ b3^2+ b4^2+ b5^2)/5 [a6]	
Mean parameter estimate	=(a1+a2+a3+a4+a5)/5 [a7]	
Measurement variance	=((a1-a7)^2+(a2-a7)^2+(a3-a7)^2+(a4-a7)^2+(a5-a7)^2)/4 [a8]	
Variance of parameter estimate	=a6+(1.2*a8) [a9]	
Corrected standard error	=sqrt(a9)	

* If SAS® or SPSS® are used, the SE should be estimated as previously described.

[4] Formulae kindly provided by Keith Rust and Sheila Krawchuk of Westat, Inc.

How to analyse the data using the proficiency scale

PISA 2000 assessed reading literacy as the major domain, while keeping mathematics and science as minor domains. That means that two-thirds of the assessment was in reading literacy tasks. The reading scales were divided into five levels of knowledge and skills, facilitating their interpretation, and because of the manner in which the PISA performance data have been scaled, it is possible to describe what students scoring at around a particular point are able to do. Because both item difficulties and student performance scores are scaled to the same metric, one can examine items of similar difficulty and make inferences about the underlying skills and complexity of reasoning that are required to respond correctly to such clusters of items. Therefore, the application of techniques associated with item response theory to the PISA performance data means that it is possible to generate a criterion-referenced interpretation of student proficiency. The creation of proficiency levels is extremely useful from a policy and pedagogical point of view because it provides a shorthand description of what students in each group are likely to be able to do. Comparisons of the proportions of students at each proficiency level within and between countries can yield useful information about the relative strengths and weaknesses of groups of students. The development of the proficiency levels for PISA involved establishing appropriate cut-off points for each level, and developing a substantive description of the skills and knowledge associated with each level through a detailed examination of the items associated with these levels. The process of developing proficiency levels is thus an iterative one. Subject-matter experts and technical experts of the PISA consortium worked together to produce them.

PISA proficiency levels were defined in such a way that a student with a reading score at the bottom of a level has an average probability of .50 of correctly responding to all items at that level. Application of this criterion, and a further criterion that proficiency levels should be of fixed width (.80 logits), led to the establishment of a response probability convention of .62[5]. The label ‘below Level 1’ is assigned to students who did not meet the criterion for Level 1 (*i.e.*, the estimated probability of these students responding correctly to items at the bottom of Level 1 is less than .50). PISA does not describe what students below Level 1 can accomplish[6]. Similarly, PISA does not describe the upper limits of knowledge and skills of students at Level 5 on the scales (*i.e.*, students at this level may have additional skills not assessed by PISA).

The cut-off points for the reading scales and its three subscales are 334.75, 407.47, 480.18, 552.89 and 625.61. The five levels are defined in Figure 2.

[5] For analysts familiar with the International Adult Literary Survey (IALS) it is pertinent to point out that the response probability associated with the IALS proficiency levels was set at .80. This more stringent criterion means, in effect that one must be more certain that a person can correctly respond to items associated with a particular proficiency level in order to categorise that individual as belonging to that level. This is especially relevant if analysts of the PISA 2000 international database are making comparisons between performance on the PISA assessment of reading literacy and performance on IALS.

[6] Referring again to the IALS study, no distinction was made between students whose scores were below level 1. This may also be relevant to those wishing to make comparisons between the two studies.

Figure 2: *Cut points for proficiency levels for the PISA combined literacy scale and the three literacy subscales*

Level 0: the reading score is equal to or below 334.75;

Level 1: the reading score is greater than 334.75 and equal to or below 407.47;

Level 2: the reading score is greater than 407 .47 and equal to or below 480.18;

Level 3: the reading score is greater than 480.18 and equa l to or below 552.89;

Level 4: the reading score is greater than 552.89 and equal to or below 625.61;

Level 5: the reading score is greater than 625.61.

To estimate the percentages of students in each of the six levels, five new categorical variables should be computed, one for each of the five plausible values provided by each scale or subscale, using the type of syntax shown in Figure 3, taken from SPSS®. It is acceptable to combine levels, such as Level 1 and below Level 1, but advisable that explicit note of this is made to prevent misinterpretation of results.

Figure 3: *SPSS® syntax used to create six proficiency levels for each plausible value*

```
*individual plausible values: proficiency levels for overall reading produces a proficiency

*level pv1rlev, pv2rlev, etc. associated with each plausible value, pv1read, pv2read, etc.

IF (pv1read le 334.75) pv1rlev = 0.

IF (pv1read gt 334.75) pv1rlev = 1.

IF (pv1read gt 407.47) pv1rlev = 2.

IF (pv1read gt 480.18) pv1rlev = 3.

IF (pv1read gt 552.89) pv1rlev = 4.

IF (pv1read gt 625.61) pv1rlev = 5.

IF (pv2read le 334.75) pv2rlev = 0.

IF (pv2read gt 334.75) pv2rlev = 1.

IF (pv2read gt 407.47) pv2rlev = 2.

IF (pv2read gt 480.18) pv2rlev = 3.
```

```
IF (pv2read gt 552.89) pv2rlev = 4.

IF (pv2read gt 625.61) pv2rlev = 5.

(... and so on for each of the five plausible values.)
```

The proficiency level variables have been calculated for the 9th grade students and 15-year-old students in Switzerland and are stored in the database. For the five plausible values on overall reading they are labelled PV1RLEV to PV5RLEV. For the five plausible values on the reading/retrieving subscale they are labelled PV1RLEV1 to PV5RLEV1. For the five plausible values on the reading/interpreting subscale they are labelled PV1RLEV2 to PV5RLEV2. For the five plausible values on the reading/reflecting subscale they are labelled PV1RLEV3 to PV5RLEV3.

Percentages and sampling variance can be estimated with WesVar® for each of these categorical variables. The results then need to be combined as described above.

It is possible to shortcut this procedure by generating for each plausible value six dichotomous variables coded 0,1 (below level 1 or not, at level 1 or not, at level 2 or not, ... at level 5 or not). Therefore, 30 dichotomous variables need to be computed. As the percentage of students for these dichotomous variables can be estimated by computing the mean, then it becomes possible to use the PV function in WesVar®. The standard error will therefore consist of the sampling variance and the measurement error. Figure 4 shows the SPSS® syntax to generate the 30 dichotomous variables.

Figure 4: *SPSS® syntax to generate the proficiency levels using 30 dichotomous variables*

```
IF (pv1rlev = 0) p1level0 = 1.

IF (pv1rlev = 1 or pv1rlev = 2 or pv1rlev = 3 or pv1rlev = 4 or pv1rlev = 5) p1level0 = 0.

IF (pv1rlev = 1) p1level1 = 1.

IF (pv1rlev = 0 or pv1rlev = 2 or pv1rlev = 3 or pv1rlev = 4 or pv1rlev = 5) p1level1 = 0.

IF (pv1rlev = 2) p1level2 = 1.

IF (pv1rlev = 0 or pv1rlev = 1 or pv1rlev = 3 or pv1rlev = 4 or pv1rlev = 5) p1level2 = 0.

IF (pv1rlev = 3) p1level3 = 1.

IF (pv1rlev = 0 or pv1rlev = 1 or pv1rlev = 2 or pv1rlev = 4 or pv1rlev = 5) p1level3 = 0.

IF (pv1rlev = 4) p1level4 = 1.

IF (pv1rlev = 0 or pv1rlev = 1 or pv1rlev = 2 or pv1rlev = 3 or pv1rlev = 5) p1level4 = 0.

IF (pv1rlev = 5) p1level5 = 1.

IF (pv1rlev=0 or pv1rlev = 1 or pv1rlev = 2 or pv1rlev = 3 or pv1rlev = 4) p1level5 = 0.
```

IF (pv2rlev = 0) p2level0 = 1.

IF (pv2rlev = 1 or pv2rlev = 2 or pv2rlev = 3 or pv2rlev = 4 or pv2rlev = 5) p2level0 = 0.

IF (pv2rlev = 1) p2level1 = 1.

IF (pv2rlev = 0 or pv2rlev = 2 or pv2rlev = 3 or pv2rlev = 4 or pv2rlev = 5) p2level1 = 0.

IF (pv2rlev) = 2 p2level2 = 1.

IF (pv2rlev = 0 or pv2rlev = 1 or pv2rlev = 3 or pv2rlev = 4 or pv2rlev = 5) p2level2 = 0.

IF (pv2rlev = 3) p2level3 = 1.

IF (pv2rlev = 0 or pv2rlev = 1 or pv2rlev = 2 or pv2rlev = 4 or pv2rlev = 5) p2level3 = 0.

IF (pv2rlev = 4) p2level4 = 1.

IF (pv2rlev = 0 or pv2rlev = 1 or pv2rlev = 2 or pv2rlev = 3 or pv2rlev = 5) p2level4 = 0.

IF (pv2rlev = 5) p2level5 = 1.

IF (pv2rlev = 0 or pv2rlev = 1 or pv2rlev = 2 or pv2rlev = 3 or pv2rlev = 4) p2level5 = 0.

(... and so on for each of the five plausible values.)

The 30 dichotomous variables are not stored in the French Subsample database. Analysts who want to use these variables have to calculate them by themselves.

Once these 30 variables are imported into WesVar®, then the PV function can be used and results do not need to be imported in an Excel® spreadsheet to be combined.

Analysis of the questionnaire data

This section presents a suggestion for analysing the questionnaire data through the aggregation of variables.

Aggregating variables

Some variables from the student questionnaire can be aggregated to the school level for specific analysis since they represent measures of school climate or provide a proxy for the socio-economic status of the student body. Aggregation can be especially useful if one is carrying out multilevel analyses of performance. The amount of between-school variation with respect to these variables may also be of interest in and of itself (*i.e.*, outside student performance); for example, the between-school variability associated with the International Socio-Economic Index of Occupational Status (ISEI) gives an indication of the extent to which segregation by socio-economic levels occurs between schools. An added advantage of aggregation is that missing data items are reduced to zero at the school level. The variables in the student file that could provide useful school-level indicators include:

- School level International Socio-Economic Index of Occupational Status (ISEI on HISEI)
- Index of family wealth of the student body (WEALTH)
- Index of teacher support (TEACHSUP)
- Index of disciplinary climate (DISCEIMA)
- Index of teacher-student relations (STUDREL)
- Index of achievement press (ACHPRESS)
- Index of students’ sense of belonging in the school (BELONG).

An example of a SPSS® syntax for aggregating ISEI is provided in Figure 5.

Figure 5: *SPSS® syntax used to aggregate the International Socio-Economic Index of Occupational Status (ISEI) of the student level to the school level*

```
get file = ‘file with variable to be aggregated’.
sort by schoolid (a).
aggregate
        / outfile = ‘new file to contain aggregate variable(s)’
        /break = schoolid
        /schisei = mean(isei).

 *schisei is thus the aggregated isei.

get file = ‘school file to which aggregate isei is to be matched’.
sort by schoolid (a).

*both files need to be sorted in ascending order by the variable on which they are matched.
*the match variable must be a unique identifier for the school and in the same format in both *files. The variable schoolid is the match variable in this example. It was created by multiplying *the stratum ID (stidstrt) by 1.000.000 and adding it to the school ID (stidsch) in both files.
match files

        /file=*
        / table = ‘new file containing aggregate variable(s)’
        /by schoolid.

save outfile = ‘new file containing original school file plus new aggregate variables’.
execute.
```

The School File of the French Subsample

In the Swiss French Subsample database there is a school file for the French Subsample. There is also a school file for the national sample aviable at SIDOS. Persons interested in the international school file must consult the PISA web page of the OECD.

The responses to the school questionnaire

The school file contains the original variables collected through the school context questionnaire.

The names which are used to represent these variables in the database are directly related to the version of the school questionnaire. Each variable name consists of seven characters.

SC??Q??

The sixth and seventh characters refer to the item number of the question. For instance, SC02Q0l is the number of boys and SC02Q02 is the number of girls enrolled in the school.

The third and fourth characters refer to the question number as it appears in the school questionnaire. For instance, SC02 refers to the second question in the school questionnaire relating to enrolment.

The first two characters refer to the instrument:
SC for the school questionnaire.

The combination of the two files, school data and the student data

A recommendation is to analyse the school data at the student level. From a practical point of view, it means that the school data should to be imported into the student data file. From a theoretical point of view, while it is possible to estimate the percentages of schools following a specific school characteristic, it is not meaningful. Instead, the recommendation is to estimate the percentages of students following the same school characteristic. For instance, the percentages of private schools versus public schools will not be estimated, but the percentages of students attending a private school versus the percentage of students attending public schools will.

As school data will be imported in the student data file, the final weight and the replicates will be used in a similar way to how they are used for the student data.

The school questionnaire indices

As in the student questionnaire data file, two kinds of indices were derived from the school questionnaire data.

Indices derived through a direct combination of the answers from the school questionnaire

These indices, derived from the school questionnaire, are mainly related to the school size, the computer environment of the school and school staffing.

- ***School size (SCHLSIZE).*** This index represents the total enrolment in the school and is the sum of the number of boys (SC02Q0l) and the number of girls (SC02Q02) enrolled in the school.

- ***Percentage of girls (PCGIRLS)***. This index is the ratio between the number of girls and the total enrolment number of boys (SC02Q01) plus number of girls (SC02Q02) *i.e.*, the number of girls in the school divided by the total enrolment.

- ***School type (SCHLTYPE)***. A school was classified as either public or private according to whether a public agency or a private entity had the ultimate power to make decisions concerning its affairs. It was based on SC03Q01 and SC04Q0l to SC04Q04. It was further divided into three categories:[7]

  - Government-independent private schools were coded as 1, if the school principal reported that the school was controlled and managed by a non-governmental organisation (*e.g.*, a church, a trade union or a business enterprise) or if its governing board consisted mostly of members not selected by a public agency, where it received less than 50 per cent of its core funding from government agencies.

  - Government-dependent private schools were coded as 2, if the school principal reported that the school was controlled and managed by a non-governmental organisation (*e.g.*, a church, a trade union or a business enterprise) or if its governing board consisted mostly of members not selected by a public agency, where it received 50 per cent or more of its core funding from government agencies.

  - Government or public schools were coded as 3, if the school principal reported that the school was: controlled and managed directly by a public education authority or agency; or controlled and managed either by a government agency directly or by a governing body (council, committee, etc.), most of whose members were either appointed by a public authority or elected by public franchise.

- ***Hours of schooling per year (TOTHRS)***. This index was derived from the information which principals provided on: the number of weeks in the school year for which the school operates; the number of <class periods> in the school week; and the number of teaching minutes in a single <class period>. It consists of the total number of 60-minute hours of schooling per year. It was based on the product of the three factors, SC06Q01, SC06Q02, SC06Q03, divided by 60.

[7] For a definition of the types of school, see OECD (1998, p. 422).

- ***Number of computers per student per school (TCOMP)***. This index is the total number of computers in the school (SC13Q01), divided by the school size (SCHLSIZE).

- ***Proportion of computers available to l5-year-olds (PERCOMP1)***. This index is the number of computers available to 15-year-old students (SC13Q02), divided by the total number of computers in the school (SC13Q01).

- ***Proportion of computers available to teachers only (PERCOMP2)***. This index is the number of computers available only to teachers (SC13Q03), divided by the total number of computers in the school (SC13Q01).

- ***Proportion of computers available to the administrative staff (PERCOMP3)***. This index is the total number of computers available only to the administrative staff (SC13Q04), divided by the total number of computers in the school (SC13Q01).

- ***Proportion of computers with Internet access (PERCOMP4)***. This index is the number of computers connected to the Internet/World Wide Web (SC13Q05), divided by the total number of computers in the school (SC13Q01).

- ***Proportion of computers on a local network (PERCOMP5)***. This index is the number of computers connected to a local area network (LAN, Intranet) (SC13Q06), divided by the total number of computers in the school (SC13Q01).

- ***Student-teaching staff ratio (STRATIO)***. This index is the school size (SCHLSIZE) divided by the total number of teachers (SCl4Q01 + (0.5*SC14Q02), that is, part-time teachers contribute 0.5 and full-time teachers 1.0 to the total number of teachers). This rule applies to all indices based on question SC14.

- ***Proportion of teachers with a third level qualification [ISCED 5A (PROPQUAL)***. This index is the total number of teachers who have an <ISCED SA> qualification in <pedagogy> (SC14Q03 + (0.5*SC14Q04)) divided by the total number of teachers (SC14Q01 + 0.5*SC14Q02)).

- ***Proportion of teachers who are certified by the appropriate authority (PROPCERT)***. This index is the total number of teachers fully certified as teachers by <the appropriate authority> (SC 14Q05 + (0.5*SC14Q06)) divided by the total number of teachers (SC14Q01 + (0.5* SC14Q02)).

- ***Proportion of <language of assessment> teachers who have a third level qualification [ISCED 5A] (PROPREAD)***. This index is the total number of <language of assessment> teachers who have a third level qualification (SC14Q09 + (0.5*SC14Q10)) divided by the total number of teachers (SCl4Q0l + (0.5*SC14Q02)).

- ***Proportion of mathematics teachers who have a third level qualification [ISCED 5A] (PROPMATH)***. This index is the total number of mathematics teachers who have a third level qualification (SC14Q13 + (0.5*SC14Q14)) divided by the total number of teachers (SC14Q01 + (0.5*SC14Q02)).

- ***Proportion of science teachers who have a third level qualification [ISCED 5A] (PROPSCIE)***. This index is the total number of science teachers who have a third level qualification (SC14Q17 +

(0.5*5S14Q18)) divided by the total number of teachers (SCl4Q01 + (0.5*SC14Q02)).

Weighted likelihood estimate indices

The following indices from the school questionnaire were derived using the weighted estimate method (Warm, 1985):

- ***Index of the quality of schools' educational resources (SCMATEDU)***. This index is derived from school principals' reports on the extent to which learning by 15-year-olds in their school was hindered by: lack of instructional material; not enough computers for instruction; lack of instructional materials in the library; lack of multi-media resources for instruction; inadequate science laboratory equipment; and, inadequate facilities for the fine arts. It was based on questions SC11Q04 to SC11Q09. This index was inverted during reporting so that low values indicate a low quality of educational resources (OECD, 2001).

- ***Index of the quality of schools' physical infrastructure (SCMATBUI)***. This index was derived from principals' reports on the extent to which learning by 15-year-olds in their school was hindered by: poor condition of buildings; poor heating and cooling and/or lighting systems; and, lack of instructional space (*e.g.*, in classrooms). It was based on questions SC11Q01 to SC11Q03. This index was inverted during reporting so that low values indicate a low quality of physical infrastructure (OECD, 2001).

- ***Index of teacher shortage (TCSHORT)***. This index was derived from principals' views on how much learning by 15-year-old students was hindered by: shortage or inadequacy of teachers in general and shortage of teachers in the <language of assessment>, mathematics or science. It was based on questions SC21Q01 to SC21Q04. This index was inverted during reporting so that low values indicate problems with teacher shortage (OECD, 2001).

- ***Index of principals' perceptions of teacher-related factors affecting school climate (TEACBEHA)***. This index was derived from principals' reports on the extent to which the learning by 15-year-olds was hindered by: low expectations of teachers; poor student-teacher relations; teachers not meeting individual students' needs; teacher absenteeism; staff resisting change; teachers being too strict with students; and students not being encouraged to achieve their full potential. It was based on questions SC19Q01, SC19Q03, SC19Q07, C19Q08, SC19Q11, SC19Q14 and SC19Q16. This index was inverted during reporting so that lower values indicate a poorer disciplinary climate (OECD, 2001).

- ***Index of principals' perceptions of student-related factors affecting school climate (STUDBEHA)***. This index was derived from principals' reports on the extent to which learning by 15-year-olds in their school was hindered by: student absenteeism; disruption of classes by students; students skipping classes; students lacking respect for teachers; the use of alcohol or illegal drugs; and students intimidating or bullying other students. It was based on questions SC19Q02, SC19Q06, SC19Q09, SC19Q10, SC19Q13 and SC19Q15. This index was invented during reporting so that low values indicate a poorer disciplinary climate (OECD, 2001).

- ***Index of principals' perceptions of teachers' morale and commitment (TCMORALE)***. This index was derived from the extent to which school principals agreed with the following statements: the morale of the teachers in this school is high; teachers work with enthusiasm; teachers take pride in this school; and, teachers value academic achievement. It was based on questions SC20Q01 to SC20Q04.

- ***Index of school autonomy (SCHAUTON)***. School principals were asked to report whether teachers, department heads, the school principal, an appointed or elected board or an education authority at a higher level had the main responsibility for: appointing teachers; dismissing teachers; establishing teachers' starting salaries; determining teachers' salary increases; formulating school budgets; allocating budgets within the school; establishing student disciplinary policies; establishing student assessment policies; approving students for admittance to school; choosing which textbooks to use; determining course content; and deciding which courses were offered. The PISA index of school autonomy was derived from the number of categories that principals classified as not being a school responsibility. It was based on questions SC22Q011 to SC22Q12. This index was inverted during reporting so that high values indicate a high degree of autonomy.

- ***Index of teacher autonomy (TCHPARTI)***. School principals were asked to report whether teachers, department heads, the school principal, an appointed or elected board or an education authority at a higher level had the main responsibility for: appointing teachers; dismissing teachers; establishing teachers' starting salaries; determining teachers' salary increases; formulating school budgets; allocating budgets within the school; establishing student disciplinary policies; establishing student assessment policies; approving students for admittance to school; choosing which textbooks to use; determining course content; and deciding which courses were offered. The PISA index of teacher autonomy was derived from the number of categories that principals classified as being mainly the responsibility of teachers. It was based on questions SC22Q01 to SC22Q12.

These indices, based on weighted estimates (Warm, 1985), were standardised to have a mean of 0 and a standard deviation of 1 at the international level using the same procedures as were applied to the performance variables, *i.e.*, each OECD country, except the Netherlands[8], contributed equally to the standardisation.

[8] Response rate too low to ensure comparability (Annex A3, OECD (2001)).

The File with the Student Test Data

In the French Subsample database there is a item file for the French Subsample only. SIDOS provides also a file with the student test data of the national sample. Persons interested in the international item file must consult the PISA web page of the OECD.

The file with the test data (filename: FS ITEMS Grade9.POR) contains individual students' responses to all items used for the international item calibration and in the generation of the plausible values. All item responses included in this file have a one-digit format, which contains the score for the student on that item.

The PISA items are organised into units. Each unit consists of a piece of text or related texts, followed by one or more questions. Each unit is identified by a short label and by a long label. The units' short labels consist of four characters. The first character is R, M or S respectively for reading, mathematics or science. The three next characters indicate the unit name. For example, R083 is a reading unit called 'Household'. The full item label (usually seven-digit) represents each particular question within a unit. Thus items within a unit have the same initial four characters: all items in the unit 'Household' begin with 'R083', plus a question number: for example, the third question in the 'Household' unit is R083Q03.

Users may notice that the question numbers in some cases are not sequential, and in other cases, that question numbers are missing. The initial item numbering was done before the field trial, with some changes occurring after it (the field trial took place a year before the main assessment). For example, during the development of the main study instruments, some items were re-ordered within a unit, while others were deleted from the item pool.

In this file, the items are sorted by domain and alphabetically by short label within domain. This means that the mathematics items appear at the beginning of the file, followed by the reading items and then the science items. Within domains, units with smaller numeric IDs appear before those with larger IDs, and within each unit, the first question will precede the second, and so on.

Recoding of the assessments items

Some of the items needed to be recoded prior to the national and international scaling processes.

- Double-digit coded items (mathematics and science only) were truncated by retaining only the first digit, which corresponds to the score initially assigned to the item.

- Other items were recoded and/or combined. These items have been re-labelled. The character "T" was added to the end of the previous short label for such items.

- Numerical variables were recoded into scores, *i.e.*, incorrect answer (0), correct answers (1), missing answer (9) or not applicable (7).

- Some questions consisted of several true/false or yes/no items. Two questions were also composed of several multiple choice items (R088Q04 and R099Q03). These items were combined into new variables. The new codes correspond to the number of correct answers on the subset of items.

- Finally the items, which comprised a subset of items (R119Q09, R122Q01, R216Q03, R219Q01 and M192Q01), were combined to form new variables. The combined codes correspond to the number of correct answers to each of the sub-items included in these five items.

International scores assigned to the items

For the final scores allocated to the different categories, see OECD (2002c, Appendix 8). The codes are grouped according to the scores they were assigned for the final international calibration.

Making Comparisons

To test whether the means for two sub-groups (A and B) of students are different a *t*-test needs to be performed. The formula for the *t*-test is:

$$T = \frac{(\hat{m}_A - \hat{m}_B)}{\sqrt{\hat{s}^2_{(\hat{m}_A - \hat{m}_B)}}}$$

where $\hat{m}_A$ is the estimated mean of group A, $\hat{m}_B$ is the estimated mean of group B, and $\hat{s}^2$ is the estimated sampling variance for the difference in the means. The null hypothesis of equal means is rejected at the a level if $|T| > t_v(a)$, where $t_v(a)$ is the critical a value for the *t* distribution with ? degrees of freedom.

In general

$$\hat{s}^2_{(\hat{m}_A - \hat{m}_B)} = \hat{s}^2_{(\hat{m}_A)} + \hat{s}^2_{(\hat{m}_B)} - 2\,\mathrm{cov}(\hat{m}_A, \hat{m}_B)$$

where $\hat{s}^2_{(\hat{m}_A)}$ is the sampling variance for the estimated mean of group A, $\hat{s}^2_{(\hat{m}_B)}$ is the sampling variance for the estimated mean of group B, and $\mathrm{cov}(\hat{m}_A, \hat{m}_B)$ is the sampling covariance for the estimates of the two means. That is, the sampling variance for the difference between two means is equal to the sampling variance on the first mean (Group A), plus the sampling variance on the second mean (Group B), minus two times the covariance between the two means. If the two samples are independent, this covariance is 0, and the sampling variance of the difference simplifies to be the sum of the sampling variance for the estimates of the performance for each of the two groups.

Dependent versus independent samples

If the *samples are independent*, as is the case for countries in the PISA, the sampling variance for the difference between two countries will be the sum of their respective sampling variances.

If the *samples are not independent*, the covariance will need to be computed to accurately estimate the sampling variance of the difference. Two examples of dependent samples are: i) the sample of boys and the sample of girls within a particular country, and ii) the country sample and the OECD sample as the country sample contributes to the OECD parameter estimates (*e.g.*, when comparing the country mean estimate with the OECD average).

When samples are not independent, a way to estimate the sampling variance for the difference is to use the Fay's replicates (*e.g.* for reading: variables W_FSTR1 to W_FSTR80) included in the database. In the case of comparing a country mean estimate with the OECD mean estimate, the final estimate for the difference will be the difference between the country estimate and the OECD estimate, using the student final weight *i.e.* WTREAD, WTMATH or WTSCIE. To compute the sampling variance for the difference, it will be necessary to compute the difference for each replicate; then use these 80 estimates for the difference to compute the sampling variance on the difference, as mentioned on page 10 of this manual.

Another way to compute the sampling variance for the difference is to use the cell function in WesVar®.

Note: It is worth noting that the sampling variance for the difference between two independent samples can also be computed in WesVar®, using the replicates. But, given that a small covariance may be observed by chance, the results will be slightly different than when using the formulae for two independent samples.

The Bonferroni Adjustment

In the publication *Knowledge and Skills for Life First – Results from PISA 2000* (OECD, 2001) the Bonferroni adjustment was used in the test of significance in the multiple comparison tables and in the figures comparing each country mean estimate with the mean estimate of other countries (Figure 2.4, Figure 3.2, Figure 3.6, Table 2.2a, Table 2.2b, and Table 2.2c). The Bonferroni adjustment was not applied to the tests of significance included in any other tables or figures, including those that compare the country mean estimate and the OECD mean estimates.

In the table of multiple comparisons of achievement, the reader is more likely to compare one country with each of the other countries one at a time. Therefore, the Bonfennoni adjustment is based on 31 comparisons (that is, one country with the other 31 countries) and not 496 comparisons (that is, all possible pairwise comparisons (32*31)/2). With a type I error rate of 0.05, the critical value adjusted for 31 comparisons is approximately equal to 3.154, instead of 1.960.

Additional technical information and glossary

Calculation of correlation using plausible values

Let us suppose that one is interested in the correlation between the student reading ability, denoted X, and a context variable Y, collected through the student questionnaire. The correlation between X and Y, denoted *r*(X,Y)*, should be computed for each of the five plausible values. The correlation that has to be reported will be the average of the five computed correlations:

$$r^*(X,Y) = \frac{1}{5}\sum_{m=1}^{5} \hat{r}_m,$$

where $\hat{r}_m$ is the estimate of r computed using the m[th] plausible value.

The final estimate of *r* is the average of the estimates computed using each plausible value in turn. If U_m is the sampling variance for $\hat{r}_m$ then the sampling variance of r^* is:

$$V = U^* + (1 + M^{-1})B_M,$$

where $U^* = \frac{1}{m}\sum_{m=1}^{M} U_m$ and $B_M = \frac{1}{M-1}\sum_{m=1}^{M}(\hat{r}_m - r^*)^2$.

An a-% confidence interval for r^* is $r^* \pm t_v[(1-a)/2]V^{1/2}$ where $t_v(s)$ is the s percentile of the *t*-distribution with *V* degrees of freedom. $V = \dfrac{1}{\dfrac{f_M^2}{M-1} + \dfrac{(1-f_M)^2}{d}}$, $f_M = (1+M^{-1})B_M / V$

and *d* is the degrees of freedom that would have applied if q_m had been observed. In PISA the value of *d* will be equal to 80.

It is worth noting that the use of one plausible value will provide unbiased estimates of population parameters. However, the standard error estimated from the use of just one plausible value will contain the sampling variance component and not the measurement variance. It will therefore slightly underestimate the total uncertainty in the estimate.[9]

Double-digit coding

Students' responses could give valuable information about their ideas and thinking, besides being correct or incorrect. The marking guides for mathematics and science included a system of two-digit coding for marking so that the frequency of various types of correct and incorrect responses could be recoded. The first digit is the actual score. The second digit is used to categorise the different kinds of responses on the basis of the strategies used by the student to answer the item. There are two main advantages of using double-digit codes. Firstly, more information can be collected about students' misconceptions, common errors and different approaches to solving problems. Secondly, double-digit coding allows a more structured way of presenting the codes, clearly indicating the hierarchical levels of groups of codes. The assessment data files including the second digit were available to national centres.

Replication methodology for calculation of variance

The approach used for calculating sampling variances is known as Balanced Repeated Replication (BRR), or Balanced Half-Samples. A particular variant, known as Fay's method, has been used.

The variance estimator is:

$$V_{BRR} = \frac{1}{T(1-K)^2} \sum_{t=1}^{T} \left\{ \left(V_t^* - X^*\right)^2 \right\}$$

where X* is the estimate of a given statistic from the full sample, V_t^* a set of *T* replicate estimates and *K* the Fay's coefficient. For PISA 2000, 80 replicates were computed and the Fay's coefficient was set to *K* = 0.5. Therefore, the factor $\dfrac{1}{T(1-K)^2}$ is equal to $\dfrac{1}{20}$.

[9] B_M cannot be computed if just one plausible value is used.

SAS®

SAS® is a statistical package. For further information: http://www.sas.com.

SPSS®

SPSS® is a statistical package. For further information: http://www.spss.com.

WesVar®

WesVar® is a statistical package that computes estimates and their variance estimates from survey data using replication methods. The information generated can then be used to estimate sampling errors for different types of survey statistics. It can be used in conjunction with a wide range of complex sample designs, including multistage, stratified, and unequal probability samples. For further information: http://www.westat.com/wesvar.

References

Baumert, J., Gruehn, S., Heyn, S., Köller, O. and Schnabel, K.U. (1997), *Bildungsverläufe und Psychosoziale Entwicklung im Jugendalter (BIJU): Dokumentation — Band 1* , Max-Planck-Institut für Bildungsforschung, Berlin.

Baumert, J. , Heyn. S. and KöIler, O. (1994), *Das Kieler Lernstrategien-Inventar (KSI)*, Institut für die Pädagogik in den Naturwissenschaften an der Universität Kiel, Kiel.

Eignor, D., Taylor, C., Kirsch, I. and Jamieson, J. (1998), *Development of a Scale for Assessing the Level of Computer Familiarity of TOEFL Students*, TOEFL Research Report No, 60 Educational Testing Service, Princeton, NJ.

Ganzeboom, H.B.G., De Graaf, P. and Treiman, D.J. (with De Leeuw, J.) (1992), "A standard international socio-economic index of occupational status", *Social Science Research*, Vol. 21 (1), pp. 1 -56.

Marsh, H. W., Shavelson, R.J. and Byrne, B.M. (1992), "A multidimensional, hierarchical self- concept", in R. P. Lipka and T. M. Brinthaupt (Eds.), *Studying the Self: Self-perspectives across the life-span*, State University of New York Press, Albany.

Moser, U. (2001) *Für das Leben gerüstet? Die Grundkompetenzen der Jugendlichen – Kurzfassung des nationalen Berichtes PISA 2000*, edited by BFS/EDK, Reihe Bildungsmonitoring Schweiz, Neuchâtel.

OECD (1998), *Education at a glance — OECD indicators*, Paris.

OECD (1999a), *Classifying educational programmes: Manual for ISCED— 97 implementation in OECD countries*, Paris.

OECD (1999b), *Measuring student knowledge and skills: A new framework for assessment*, Paris.

OECD (2000), *Measuring student knowledge and skills: The PISA 2000 assessment of reading, mathematical and scientific literacy*, Paris.

OECD (2001), *Knowledge and skills for life: First results from PISA 2000*, Paris.

OECD (2002a), *Sample tasks from the PISA 2000 assessment: Reading, mathematical and scientific literacy*, Paris.

OECD (2002b), *PISA 2000 Technical Report*, Paris.

OECD (2002c), *Manual for the PISA 2000 Database*, Paris.

Owens, L. and Barnes, J. (1992), *Learning Preferences Scales*, ACER, Victoria, Australia.

Renaud, A. (2002). *Programme international pour le suivi des acquis des élèves (PISA). Plan d'échantillonnage pour PISA 2000 en Suisse*. Rapport de méthodes édité par l'Office fédéral de la statistique, Neuchâtel.

Warm, T.A. (1985), "Weighted maximum likelihood estimation of ability in Item Response Theory using tests of finite length", *Technical Report CGI- TR-85-08*, U.S. Cost Guard Institute, Oklahoma City.

Westat (2000), *WesVar complex samples 4.0, Rockville*, MD.

Zahner, C. et al. (2002*) Für das leben gerüstet? Die Grundkompetenzen der Jugendlichen – Nationaler Bericht der Erhebung PISA 2000*, edited by BFS/EDK, Serie Bildungsmonitoring Schweiz, Neuchâtel.