Chapter 2
Data and Methods Used for ICILS 2013

Abstract IEA's International Computer and Information Literacy Study (ICILS) was designed to establish how well students around the world were prepared for study, work, and life in the digital age. This chapter describes the ICILS 2013 study design, the sample design, scaling methods, and the variables used, and outlines the practical significance of particular results.

Keywords Computer and information literacy (CIL) · Gender differences · Information and communications technologies (ICT) · International Computer and Information Literacy Study (ICILS) · International large-scale assessments · Methodology

2.1 Sampling

This report is based on secondary analyses of student and teacher data from ICILS 2013 (Fraillon et al. 2015). ICILS 2013 gathered data from almost 60,000 grade eight (or equivalent) students and 35,000 teachers of grade eight students in more than 3300 schools from 21 countries. In each country, the samples were designed as two-stage cluster samples. During the first stage, schools were sampled with a probability proportional to the numbers of students enrolled in a school. Twenty students were then randomly sampled from all students enrolled in the target grade. In schools with fewer than 20 students, all students were invited to participate (Meinck 2015). These student data were augmented by data from almost 35,000 teachers in those schools. From the sampled schools, a minimum of 15 teachers was selected at random from all teachers teaching the target grade, but in schools with 20 or fewer such teachers, all teachers were invited to participate (Meinck 2015).

E. Gebhardt et al., *Gender Differences in Computer and Information Literacy*,
IEA Research for Education 8, https://doi.org/10.1007/978-3-030-26203-7_2

2.1.1 Data Collection

The main ICILS survey took place in the 21 participating education systems (18 countries and three benchmarking education systems) between February and December 2013 (the survey took place between February and June 2013 in the Northern Hemisphere countries, and between October and December 2013 in the Southern Hemisphere countries).

Students completed a computer-based test of CIL that consisted of questions and tasks presented in four 30-min modules. Each student completed two modules randomly allocated from the set of four, so that the total assessment time for each student was one hour (Fraillon et al. 2015). The psychometric properties of the student assessment have been reported by Gebhardt and Schulz (2015). After completing the two test modules, students completed a 30-min questionnaire (again on computer) that included questions relating to students' background characteristics, their interest in and enjoyment of using ICT, their experience and use of computers and ICT to complete a range of different tasks in school and out of school, and use of ICT during lessons at school (Schulz and Ainley 2015).

Teachers completed a 30-min online questionnaire about their background and familiarity with ICT, their confidence in using ICT, and their use of ICT in teaching in general and with a randomly-selected reference class. In this questionnaire, teachers were asked about the emphasis they placed on developing students' CIL, their views about the use of ICT in teaching and their participation in professional learning relating to pedagogical use of ICT. The properties of the student and teacher-based scales have been reported by Schulz and Friedman (2015).

2.1.2 Participation and Response Rates

Despite the efforts of participating countries and educational systems to meet the minimum response rates required, not all countries who participated in ICILS 2013 had data that allowed for further investigation in the current report. Fourteen countries met the minimum participation requirements for comparing student achievement and 12 countries met the minimum response rate requirement for teacher responses (Table 2.1). Germany and Norway met the student response rate criteria but failed to meet the teacher response rate criteria. Three benchmarking participants (Ontario in Canada, Newfoundland and Labrador in Canada, and the city of Buenos Aires in Argentina) also participated in ICILS 2013, however, in this report we focus only on full country participants.

Table 2.1 ICILS 2013 weighted survey response rates

Country	Overall student participation rate (%)	Met criteria for student survey	Overall teacher response rate (%)	Met criteria for teacher survey
Australia	86.3	Yes	79.0	Yes
Chile	93.4	Yes	95.9	Yes
Croatia	81.1	Yes	96.0	Yes
Czech Republic	93.7	Yes	99.9	Yes
Denmark	64.1	No	49.7	No
Germany	75.2	Yes (with replacements)	64.9	No
Hong Kong SAR	68.6	No	58.3	No
Republic of Korea	96.3	Yes	99.9	Yes
Lithuania	88.8	Yes	85.6	Yes
Netherlands	71.9	No	49.5	No
Norway (grade nine)	83.4	Yes	64.5	No
Poland	86.3	Yes	93.6	Yes
Russian Federation	92.8	Yes	98.4	Yes
Slovak Republic	92.3	Yes	97.7	Yes
Slovenia	90.0	Yes	88.1	Yes
Switzerland	43.5	No	27.2	No
Thailand	88.8	Yes	85.4	Yes
Turkey	85.8	Yes	95.8	Yes

Only those countries that met the following response rate requirements, either initially or after replacements were recruited, were included in the analyses in this report:

- an unweighted school response rate without replacement of at least 85% (after rounding to the nearest whole percent) and an unweighted overall student/teacher response rate (after rounding) of at least 85%, or
- a weighted school response rate without replacement of at least 85% (after rounding to the nearest whole percent) and a weighted overall student/teacher response rate (after rounding) of at least 85%, or
- the product of the (unrounded) weighted school response rate without replacement and the (unrounded) weighted overall student/teacher response rate of at least 75% (after rounding to the nearest whole percent).

2.1.3   Weighting of Data

One of the main objectives of any large-scale international study is to obtain estimates of population characteristics. In order to draw accurate conclusions about the population, researchers need to take into account the complex sample design implemented in all countries, in particular, the critical characteristic that sampling units do not have equal probability of selection. In addition, nonparticipation of schools, teachers, and students, in particular differential patterns of nonresponse, have the potential to bias results. To account for these complexities, sampling weights and nonresponse adjustments were calculated for each country, leading to an estimation (or "final") weight for each sampled unit. Further detailed information on the weighting procedures used in ICILS 2013 are available in the ICILS 2013 technical report (Fraillon et al. 2015). All findings presented in this report are based on appropriately weighted data.

2.2   Measures and Scales

In our analyses we used measures (based on responses to single items) and scales (constructed from responses to a number of similar items) that were derived for the ICILS 2013 international student assessment, and the student and teacher survey questionnaires. No new scales were created for the analyses reported in this volume. In this report, we considered four variables derived from the international student assessment.

2.2.1   Student Computer Literacy

The Rasch item response model (Rasch 1960) was used to derive the CIL scale from student responses to the 62 test questions and large tasks (which corresponded to a total of 81 score points). The final reporting scale was set to a metric that had a mean of 500 (the ICILS average score) and a standard deviation of 100 for equally-weighted national samples. Plausible value methodology with full conditioning was used to derive summary student achievement statistics. Student computer literacy is a dependent variable.

2.2.2   Student Performance Measures on CIL Strand Items

Similarly to the full measure of CIL, students' performance on seven strands of CIL items (creating information, transforming information, sharing information, accessing and evaluating information, managing information, knowing about and

understanding computer use, and using information safely and securely) was scaled to a mean of 500 with a standard deviation of 100. Student performances on different strand items were considered to be dependent variables.

2.2.3   Student Performance on CIL Item Types

As already noted, student performance on the three types of CIL items (large task, multiple choice, and constructed response items) was scaled to the common metric and these measures of student performance were considered to be dependent variables in some analyses.

2.2.4   Time Taken to Respond to Items

ICILS 2013 recorded the amount of time taken by students (in seconds) to respond to each test item. Time taken to respond to test items is used as a dependent variable in our analyses.

We used a number of other scales derived for ICILS 2013 for our analyses (Table 2.2). These are described in more detail in the relevant chapter of this report.

Table 2.2   ICILS 2013 scales used in this report

Chapters	Description of ICILS 2013 scale used
3	Students' confidence (ICT self-efficacy) in solving basic computer-related tasks (S_BASEFF)
3	Students' confidence (ICT self-efficacy) in solving advanced computer-related tasks (S_ADVEFF)
4	Students' interest and enjoyment in using computers and computing (S_INTRST)
4	Students' use of specific ICT applications (S_USEAPP)
4	Students' use of ICT for social communication (S_USECOM)
4	Students' use of ICT for exchanging information (S_USEINF)
4	Students' use of ICT for recreation (S_USEREC)
4	Students' use of ICT for (school-related) study purposes (S_USESTD)
4	Students' use of ICT during lessons at school (S_USELRN)
4	Students' reports on learning ICT tasks at school (S_TSKLRN)
5	Teachers' ICT self-efficacy (T_EFF)
5	Teachers' positive views on using ICT in teaching and learning (T_VWPOS)
5	Teachers' negative views on using ICT in teaching and learning (T_VWNEG)

Notes All ICILS scales referred to here are described in detail in chapter 12 of the ICILS 2013 technical report (Schulz and Friedman 2015)

2.3   Measures of Significance and Effect

In large-scale studies with many thousands of respondents, even small differences or correlations can be significant. An effect size provides a quantitative measure of the magnitude of the difference or correlation. In this report we use a "rule of thumb" measure of effect when we talk about the sizes of the statistically significant differences on either the CIL scale or the questionnaire scales as follows:

- We refer to the differences as "large" if the differences are larger than 50 points on the ICILS 2013 CIL scale (the international standard deviation was 100) or larger than five points on the ICILS 2013 questionnaire scales (the international standard deviation for these was 10);
- We refer to the differences as "moderate" if the differences are between 30 and 50 points on the ICILS 2013 CIL scale or between three and five points on the ICILS 2013 questionnaire scales;
- We refer to the differences as "small" if the differences are between 10 and 30 points on the ICILS 2013 CIL scale or between one and three points on the ICILS 2013 questionnaire scales; and
- We refer to the differences as "not meaningful" or "negligible" if the differences are less than 10 points on the ICILS 2013 CIL scale or less than one point on the ICILS 2013 questionnaire scales.

For correlations, we also provide Cohen's d as a measure of effect size. Cohen (1988) suggested the following labels for effect sizes for correlations:

- Strong if Cohen's $d = 0.8$;
- Moderate if Cohen's $d = 0.5$; and
- Insubstantial if Cohen's $d = 0.2$.

For further information about the development of the scales for ICILS 2013, and their psychometric properties, please refer to the ICILS 2013 technical report (Fraillon et al. 2015).

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY, USA: Routledge Academic.

Fraillon, J., Schulz, W., Friedman, T., Ainley, J., & Gebhardt, E. (2015). *ICILS 2013 technical report*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.

Gebhardt, E., & Schulz, W. (2015). Scaling procedures for ICILS test items. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report* (pp. 155–176). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.

Meinck, S. (2015). Sampling design and implementation. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report* (pp. 67–86). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.

Schulz, W., & Ainley, J. (2015). ICILS questionnaire development. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report* (pp. 23–36). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.

Schulz, W. & Friedman, T. (2015). Scaling procedures for ICILS questionnaire items. In J. Fraillon, W. Schulz, T. Friedman, J. Ainley & E. Gebhardt (Eds.), *ICILS 2013 technical report* (pp. 177–220). Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA). Retrieved from https://www.iea.nl/publications/technical-reports/icils-2013-technical-report.